Computing Assignment VIII

redSloth: Tyler Hoppenfield, Daniel Mather, Iwunze Ugo

March 15, 2018

Wald Test

- Our distance metric will be $\frac{\hat{\beta}_1}{\hat{\beta}_2} - \frac{.3}{1.1}$
- using the delta method, this is asymptotically distributed as follows:

$$N(0, R' n \hat{V}_{\hat{\theta}}^{-1} R))$$

$$R' = (0, \frac{1}{\beta_2}, -\frac{\beta_1}{\beta_2^2})$$

$$\hat{\theta} = \frac{\hat{\beta}_1}{\hat{\beta}_2}$$

Simulations

Our exact variance of the ratio is:

```
##      ratio
## 0.03564622
```

Our asymptotic variance is:

```
##          v
## 0.02311099
```

our bootstrapped variance is:

```
## boot_mean
## 0.2660209
```

Based on these, we expect to reject the null hypothesis too often with the bootstrap Wald test.

Actual wald distributions with entire sample

```
##   W_test
## 1.502106
```

Boostrap Wald statistics

Bootstrap centered on the correct null hypothesis (95th percentile of bootstrap values, percent reject using the correct distribution)

```
## wtest_reject
##        0.226

## wtest2_cval
##    3.514044
```

Bootstrap centered on the sample mean null hypothesis (95th percentile of bootstrap values, percent reject using the correct distribution)

```
## wtest2_reject
##        0.0465

##      wald_q4
## 0.003348813
```

We print the null too, often with bootstrap methods,

Question 4: modified Wald test with bootstrap variance estimator

```
##   boot_var
## 0.02662385
```

Question 5

Since the 95th percentile of the bootstrap Wald is so far out, we naturally fail to reject the null.

Cross-Validation

The data frame, "crossval_df", is 100 rows by 3 columns. Each row in the column, "test" contains a randomly chosen triple (y, x1, x2) from the 100 simulated observations. It is 1 row by 3 columns. Each row in the "train" column contains the other 99 observations. The column, "id" specifies which of the 100 observations is selected. It is 99 rows by 3 columns.

The column, "model" in the data frame, "crossval" contains an indicator for the type of regression which is run on the data contained in "train". The column, "mse" similarly contains an indicator for the type of mse's that are calculated (one for the in sample regressions, and another for the out-sample predictions). The column, "value" contains the mse's for the in-sample and out-sample regressions for both of the regression models.

The column, "resids_out" gives the residuals from the out-sample regression. It takes the difference in the predicted value of y and the actual value of y. The ".x" is the independent variable from the excluded observation, which is used to predict dependent variable.

Mean squared errors

reg1.in	reg2.in	reg1.out	reg2.out
1.198	1.164	1.25	1.241