Econometrics ECON 662D1

Assignment 3

Jean-Sébastien Matte

12/01/2020

Considering the linear regression

$$y_t = \alpha + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \beta_3 \boldsymbol{x}_3$$

we present the null and alternative hypotheses for case I and II:

Case I

For case I, we want to test the hypothesis that $\beta_1 = 0$. Therefore, the null and alternative hypotheses are:

$$H_0 : \beta_1 = 0 \implies y_t = \alpha + \beta_2 \boldsymbol{x}_2 + \beta_3 \boldsymbol{x}_3$$
$$H_A : \beta_1 \neq 0 \implies y_t = \alpha + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \beta_3 \boldsymbol{x}_3$$

Case II

For case II, we want to test the hypothesis that $\beta_1 = 0.1$. Therefore, the null and alternative hypotheses are:

$$H_0 : \beta_1 = 0.1 \implies y_t = \alpha + 0.1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \beta_2 \boldsymbol{x}_2$$
$$\implies y_t - 0.1 \boldsymbol{x}_1 = \alpha + \beta_2 \boldsymbol{x}_2 + \beta_2 \boldsymbol{x}_2$$
$$H_A : \beta_1 \neq 0.1 \implies y_t = \alpha + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \beta_3 \boldsymbol{x}_3$$

Models under the null hypothesis are *restricted*, while the models under the alternative hypotheis are *unrestricted.* From this point onwards, all regression results from the *unrestricted* models will be denoted with a hat over the variables, and all regression results from the *restricted* models will be denoted with a tilde over the variables.

Furthermore, we denote $\boldsymbol{X}_r$ the restricted matrix of covariates. For both cases, this matrix corresponds to the original matrix, with the second column, i.e. the column corresponding to covariate x_1, removed.

The procedures followed to obtain results for each of the three methods we were asked to consider is presented below.

Simple Student t:

The p-value for the simple Student t test is obtained directly from the regression summary. This test consitsts at calculating a t statistic from the least-quare estimate $\hat{\beta}_1$ against the null hypothesis such that

$$t = \frac{\hat{\beta}_1 - \beta_{10}}{\hat{s}_{\hat{\beta}_1}}$$

and obtaining the p-value from Student t distribution with $n - 4$ degrees of freedom, where n is the sample size. Note that in the notebook attached, for the output of the t test for β_1, labelled 'Test for Constraints', `'c0'` corresponds to β_1. It was impossbible to relabel the constraint to follow the notation of our model.

Bootstrap using rescaled, restricted residuals:

We proceed the following way to perform the bootstrap of rescaled, restricted disturbances.

1. We first obtain the least-square residuals $\tilde{\boldsymbol{u}}$ and regression coefficients $\tilde{\boldsymbol{\beta}}$ directly from the statistical package by regressing the original observations $\boldsymbol{y}$ on the covariates $\boldsymbol{X}_r$ of the *restricted* model. Recall that the *restricted* model, for both cases, corresponds to the model of the null hypothesis. We then rescale the residuals with scale factor
$$\frac{n}{n-k}$$
with n equal to the sample size (ranging from 20 to 100), and $k = 3$.

2. For each of the $B = 999$ bootstrap iterations:

   (a) Randomly sample the rescaled, restricted residuals. The random sample has the same number of residuals than the original regression, i.e. we resample n rescaled restricted residuals. Denote the n-vector of resampled rescaled restricted residuals as $\tilde{\boldsymbol{u}}^*$.

   (b) Calculate a vector of updated observations, denoted $\boldsymbol{y}^*$, obtained from
$$\boldsymbol{y}^* = \boldsymbol{X}_r\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{u}}^*$$

   (c) Regress the updated observations $\boldsymbol{y}^*$ on the original covariate matrix $\boldsymbol{X}$.

   (d) Obtain the t statistic for coefficient β_1 directly from the regression summary. Note that the calculation to get the t statistic is identical to that presented for the simple Student t test above. Store the statistic.

