Answers to the Problem Set 4

Econometrics (Ph.D.)

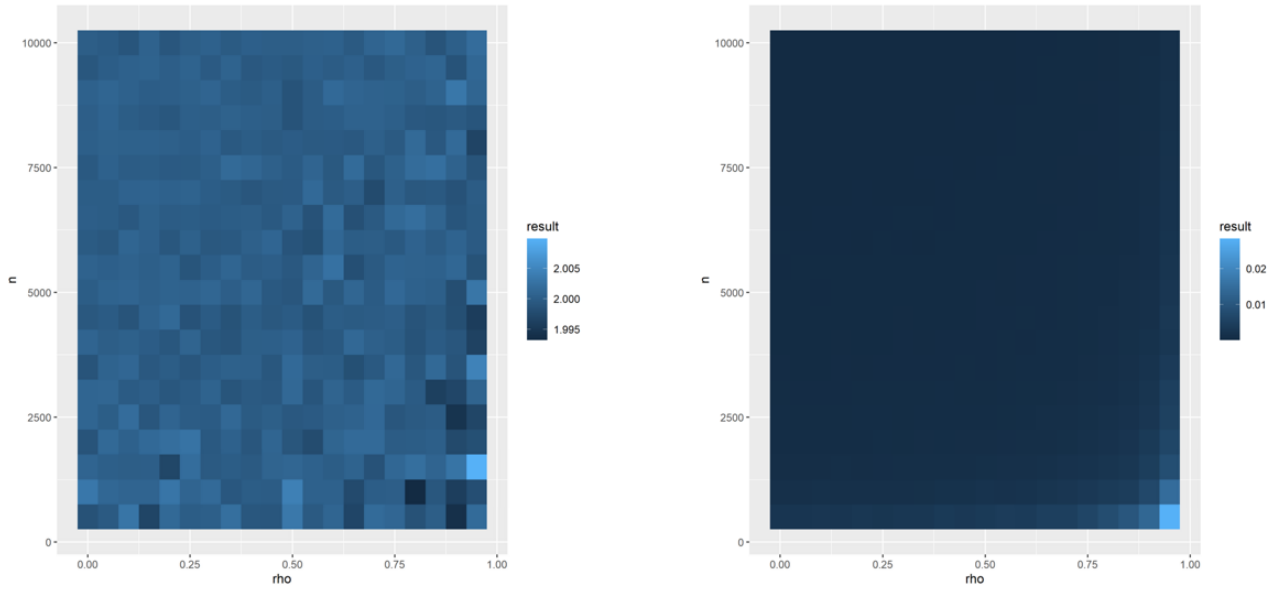
Sugarkhuu Radnaa

13 January, 2022

Problem 1

In Figure 1 below, we see the distribution (heatmap) of mean and variance of β_1 over different numbers of sample and correlation between X_1 and X_2. There is no clear pattern in the heatmap of the mean. This is simply due to the fact that the OLS estimator is unbiased. However, the heatmap of the variance shows that with lower number of sample and higher correlation between Xs (i.e., ρ), the variance of β_1 tends to be higher. This owes to the violation of the multicolinearity assumption of the Gauss-Markov theorem since the Xs are highly correlated when ρ is high, thus making it harder to pin down the value of the parameter.

Figure 1: Mean (left) and Variance (right) of $beta_1$ over different number of draws (from 500 to 10000, bottom-top) and rho (from 0 to 0.95, left-right)



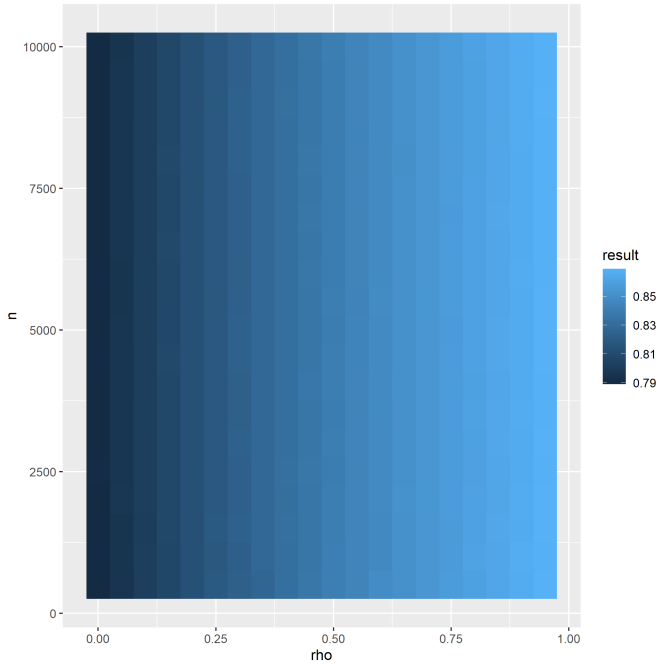
The issue can be partly amended by increasing the number of samples as we see that the variance decreases gradually even when ρ is almost equal to 1 with the increase in the sample length. However it is not nearly equal to the case of low correlations as shown in Table 1 below.

