ADVANCED ECONOMETRICS

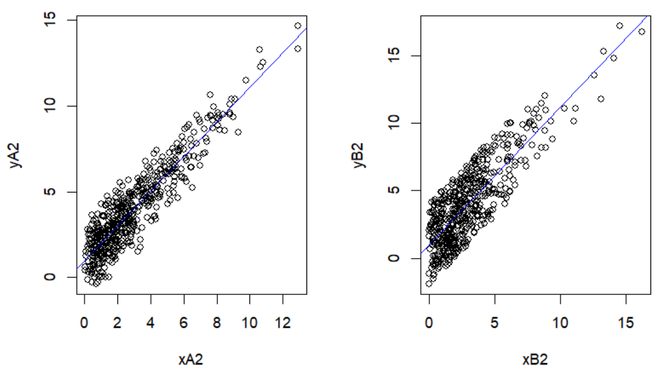
Homework 2

Fabio Marcaurelio

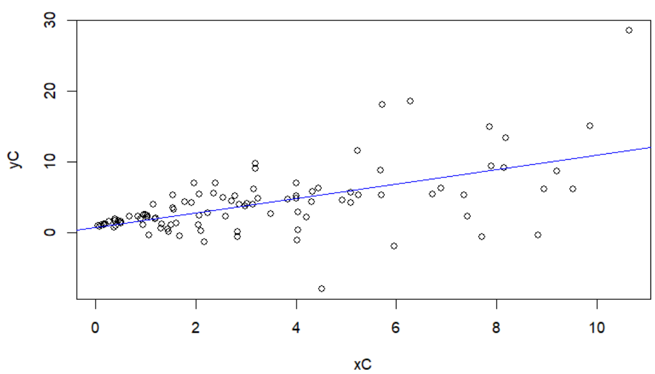
1 GENERAL DISCUSSION OF THE RESULT

1.1 does homoskedasticity hold in this case?

we simulate 1000 samples from three types of distribution. In the first we assume that $u_i \sim N(0,1)$ and $x_i \sim \chi^2_3$, in this case we can assume the homogeneity of the variance of u_i with respect to x_i since we know by construction that $E[u_i|x_i] = 0$. From these assumption we can deduce that homoskedasticity holds. In R we take the difference between the variance of coefficients under the assumption of homoskedasticity and under the robust heteroskedasticity, difference is almost zero. if we consider the second case in which $u_i \sim U(-3,3)$ and $x_i \sim \chi^2_3$ result is not different since the assumption on the relation in u_i and x_i is the same, then homoskedasticity holds.
If we consider $n = 50$ and $n = 500$ we know that the result on variances can be different, if we increase n to a large number (500) variances of simulations decrease and then the distribution of the coefficients will be approximated to $N(\beta_i, V)$
For the third distribution we have a different assumption, x_i does not change but $u_i \sim N(0, x_i^2)$. This breaks the assumption of the independence of variance then homoskedasticity can not works. this is verifiable from the different variance distribution of estimated beta under the assumption of homoskedasticity and robust to heteroskedasticity.
We can also check it by the plot of y_i for all of three cases, in the first two pictures we have the plot for ya and yb for $n = 500$, for n=50 result is similar.



in the last picture we can deduce that homoskedaticity assumption can not holds:



At the same time, if we calculate the variance covariance matrix both for homoskedasticity and robust to heteroskedasticity with $R = 1000$ simulations we can show that for A and B there are no Significant differences in these matrices. for scenario C these results are not true and then this is another prove of the presence of heteroskedasticity.

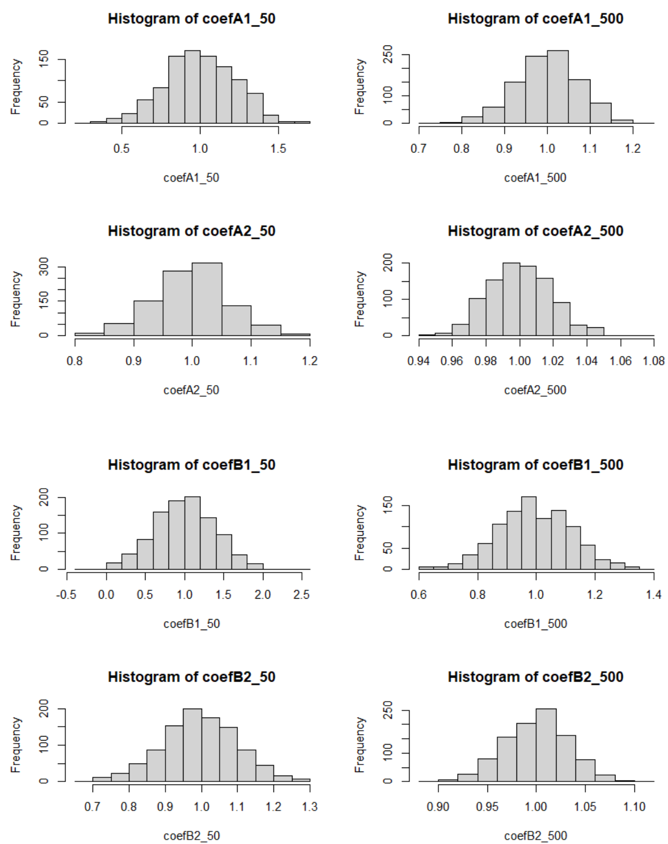
1.2 quantify the quality of the asymptotic approximation when n = 50 and n = 500