3. With the B bootstrap staticstics collected from each iteration, we are able to compute the bootstrap p-value from the follwing equation
$$p_{boot}(\hat{\tau}) = 2\min\left\{\frac{1}{B}\sum_{b=1}^{B}\mathbb{1}_{(\hat{\tau}_b^* \leq \hat{\tau})}; \frac{1}{B}\sum_{b=1}^{B}\mathbb{1}_{(\hat{\tau}_b^* > \hat{\tau})}\right\}$$
where $\hat{\tau}_b^*$ correspond to each of the bootstrap t statistics collected, and $\hat{\tau}$ to the t statistic between the original *unrestricted* model and the null hypothesis. The latter corresponds to the t statistic calculated for the previous method presented above.

Wild Bootstrap:
We perform the following way for the Wild Bootstrap. For each of the $B = 999$ bootstrap iterations:

1. We first define the pmf distribution that will be used to rescale the restricted disturbances. Here, we use the pmf distribution as proposed by Mammen. Denote s_t^* each drawing from this pmf distribution corresponding to each observation in the sample.

2. Multiply the *restricted* residuals $\tilde{u}_t$ of each of the observations in the samples with the corresponding s_t^*.

3. Generate a set of updated observations, labelled $\boldsymbol{y}^*$, obtained from
$$y_t^* = \boldsymbol{X}_t\tilde{\boldsymbol{\beta}} + s_t^*\tilde{u}_t$$

4. Regress the updated observations $\boldsymbol{y}^*$ on the original covariate matrix $\boldsymbol{X}$.

5. Obtain the t statistic for coefficient β_1 directly from the regression summary. Note that the calculation to get the t statistic is identical to that presented for the simple Student t test above. Store the statistic.

With the B bootstrap staticstics collected from each iteration, we are able to compute the bootstrap p-value from the follwing equation
$$p_{boot}(\hat{\tau}) = 2\min\left\{\frac{1}{B}\sum_{b=1}^{B}\mathbb{1}_{(\hat{\tau}_b^* \leq \hat{\tau})}; \frac{1}{B}\sum_{b=1}^{B}\mathbb{1}_{(\hat{\tau}_b^* > \hat{\tau})}\right\}$$

where $\hat{\tau}_b^*$ correspond to each of the bootstrap t statistics collected, and $\hat{\tau}$ to the t statistic between the original *unrestricted* model and the null hypothesis. The latter corresponds to the t statistic calculated for the previous method presented above.

Tables **??** and **??** summarize the results of the p-values for the 17 different sample sizes, for each case studied. Note that a detailed code of how all resutls were obtained, as well as all regression summaries for OLS estimation of the *unrestricted* models for each sample size, can be found in the notebook attached.

Sample Size (n)	**Student t**	**Bootstrap**	**Wild Bootstrap**
20	0.7192	0.7267	0.649
25	0.5739	0.5946	0.470
30	0.3784	0.3623	0.258
35	0.4822	0.4965	0.330
40	0.4426	0.4464	0.338
45	0.5183	0.5205	0.478
50	0.4841	0.4845	0.420
55	0.4980	0.5225	0.452
60	0.5385	0.4985	0.462
65	0.3234	0.3624	0.284
70	0.3323	0.2903	0.302
75	0.3656	0.3443	0.338
80	0.3508	0.3644	0.298
85	0.2184	0.2342	0.188
90	0.2883	0.2482	0.246
95	0.1404	0.1141	0.088
100	0.2224	0.2082	0.186

Table 1: Summary of Results for Case I

Figure **??** presents plots of the p-values obtained from the different methods, as a function of the sample size for Case I. We do not present a similar figure for Case II as all resutls are 0.

From Figure **??**, we see that the three methods yield similar p-values for the hypothesis, which is what one would expect to get for running a linear regression on a (relatively) small dataset. Before looking further into the similarity between the methods, we present in Figure **??** the distribution of the residuals for the unrestricted model, and the restricted model of Case I, and in Figure **??** the residuals plot, again for Case I only.

