Macroeconomic Theory II
Julio Garín
Spring 2017
Problem Set 0

Due Date: Beginning of class on Tuesday, January 10.
Instructions: Use LaTeX. Label all plots, series, and axes accordingly.

1. **GDP and Productivity.** The file `PS0_Data.txt` data that I have given you includes quarterly data on US real GDP per capita and US labor productivity (an index of output per hour) from the first quarter of 1947 to the third quarter of 2016.

   (a) Read the two time series into Matlab using the `load('PS0_Data.txt')` command, which will create a 275x2 matrix called `PS0_Data`, with the GDP data in the first column and the labor productivity index in the second column.

   (b) Extract the first column of the matrix `PS0_Data` into a vector called `gdp` and extract the second column into a vector called `prod`.

   (c) Calculate growth rates of each series two different ways: $gr_t^x = (x_t - x_{t-1})/x_{t-1}$ and $\ln gr_t^x = \ln x_t - \ln x_{t-1}$, where x is our variable of interest. Hint: you can calculate these growth rates as matrix operations; for example, `log(gdp(2:end))` is the vector of values 2 through 274 and `log(gdp(1:end-1))` is the vector of values 1 through 274 and so subtracting the latter from the former gives the entire sequence of growth rates.

   (d) Calculate the difference between the two growth rate measures (e.g. let `diff_gdp` equal the difference between the two growth rate measures for GDP). Then plot both measures, as well as their difference, all in one figure. Do this separately for `gdp` and for `prod`. When does the discrepancy between the two measures become noticeable?

   (e) In the mid-1990s, economists were trying to understand the 'productivity slowdown'—the observation that average productivity growth appeared to have slowed down in about 1974. By 2001, things appeared to have swung in the opposite direction—the average growth rate in the late 1990s appeared to have increased. To examine this, first plot the level (as opposed to the growth rate) of `prod`. Ignoring the 'wiggles' associated with business cycles, does the plot have the shape of an exponential function, as it would if there were roughly constant growth? Now plot the natural log of the level. Judging from this plot, does there appear to have been constant growth?

   (f) Calculate the mean of the growth rate of `prod` (using the `gr_prod` measure) for the full sample, and then for three subsamples: 1947-1974 (i.e. the first 108 growth rates), 1974-1997 (growth rates 109 through 200), and 1997-2015 (growth rates 201 through 265). Do your results support the idea of a 1974 'productivity slowdown' and a 1997 'productivity speed-up'?

   (g) Economists have devoted a lot of attention in recent years to what has been dubbed the 'Great Moderation,' that is, the observation that GDP appears to have become much less volatile since 1984. To examine this, calculate the standard deviation of the growth rate of GDP (using the `gr_gdp` measure) for the full sample, and for two sub-samples: pre-1984 (the first 148 growth rates) and post-1984 (all subsequent growth rates). Do you see the 'Great Moderation'?

2. **Sample Size, Law of Large Numbers, etc.** This problem has you create several samples of data, calculate sample statistics, and examine what happens to the statistics (e.g. in terms of their precision) as the sample size grows.

   (a) Use the `randn` command to create a 1,000x1 matrix (call it x) of a sample of Normally distributed (i.e. the Bell Curve) random variables. By default, `randn` creates random variables with a mean of 0 (the Bell curve is centered on 0) and a standard deviation of 1 (the standard deviation controls how 'dispersed' the Bell curve is around the mean).

(b) Use the `hist(x,20)` command to plot the histogram that places the values of the x sample into 20 bins. Does it look like a Bell Curve?

(c) Calculate the mean and standard deviation using the `mean` and `std` commands. Are they exactly equal to 0 and 1?

(d) Now we want to scale up these sample data so that they have mean 2 and standard deviation 3. If $x \sim N(0,1)$, then to create $y \sim N(2,3)$ we just need to multiply each data point by 3 and then add 2. You could do this as one big matrix operation (i.e. `y=2+3*x`), and that's the most efficient way to do it, but let's take advantage of this as an opportunity to do a loop in Matlab. That is, construct a loop that creates the matrix `y` by multiplying each element of `x` by 3 and then adding 2.

(e) Again use the `hist(y,20)` command to plot the histogram that places the values of the y sample into 20 bins. Do the mean and standard deviation seem to have changed as expected, judging from the change in the histogram relative to the first histogram?

(f) Now create a longer sample of 100,000 data points and again re-scale it so that they are from a distribution with a mean of 2 and standard deviation of 3. Plot the histogram again. Does it look more precisely like a Bell curve? Calculate the sample mean and sample standard deviation. Are they closer to 2 and 3 than they were in the case of the smaller sample?

(g) Now let's look at how the distribution of the sample statistics—the sample mean and the sample standard deviation—change with sample size. To do this, we'll now construct 500 samples of 1,000 observations, or data points, each. That is, use `randn(1000,500)` to construct a matrix, with each column representing one sample. Then transform the data again so that they come from a Normal distribution with mean of 2 and standard deviation 3. For each sample, calculate the mean and standard deviation (giving you 500 of each). Plot the histogram of both statistics; these two histograms show how the *sample statistics* are distributed. That is, the sample statistic is just a random variable that is a function of the actual sample. As each sample is different, so will the sample mean and sample standard deviation be different. Thus, the sample mean from a finite sample will not be exactly equal to the true mean. The histogram for the mean shows how the sample means are distributed.

(h) Now repeat the previous step, but let each of the 500 samples have 50,000 data points instead of just 1,000. Because the sample is larger, we expect—because of what is known as the *law of large numbers*—that the sample mean and the sample standard deviation will tend to be closer to the true mean and standard deviation of the distribution from which the data were generated. Do the results confirm this? That is, do the histograms show less dispersion around the true mean and standard deviation than in the case where there are just 1,000 data points per sample?

(i) We also expect, due to something known as the *central limit theorem*, that the distribution of the sample statistics will itself look more and more like a Normal distribution (Bell Curve) as the size of each sample gets larger. Note with care what this says: the distribution of *the sample statistics*, (i.e. not of the samples) converges toward the Normal distribution as the sample gets large. This would still be true even if the distribution of the data were different than Normal—for example the uniform distribution. Try this by using `rand` instead of `randn` to generate uniformly distributed samples to convince yourself of this point. You can just use the default for `rand`, which draws from a uniform distribution between 0 and 1 (and thus has a mean of 0.5 and a standard deviation of $\sqrt{1/12} \sim 0.2887$). Do the sample means and sample standard deviations look Normally distributed?