We know that the distribution of β_i can change with respect to the size of n. In particular, if we fix n (not $+\infty$) central limit theorem can not works, then we have to assume that $u_i|x_i \sim N(0, \sigma^2)$ under homoskedasticity, $\hat{\beta} \sim N(\beta, v)$

and finally that

$$\frac{\hat{\beta}_i - \beta}{se(\hat{\beta}_i)} \sim t_{n-k}$$

. If $n + \infty$ then t collapse to a normal, in general this is true from $n = 100$. graphically if we plot the distribution of $\hat{\beta}_i$ we can check that when $n = 50$ tails are more "fat". We can check it both for A and B:



if we consider a confidence interval, we can compare for A and B CI with normal approximation and with t-student approx. if we do that for $n = 500$ there are no different between the two methods, for $n = 50$ different for both A and B are negligible but exist.
For reach this result i simulate $R = 1000$ confidence interval for normal approximation and t-student, both under homoskedasticity and robust to heteroskedasticity. in this way i can compare the coverage of beta in each simulation and for different assumption.

1. Confidence Interval under N

$$(\hat{\beta}_i - z_{1-\alpha/2} \cdot SE(\hat{\beta}_i) \leq \beta \leq \hat{\beta}_i + z_{1-\alpha/2} \cdot SE(\hat{\beta}_i)$$

2. Confidence interval under $t - student$

$$(\hat{\beta}_i - t_{1-\alpha/2,n-1} \cdot SE(\hat{\beta}_i) \leq \beta \leq \hat{\beta}_i + t_{1-\alpha/2,n-1} \cdot SE(\hat{\beta}_i)$$

For scenario C we have some differences, these are given by the presence of heteroskedasticity and then we have to consider the robust confidence interval. For $n = 100$ we can see that CI that contains mean of beta are more close to 0.95 in normal approximation with respect to $t - student$ for a very small difference.

1.3 Explore the 0.95 CI coverage with the two ways of computing the standard errors (under homoskedasticity and heteroskedasticity).

We can simulate $R = 1000$ confidence interval and test the coverage level by the method introduced above. Result are interesting, we know that for scenario A and B homoskedasticity assumption holds, in fact we have no significant differences between CI tested under this assumption and under robust one. For C, as we can expected, confidence intervals under homoskedasticity are far from the 0.95 coverage. If we consider these CI under robust assumption we reach a 0.95 covarage.
Previous result are very coherent with the assumption that we have done in the first point.
for completeness we do the same test for zero, we check for the coverage of confidence interval at 0.95 under homoskedasticity and robust to heteroskedasticity and the reault are the opposite with respect to one. In this case all scenarios given coefficient in $n = 50$ and $n = 500$ different from zero.

1.4 Build the test statistic for the null hypothesis H0 : BETA(2) = 1. What is the share of rejection of the null hypothesis in the simulations?

For these test we take always the robust standard error. we start defining a t-test as:

$$t = \frac{\hat{\beta}_2 - 1}{SE(\hat{\beta}_2)}$$

under $H_0 : \beta_2 = 1$. results are:

1. A- $n = 50$
   we don't reject the null hypothesis at -1.200968 level of significant, since $t > -2.009$

2. A- $n = 500$
   we don't reject the null hypothesis at 0.5876805 level of significant, since $t < 1.96$

3. B- $n = 50$
   we don't reject the null hypothesis at 0.3154873 level of significant, since $t < 2.009$

4. B- $n = 500$
   we don't reject the null hypothesis at 0.5702663 level of significant, since $t < 1.96$

5. C- $n = 100$
   we don't reject the null hypothesis at 0.1130735 level of significant, since $t < 1.96$

These results are strictly related to the coverage of above confidence intervals, as in the previous point we obtain a 0.95 coverage and now we prove that these coefficient are equal to one at 0.95 confidence level.

1.5 Build the test statistic for the null hypothesis H0 : BETA(2) = 0. What is the share of rejection of the null hypothesis in the simulations?

For these test we take always the robust standard error. we start defining a t-test as:

$$t = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)}$$

under $H_0 : \beta_2 = 0$. results are:

1. A- $n = 50$
   we reject the null hypothesis at 14.97889 level of significant, since $t > 2.009$

2. A- $n = 500$
   we reject the null hypothesis at 56.03694 level of significant, since $t > 1.96$

3. B- $n = 50$
   we reject the null hypothesis at 7.16864 level of significant, since $t > 2.009$

4. B- $n = 500$
   we reject the null hypothesis at 36.77314 level of significant, since $t > 1.96$

5. C- $n = 100$
   we reject the null hypothesis at 4.090049 level of significant, since $t > 1.96$