We first note from Figure **??** that residuals have mean 0, and that clearly, they are not distributed Normally. This would indicate that a resampling bootstrap like the one used is more appropriate in this case than an simple OLS regression. It seems however that distributions of *restricted* and *unrestricted* models are fairly similar, and that possibly because of the sample size, this explains why the OLS performs similarly to the resampling bootstrap. We further note from Figure **??** that there is no clear pattern in the residuals, and therefore not pointing to any heteroskedasticity. Foundations of Econometrics indicate that there is no reason to believe the Wild Bootstrap would yield different results than the resampling bootstrap method, even if disturbances are homoskedastic, which is clearly the case here

Finally, results show that for Case II, all methods, for all sample sizes, unanimously reject the null. Looking at the regression summary and other results for the full sample, we can see that the estimated coefficient $\hat{\beta}_1$ and its confidence interval are far from the 0.1 value the null hypothesis is assuming. From this, it is easily understood that the null will be rejected with a small p-value.

Sample Size (n)	**Student t**	**Bootstrap**	**Wild Bootstrap**
20	0	0	0
25	0	0	0
30	0	0	0
35	0	0	0
40	0	0	0
45	0	0	0
50	0	0	0
55	0	0	0
60	0	0	0
65	0	0	0
70	0	0	0
75	0	0	0
80	0	0	0
85	0	0	0
90	0	0	0
95	0	0	0
100	0	0	0

