Empirical Exercise 1, Stock and Watson Chapter 4

Econ 440 - Introduction to Econometrics

Patrick Toche, ptoche@fullerton.edu

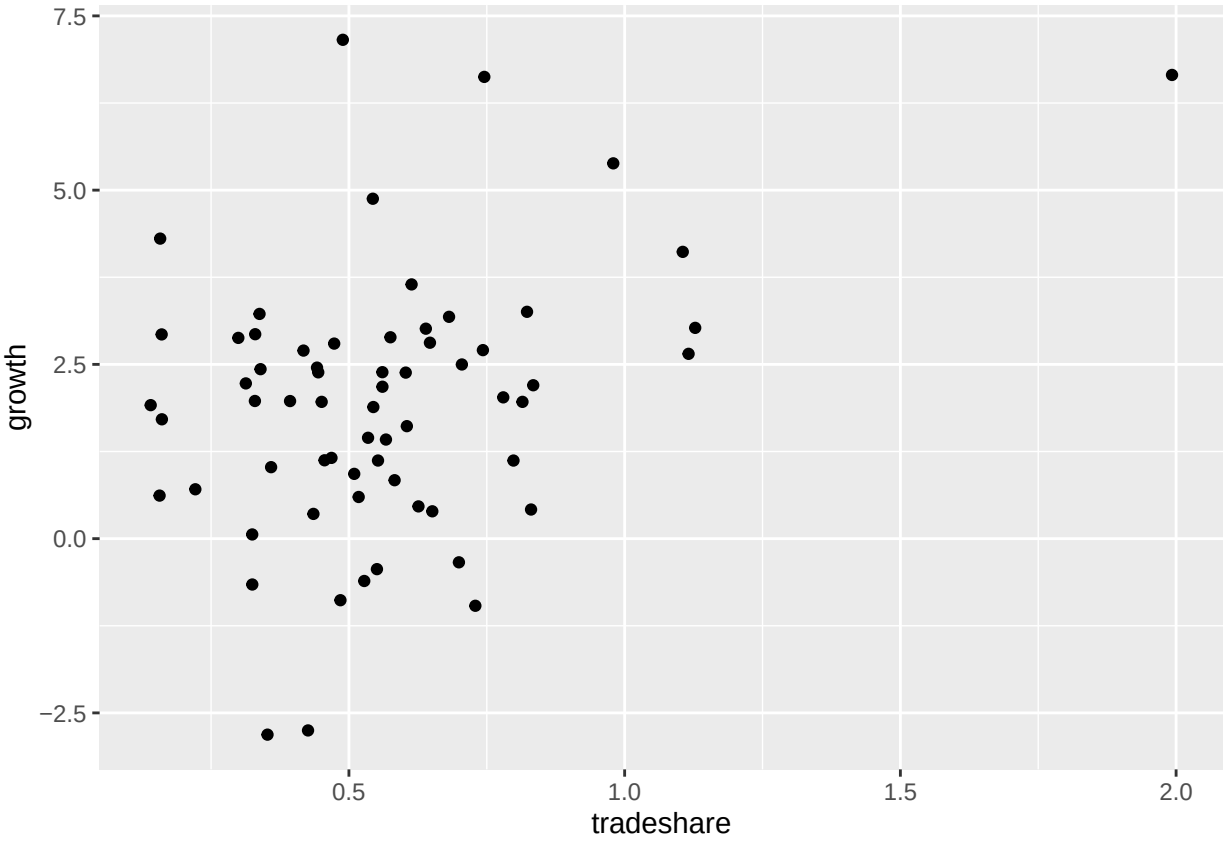
27 March 2022

In this exercise, you will investigate the relationship between growth and trade.

(a)

Construct a scatterplot of average annual growth rate (Growth) on the average trade share (TradeShare). Does there appear to be a relationship between the variables?