Table 1: Variance of $beta_1$ (times a factor of 1000)

		rho	
		0.00	**0.95**
n	**500**	2.52	28.33
	10000	0.13	13.97

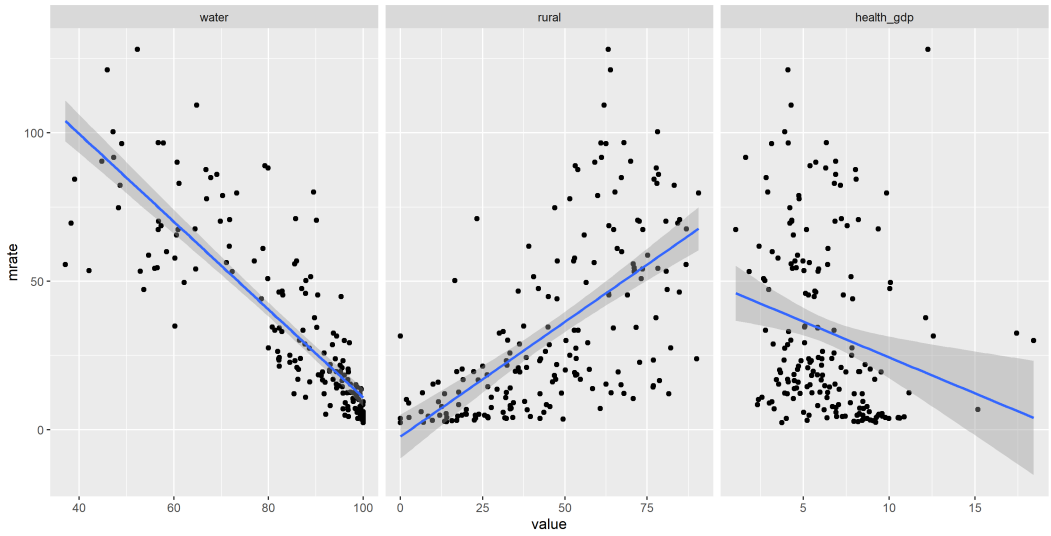
The R^2 increases as ρ increases. That is because, by construction, both of X_1 and X_2 are highly correlated and will be generating Y which is more predictable (staying similar to both than being pulled between two completely different random variables) than the otherwise.

Figure 2: Coefficient of determination (R^2)



Problem 2

Figure 1: Mortality rate versus the other indicators



The estimated model is as below:

$$mrate_i = -1.33 \cdot water_i + 0.18 \cdot rural_i + 0.22 \cdot health\_GDP_i + 136.71$$

(a) The coefficients of the "water" and "rural" are as expected. The more the clean water is available for more people, the better the situation in the country and thus the lower mortality rate. Likewise, the more people in the remote (rural) location, the less readily available are the resources, thus higher mortality rate could result. Surprisingly, the coefficient of $health\_GDP$ is positive. The more the country spends on its health in terms of GDP, the higher the mortality rate will be, the results says.

(b) β_1, the coefficient of "water" is -1.33. As the slope of conditional mean function, it says that a country that has 1.0 percentage point higher share of population that has access to a clean water resource on average has 1.33 percentage lower mortality rate.

(c) As a causal effect, it says that, ceteris paribus, 1.0 percentage point increase in the share of population that has access to a clean water resource will result in 1.33 percentage decline in the mortality rate.

Let us consider each of the three requirement in turn.

Linearity: The law of diminishing marginal productivity could be present in this case. The less the water availability, the higher the impact of increasing its availability to the population, whereas the coefficient could be just averaging over the different rate of impacts.
Homogeneity: The effect could be changing to variations in other factors or it could be different across different countries. However, with no more data available, we could assume that this

condition is hard to refute.
Exogeneity: It is hard to believe that this condition is satisfied. The water is likely to be more abundant in countries that are more developed since they should have water in the first place to be more productive as an economy. In turn, the more developed the country, the better the health system and other environment supporting the healthcare system. So β_1 might be accompanying the impact of many other correlated variables.

(d) $R^2 = 72.9\%$