Computing Assignment 8

TPP

March 15, 2018

Testing for the ratio of coefficients

```
##      EXACT VARIANCE BOOTSTRAP VARIANCE ASYMPTOTIC VARIANCE
## [1,]     0.1652598          0.2103906           2.827203
```

Question 1 Answer: The exact variance is the smallest, and not very different from the bootstrap one. The asymptotic variance is more than 10 times of the bootstrap one.Thus, we should expect to have a rejection rate well below 5%.

Question 2

```
##          95%
## 0.06434566
```

Answer: The quantile is very different from the 95% of the asymptotic $\chi^2_{(1)}$ distribution, 3.84. It is much smaller.

Question 3

```
## q3_reject
##         0
```

Answer: The average rejection rate is 0, which is reasonable since the data was generated with the ratio equals the true ratio 0.3/1.1 in each simulaion.

Question 4

```
## reject_boot_var_true
##             0.082775
```

Question 5

```
## reject_true  reject_est
##       0.120       0.305
```

Question 6 Answer: The size of the first bootstrap is smaller than 5%, while the size of the second bootstrap test, using the distribution approximated by the bootstrapped Wald is significantly larger. ##The right way of approximating the distribution is to center around the estimated ratio. Doing so gives us a false rejection rate of 30.5%.

Question 7 Answer: The results will change for sure, since Wald is not invariant to the $r(\theta)$ functional form. To be specific, here $r(\theta) = \beta_1 - 0.3/1.1\beta_2$, and $R' = (0.1 - 0.3/1.1)$

Cross-validation

1. (a) The dimension for *crossval_df* is 100×3. It is partitioned into 100 rows, one for each observation (n = 100). For each row, a single observation is removed from the overall data set, to become the "testing" data set, and the remaining n-1 observation are in the "training data set. Thus each cell in the "training" column is a list of 99 observations and each object in the "testing" column is one observation (list of a single element).

   (b) The columns *mse* indicates the MSE for both the in-sample and out-sample model. The column *model* indicates whether the regression contains X_2 or not. The column *value* contains the value mean squared error for either the "MSE in" or "MSE out" approach.

(c) The line $resids\_out = map2\_dbl(...)$ captures the residual from the test observation using the regression estimate from the training observations. Thus .x refers to the x values in the test observation and .y is the new y predicted using the test x values and the training regression coefficient(s).

2. (a) $MSE_{in_1} = 1.198 > MSE_{in_2} = 1.164$

(b) $MSE_{out_1} = 1.250 > MSE_{out_2} = 1.241$

Thus we can see that for both "MSE in" and "MSE out" including an additional regression variable decreases mean squared error. This is consistent with the intuition that, generally, as regression variables are added mean squared error is non-increasing.

(c) $MSE_{in_1} = 1.198 < MSE_{out_1} = 1.250$ and $MSE_{in_2} = 1.164 < MSE_{out_2} = 1.241$

"MSE out" generates a larger error variance than "MSE in" for both model 1 and 2. This is likely due to the fact that "MSE out" is a function of a single error estimate, and thus is not averaged over larger and smaller values for all "MSE in" estimates. On average, a greater number of lower error values are included in all "MSE in" estimates, causing this error estimate to be lower.

3. (a) Using the equation 3.47:

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \tilde{e}_i^{\,2} = \frac{1}{n} \sum_{i=1}^{n} (1 - h_{ii})^{-2} \hat{e}_i^{\,2}$$

We can calculate the out-of-sample MSE for both model. We get $MSE_{oos1} = 1.22$ and $MSE_{oos2} = 1.21$. These two numbers are lower than the MSE from simulation. This is because Hansen's out of sample MSE has only removed one observation, thus we are estimating MSE on 99 observations, while each of our out of sample MSE's was based on a single observation.