Week 3: Model Specification, Training Exercises

Coursera/Erasmus U., Econometric Methods and Applications

Anthony Nguyen

Training Exercise 3.3

Notes:

- This exercise uses the datafile `TrainExer33`and requires a computer.
- The dataset `TrainExer33` is available on the website.

Questions

(a) In Lecture 3.1 we took the first difference of the logarithmic transformed series. These two transformations combined provide the interpretation of being an (approximate) growth rate. Show this. *Hint*: Use the definition of the first difference, $log(a) - log(b) = log(a/b), log(a/b) = log(1 + (a - b)/b)$, and that $log(1 + x) \approx x$ for x small.

(b) Use dataset `TrainExer33` to regress the change in the log of the S&P500 index on a constant, the book-to-market ration, and the square of the book-to-market ratio. Is the relationship between the index and book-to-market quadratic?

(c) Define a dummy that is 1 for 1980 and all following years. Regress the change in the log of the S&P500 index on a constant, the book-to-market ration, and an interaction between the book-to-market ration and the just-defined dummy. Is the relationship between the index and book-to-market stable over the pre- and post-1980 period?

Answers

(a) In Lecture 3.1 we took the first difference of the logarithmic transformed series. These two transformations combined provide the interpretation of being an (approximate) growth rate. Show this. *Hint*: Use the definition of the first difference, $log(a) - log(b) = log(a/b), log(a/b) = log(1 + (a - b)/b)$, and that $log(1 + x) \approx x$ for x small.

$$\begin{aligned}
\Delta(y_i) &= log(y_i) - log(y_i - 1) \\
&= log(\frac{y_i}{y_i - 1}) \\
&= log(1 + \frac{y_i - y_i - 1}{y_i - 1}) \\
&= log(1 + \frac{\Delta y_i}{y_i - 1}) \\
&\approx \frac{\Delta y_i}{y_i - 1}
\end{aligned}$$

(b) Use dataset `TrainExer33` to regress the change in the log of the S&P500 index on a constant, the book-to-market ratio, and the square of the book-to-market ratio. Is the relationship between the index and book-to-market quadratic?

Similar to what we did in Exercise 3.1, we can start by making the relevant transformations to our variables and adding them to our dataframe:

```
TrainExer33 <- TrainExer33 %>% mutate(log_Index = log(Index),
                                      diff_log_Index = c(NA, diff(log_Index, lag=1)),
                                      BookMarket2 = BookMarket^2)
```

Now we can run the regression by calling:

```
mod <- lm(diff_log_Index~BookMarket+BookMarket2, data = TrainExer33)

summary(mod)
```

```
##
## Call:
## lm(formula = diff_log_Index ~ BookMarket + BookMarket2, data = TrainExer33)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58262 -0.10271  0.03059  0.14504  0.36806
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.05633    0.08900   0.633    0.528
## BookMarket   0.23684    0.28699   0.825    0.412
## BookMarket2 -0.34668    0.21266  -1.630    0.107
##
## Residual standard error: 0.1894 on 83 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1085, Adjusted R-squared:  0.08701
## F-statistic:  5.05 on 2 and 83 DF,  p-value: 0.008517
```

Compared to the regression we ran in Exercise 3.1 (where we did not include the squared BookMarket term), all of the coefficients in this regression have lower t-scores, and none of them are statistically significant. The relationship between Index and BookMarket does not appear to be quadratic.

(c) Define a dummy that is 1 for 1980 and all following years. Regress the change in the log of the S&P500 index on a constant, the book-to-market ratio, and an interaction between the book-to-market ratio and the just-defined dummy. Is the relationship between the index and book-to-market stable over the pre- and post-1980 period?

Here, we first define the dummy variable for 1980 as indicated:

```
TrainExer33 <- TrainExer33 %>% mutate(Year_1980 = ifelse(Year>=1980, 1, 0),
                                      BookYear1980 = BookMarket*Year_1980)
```

Now, we can regress the change in log_Index with BookMarket and the interaction between BookMarket and the dummy for Year_1980:

```
mod2 <- lm(diff_log_Index~BookMarket + BookYear1980, data = TrainExer33)

summary(mod2)
```

```
##
## Call:
## lm(formula = diff_log_Index ~ BookMarket + BookYear1980, data = TrainExer33)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59549 -0.09750  0.01395  0.13388  0.38866
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.16614    0.05370   3.094  0.00269 **
## BookMarket   -0.20787    0.07987  -2.603  0.01096 *
## BookYear1980  0.04859    0.08630   0.563  0.57494
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.192 on 83 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.08344,    Adjusted R-squared:  0.06136
## F-statistic: 3.778 on 2 and 83 DF,  p-value: 0.02689
```

From our regression summary, we can see that the coefficient for our interaction term is 0.049, with a t-score of 0.563, indicating it is not statistically significant. This would imply that there is not significant change between the index and the book-to-market depending on whether the year is before or after 1980, meaning that the relationship is stable over time.