Week 3: Model Specification, Training Exercises

Coursera/Erasmus U., Econometric Methods and Applications

Anthony Nguyen

Training Exercise 3.1

Notes:

- This exercise uses the datafile `TrainExer31`and requires a computer.
- The dataset `TrainExer31` is available on the website.

Questions

(a) Use dataset `TrainExer31` to regress the change in the log of the S&P500 index on a constant and the book-to-market ration, and check the result presented in Lecture 3.1 that:

$$\text{change in log(SP500 index)} = 0.177 - 0.213 \cdot \text{Book-to-market} + e$$

(b) Now regress the S&P500 index (without any kind of transformation) on a constant and the book-to-market ratio. Consider whether the effect of the book-to-market on the index is significant in this specification.

(c) Make a plot of the residuals e from both question (a) and (b) and comment on the difference.

Answers

(a) Use dataset `TrainExer31` to regress the change in the log of the S&P500 index on a constant and the book-to-market ratio, and check the result presented in Lecture 3.1 that:

$$\text{change in log(SP500 index)} = 0.177 - 0.213 \cdot \text{Book-to-market} + e$$

We need to be careful here, as the question is asking for the *change* in the log of the S&P500 index, not just the log. In `R` we can do this with the `diff()` function, being careful to pad by one row so that our vector is the same length as the rest of the dataframe:

```
#take log of Index variable
TrainExer31 <- TrainExer31 %>% mutate(log_Index = log(Index))

#take difference (change) of log of Index, adding NA to pad first row
TrainExer31 <- TrainExer31 %>% mutate(diff_log_Index = c(NA, diff(log_Index, lag=1)))
```

Now that we've calculated the change in the log of the Index, we can run the regression with our transformed y-variable:

```
mod_difflog_Index_BookMarket <- lm(diff_log_Index~BookMarket, data = TrainExer31)

summary(mod_difflog_Index_BookMarket)
```

```
##
## Call:
## lm(formula = diff_log_Index ~ BookMarket, data = TrainExer31)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58713 -0.09314  0.02152  0.13721  0.38236
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.17695    0.04994   3.543 0.000648 ***
## BookMarket  -0.21332    0.07896  -2.702 0.008347 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1912 on 84 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.07994,    Adjusted R-squared:  0.06899
## F-statistic: 7.299 on 1 and 84 DF,  p-value: 0.008347
```

(b) Now regress the S&P500 index (without any kind of transformation) on a constant and the book-to-market ratio. Consider whether the effect of the book-to-market on the index is significant in this specification.

```
mod_Index_BookMarket <- lm(Index~BookMarket, data = TrainExer31)

summary(mod_Index_BookMarket)
```

```
##
## Call:
## lm(formula = Index ~ BookMarket, data = TrainExer31)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -694.8 -252.2 -115.2  234.8 1183.7
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1035.35      95.02  10.896   <2e-16 ***
## BookMarket  -1217.68     150.80  -8.075    4e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 366.5 on 85 degrees of freedom
## Multiple R-squared:  0.4341, Adjusted R-squared:  0.4274
## F-statistic:  65.2 on 1 and 85 DF,  p-value: 4.003e-12
```

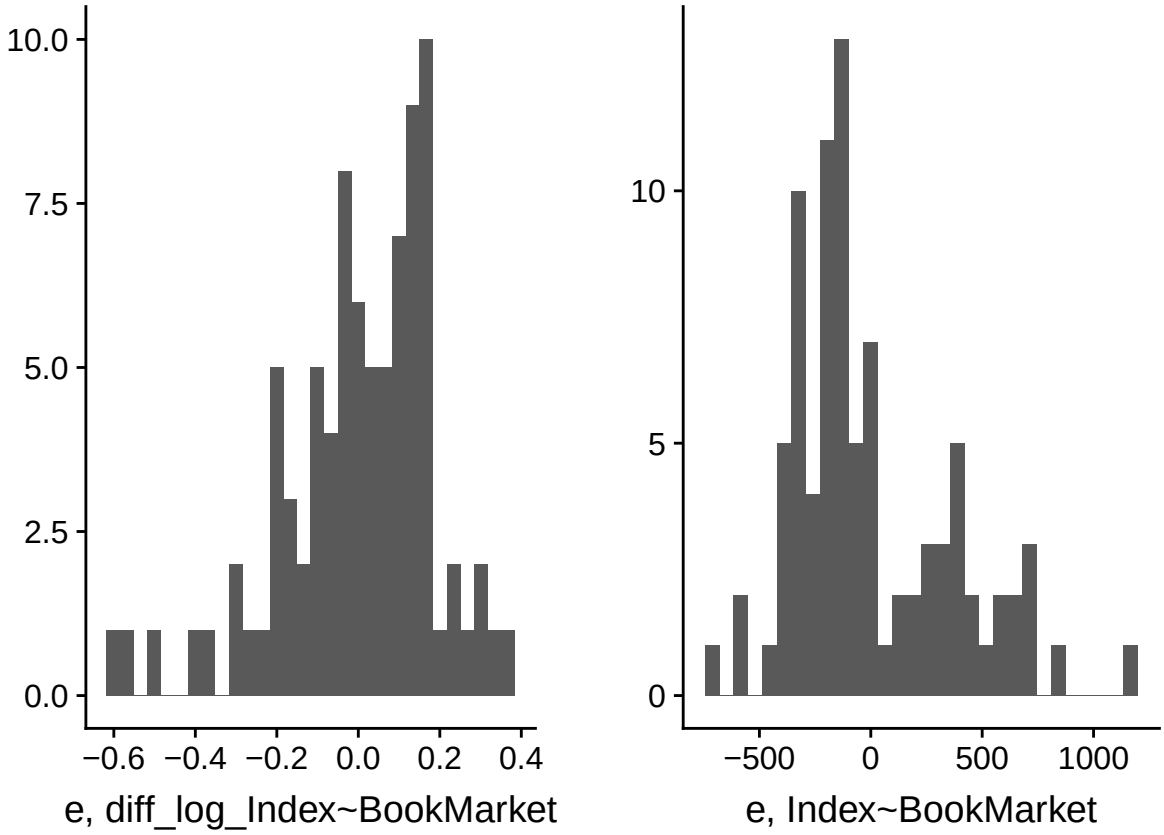
Looking at the results of this regression, the effect of BookMarket on Index has a t-value of -8.075, and p-value that is significant at the 99% level.

(c) Make a plot of the residuals *e* from both question (a) and (b) and comment on the difference.

```
res_a <- qplot(mod_difflog_Index_BookMarket$residuals, xlab = "e, diff_log_Index~BookMarket")
res_b <- qplot(mod_Index_BookMarket$residuals, xlab = "e, Index~BookMarket")

plot_grid(res_a, res_b)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



From the plot we can see that the distribution of the residuals from the two regressions have a very different distribution and use different scales.