HW2 - But We Make It Up in Volume

Emulie Chhor

30/12/2022

```r
uval <- read.csv("~/Documents/OneDrive/Bac-Maths-Info/Winter2022/CMU_36_402_UndergraduateAdvancedDataAna
kable(names(uval))
```

x
country
year
gdp
growth
underval

```r
kable(summary(uval))
```

	country	year	gdp	growth	underval
	Length:1301	Min. :1955	Min. : 285.1	Min. :-0.283416	Min. :-1.9333614
	Class :character	1st Qu.:1970	1st Qu.: 1624.4	1st Qu.: 0.002264	1st Qu.:-0.2281424
	Mode :character	Median :1985	Median : 4336.6	Median : 0.018976	Median :-0.0007638
	NA	Mean :1982	Mean : 7404.6	Mean : 0.017246	Mean :-0.0009684
	NA	3rd Qu.:1995	3rd Qu.:10488.5	3rd Qu.: 0.033947	3rd Qu.: 0.2393331
	NA	Max. :2000	Max. :62203.7	Max. : 0.256314	Max. : 1.3580288

Question 1 -

```r
model1 <- lm(growth~underval + log(gdp), data = uval)
kable(summary(model1)$coefficients)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0352453	0.0066496	-5.300375	0.0000001
underval	0.0047639	0.0021791	2.186141	0.0289834
log(gdp)	0.0062971	0.0007905	7.965909	0.0000000

Q1a)

We see that the coefficient for log(gdp) is 0.00629 with p-value of 0, which means that the log(gdp) is statistically significant. Since we say that for every increase in log(gdp), we expect the country to grow by a factor of 0.00629, the coefficient doesn't support the idea of "catching-up"

Q1b)

The coefficient for underval is 0.0047 with p-value of $0.02 < \alpha = 0.05$, which means that underval is statistically significant. We say that for every increase of the index of under-valuation, we expect the country to grow by 0.0047%, which means that the data does support the under-valuing idea.

Question 2 -

```r
model2 <- lm(growth~underval + log(gdp) + country + factor(year),
             data = uval)

# kable(summary(model2)$coefficients)
```

Q2a)

```r
kable(summary(model2)$coefficients[2:3, 1:2])
```

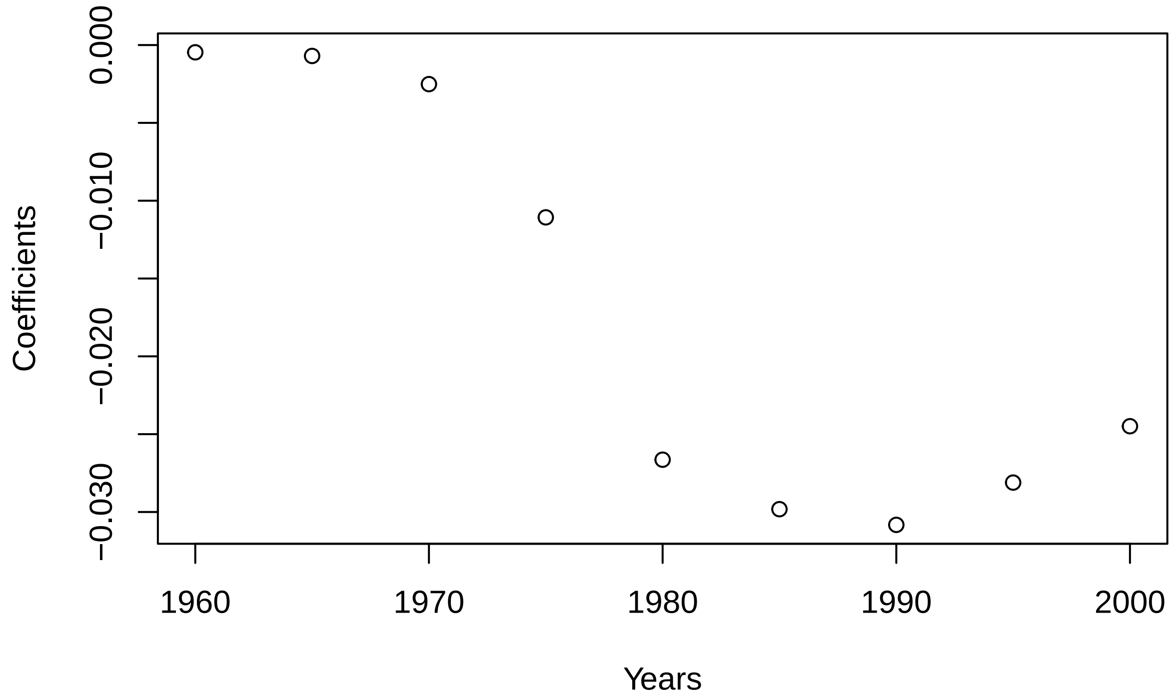
	Estimate	Std. Error
underval	0.0136094	0.0028977
log(gdp)	0.0289246	0.0031672

Q2b)

Since we only have 10 different values for year 5 years apart, we would rather consider the covariate year as a discrete value. This means that we would have a distinct slope for the 10 years value rather than for every yearly increment.

Q2c)

```r
years.coeff <- summary(model2)$coefficients[182:190, 1]
years.values <- sort(unique(uval$year))[2:10]
plot(years.values, years.coeff, xlab = "Years", ylab = "Coefficients")
```



Q2d)

The second model doesn't support the idea of catching up because, again, the log(gdp) coefficient is positive and statistically significant, which suggest that for every log(gdp) increase, the country grows by 0.0289 %. However, the model is in accord with the undervalue idea since the underval coefficient is positive and is statistically significant ($pvalue < \alpha = 0.05$)

Question 3 -

Q3a)

```
summary(model1)$r.squared
```

```
## [1] 0.04855196
```

```
summary(model1)$adj.r.squared
```

```
## [1] 0.04708594
```

```
summary(model2)$r.squared
```

```
## [1] 0.4292363
```

```
summary(model2)$adj.r.squared
```

```
## [1] 0.3321397
```

The R-squared value can be used to compare models, as it give the proportion of variance in the response variable explained by the model. Therefore, since the R-squared value (and adjusted) are both bigger in the second model, we say that the second model is the better fit.

Q3b)

Q3c)