Answers to the Problem Set 5

Econometrics (Ph.D.)

Sugarkhuu Radnaa

27 January, 2022

Problem 1

(a) The probability that β_1 is in the 95% confidence interval for different n and ρ are shown in the table below:

Figure 1: Probability that β_1 is 95% confidence interval for different values of ρ and n (1000 different sampling)

rho\n	100	500	1000	2500	5000	10000
-0.9	0.949	0.949	0.966	0.934	0.947	0.939
-0.6	0.944	0.95	0.953	0.958	0.946	0.954
-0.3	0.956	0.943	0.953	0.951	0.947	0.941
0.0	0.952	0.961	0.948	0.959	0.939	0.951
0.3	0.955	0.963	0.938	0.941	0.935	0.945
0.6	0.945	0.955	0.927	0.948	0.949	0.957
0.9	0.951	0.955	0.943	0.955	0.953	0.956

As is clear, it is almost always about 95%. If we decrease number of different draws from 1000 to less than hundred,

(b) Lower n and higher ρ gives wider confidence interval. With larger sample length, the confidence interval narrows but the impact of high ρ is still present to some extent.

Figure 2: Length of 95% confidence interval of β_1 for different values of ρ and n (1000 different sampling)

rho\n	100	500	1000	2500	5000	10000
-0.9	1.07	0.47	0.33	0.21	0.15	0.1
-0.6	0.59	0.25	0.18	0.11	0.08	0.06
-0.3	0.49	0.21	0.15	0.09	0.07	0.05
0.0	0.47	0.2	0.14	0.09	0.06	0.05
0.3	0.49	0.21	0.15	0.1	0.07	0.05
0.6	0.58	0.25	0.18	0.11	0.08	0.06
0.9	1.08	0.47	0.33	0.21	0.15	0.1

(c) Since OLS is unbiased and consistent estimator, the probability is more or less in line with the significance level of 1%.

Figure 3: Probability (1% significance) to reject $H_0 : \beta_2 = 0$ for ρ and n (1000 different sampling)

rho\n	100	500	1000	2500	5000	10000
-0.9	0.011	0.012	0.005	0.011	0.009	0.01
-0.6	0.014	0.012	0.008	0.007	0.012	0.006
-0.3	0.009	0.01	0.009	0.007	0.011	0.011
0.0	0.015	0.004	0.009	0.011	0.011	0.015
0.3	0.008	0.007	0.009	0.01	0.013	0.01
0.6	0.009	0.015	0.018	0.01	0.008	0.008
0.9	0.012	0.007	0.016	0.015	0.01	0.004

(d) When n is smaller and ρ is higher (in absolute term), uncertainty about the estimation is high (power is low) and sometimes we cannot reject the null. However, once the sample sample increases from 100 to 500, the rejection increases notably (power improves), although at high ρ, the issue is still present. Once we move beyond 2500 sample size, power of the test is above 90% for any ρ,

Figure 4: Probability (1% significance) to reject $H_0 : \beta_2 = 0$ for ρ and n (1000 different sampling)

rho\n	100	500	1000	2500	5000	10000
-0.9	0.03	0.194	0.42	0.87	0.997	1
-0.6	0.123	0.696	0.957	1	1	1
-0.3	0.132	0.861	0.996	1	1	1
0.0	0.191	0.897	0.999	1	1	1
0.3	0.16	0.875	0.994	1	1	1
0.6	0.104	0.712	0.951	1	1	1
0.9	0.031	0.188	0.409	0.906	0.998	1

Problem 2

(a) The estimated model is as below:

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

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 in the country, the results says.

(b) Homoskedastic and heteroskedastic standard errors (in yellow) of the regression parameters and other statistics are shown in the table below. All coefficients but β_4 are significantly different from 0 (below 1% percent significance level).

Figure 5: Regression output and VCOV matrix of the coef.

HOMOSKEDASTIC

	est	std	null	lower	upper	p_value
beta_1	-1.33	0.09	0.00	-1.51	-1.16	0.00
beta_2	0.18	0.06	0.00	0.06	0.30	0.00
beta_3	0.22	0.44	0.00	-0.66	1.09	0.63
beta_4	136.71	9.66	0.00	117.65	155.78	0.00

	water	rural	health_gdp	intercept
water	0.01	0.00	-0.01	-0.77
rural	0.003	0.004	0.002	-0.47
health_gdp	-0.01	0.00	0.20	-0.80
intercept	-0.77	-0.47	-0.80	93.38

HETEROSKEDASTIC

	est	std	null	lower	upper	p_value
beta_1	-1.33	0.12	0.00	-1.57	-1.09	0.00
beta_2	0.18	0.06	0.00	0.07	0.30	0.00
beta_3	0.22	0.45	0.00	-0.67	1.10	0.63
beta_4	136.71	11.65	0.00	113.74	159.69	0.00

	water	rural	health_gdp	intercept
water	0.01	0.00	-0.02	-1.35
rural	0.004	0.003	-0.004	-0.47
health_gdp	-0.02	0.00	0.20	0.90
intercept	-1.35	-0.47	0.90	135.61

(c) Taking $t_{97.5\%,183}$ and multiplying it with the standard error (either homosk. or heterosk.) of $\hat{\beta}_2$, we arrive at the 95% confidence interval for $\hat{\beta}_2$. That is [0.057;0.303] for homoskedastic and [0.065;0.296] for heteroskedastic.

(d) Using again $t_{97.5\%,183}$ and homoskedastic variance-covariance matrix of betas (with R given in the question, $R' \cdot V_\beta \cdot R$), 95% confidence interval for $\hat{\theta}$ is [30.72;35.22]. And [30.75;35.20] for heteroskedastic adjusted version. $\hat{\theta}$ itself equals 32.97.

(e) Subtracting those the values in concern for each variable from the corresponding values in X, we get something like demeaned X. Now all the coefficients are the same as previous section, but the intercept (β_4) is 32.97, the same as $\hat{\theta}$.

(f) It is given $R = [1, 10, 0, 0]'$. Homoskedastic Wald statistic is 0.47 and heteroskedastic Wald statistic is 0.50. On the other hand, 95% percentile value of χ^2_1 is 3.84.

Hence, we can not reject the null hypothesis that one percentage point increase of the population of country i that has access to a clean water is completely offset by 10 percentage point increase of the population of country i that lives in a rural location.