Table 2: Summary of Results for Case II

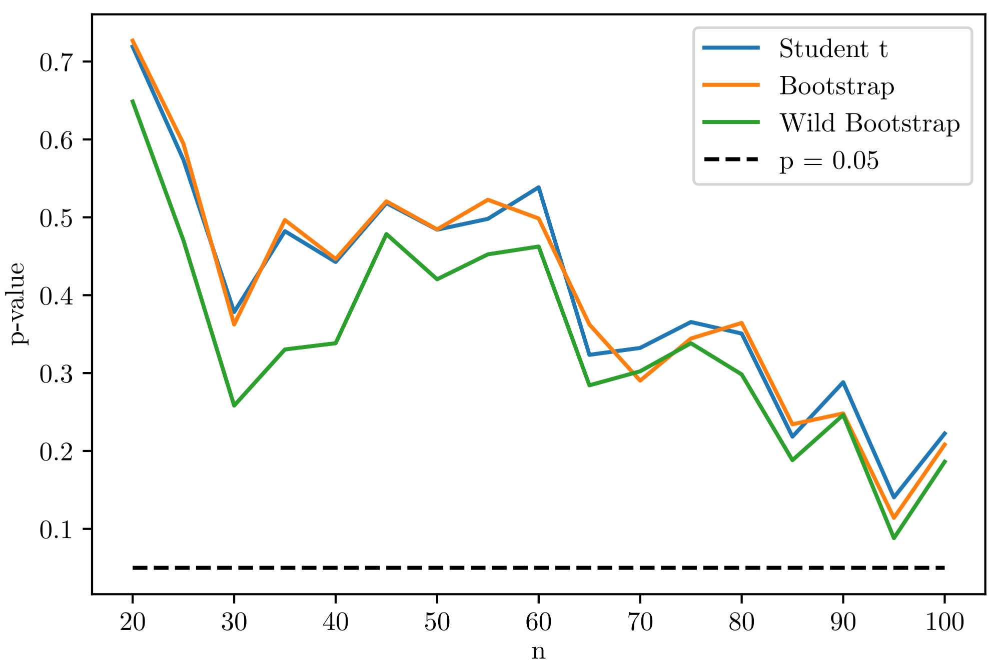

Figure 1: Variation of p-value between methods and sample size - Case I

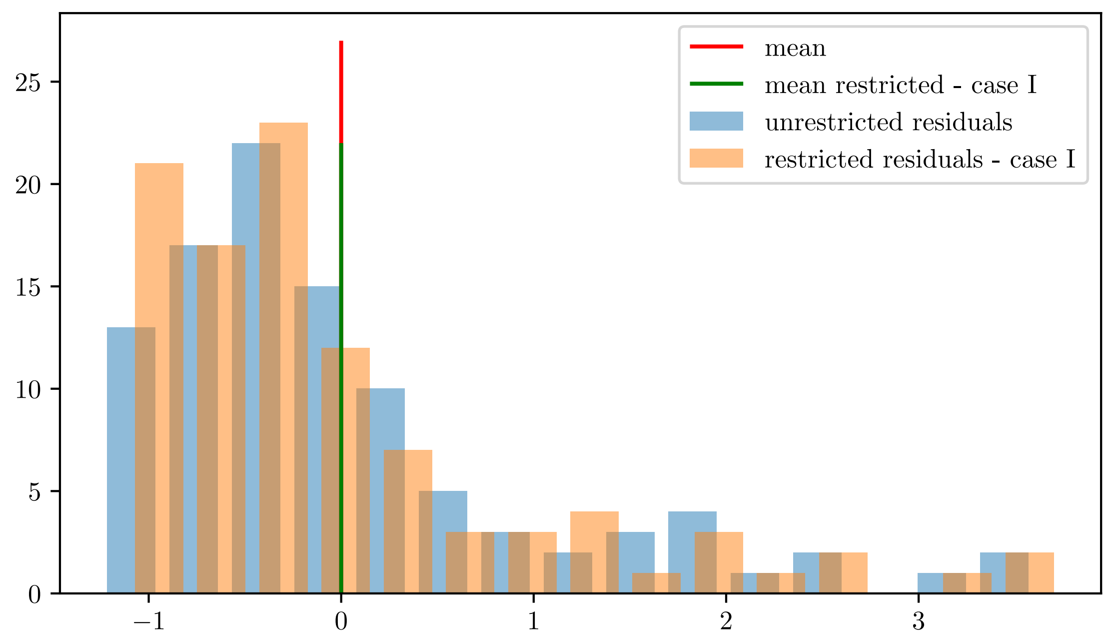

Figure 2: Residuals distribution

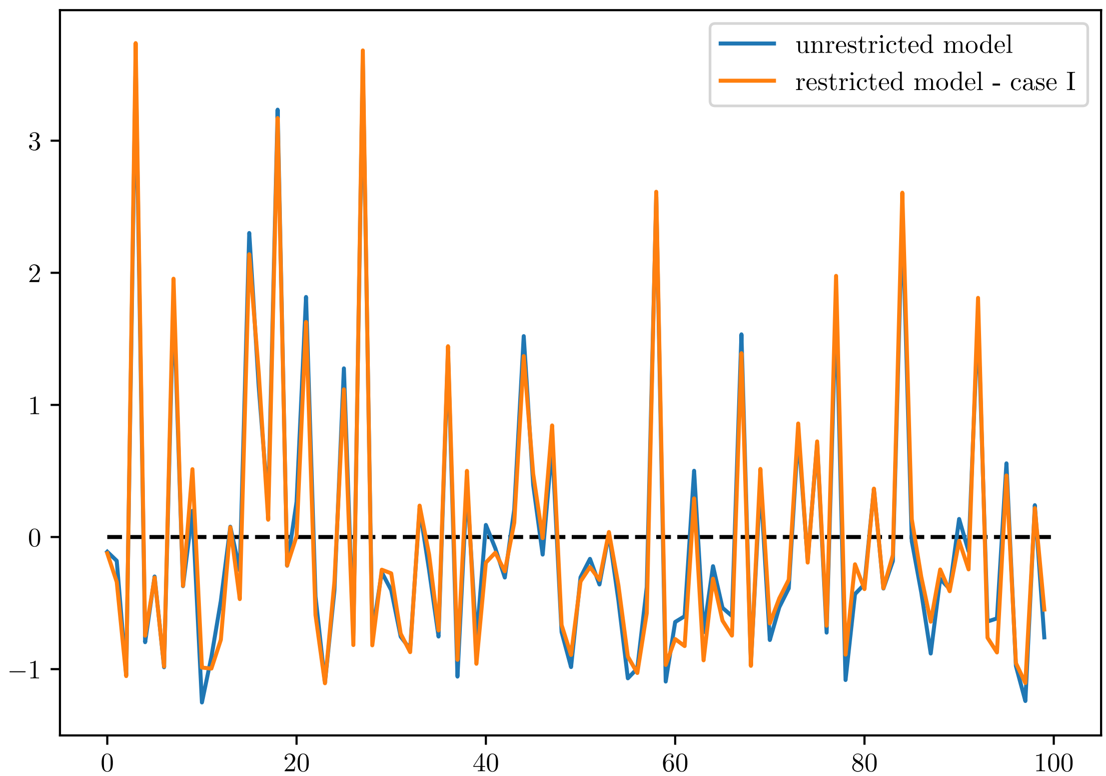

Figure 3: Residuals plot