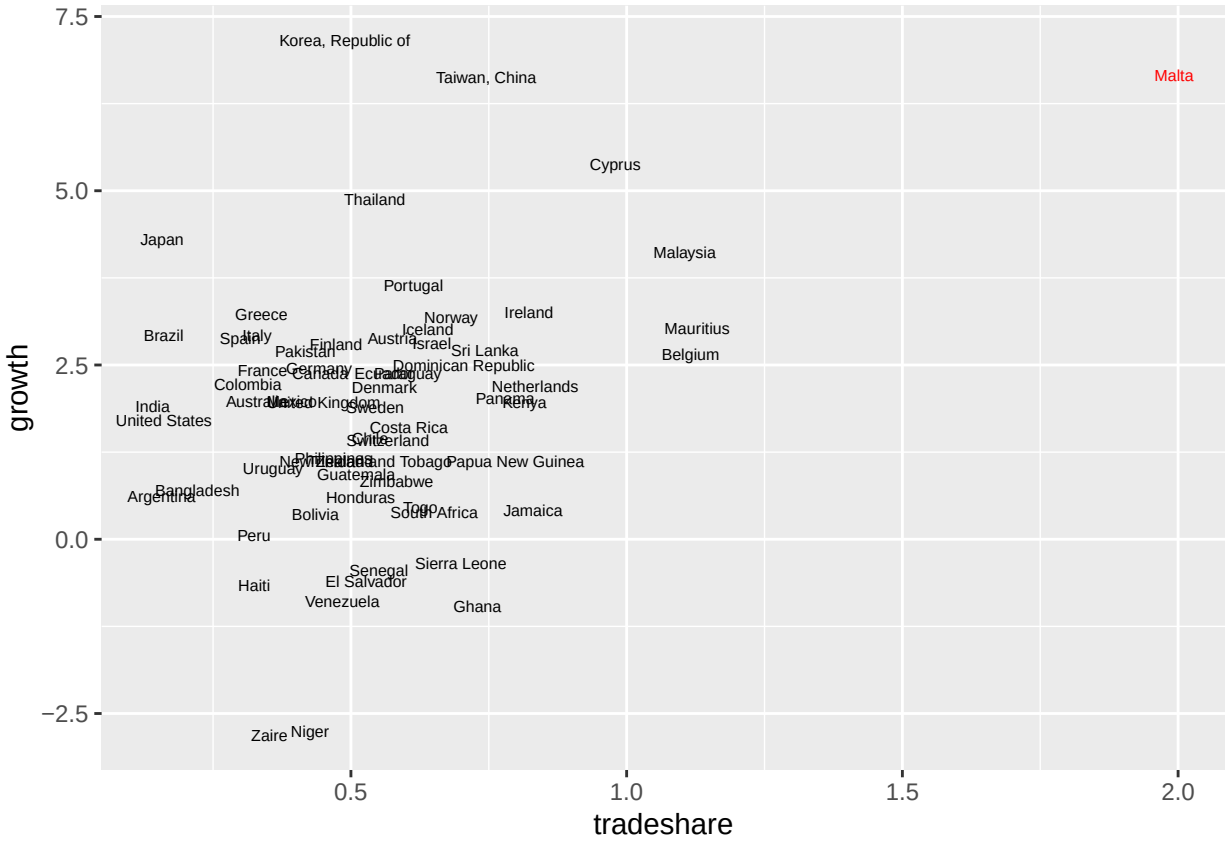
```
library(ggplot2)
df$country <- as.factor(df$country_name)
ggplot(data=df, aes(x=tradeshare, y=growth)) + geom_point()
```



(b)

One country, Malta, has a trade share much larger than the other countries. Find Malta on the scatterplot. Does Malta look like an outlier?

```
ggplot(data=df, aes(x=tradeshare, y=growth, label=country)) +
    geom_text(size=2, aes(colour = I(ifelse(country == "Malta", "red", "black"))))
```



(c)

Using all observations, run a regression of Growth on TradeShare. What is the estimated slope? What is the estimated intercept? Use the regression to predict the growth rate for a country with a trade share of 0.5 and for another with a trade share equal to 1.0.

```
ols <- lm(growth ~ tradeshare, data=df)
summary(ols)
```

```
##
## Call:
## lm(formula = growth ~ tradeshare, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.374 -0.886  0.233  0.925  5.389
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.640      0.490    1.31   0.1961
## tradeshare     2.306      0.773    2.98   0.0041 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.79 on 63 degrees of freedom
```

```
## Multiple R-squared:  0.124,  Adjusted R-squared:  0.11
## F-statistic: 8.89 on 1 and 63 DF,  p-value: 0.00407
```

(d)

Estimate the same regression, excluding the data from Malta. Answer the same questions in (c).

```
# get regression coefficients
df2 <- subset(df, country != "Malta")
ols2 <- lm(growth ~ tradeshare, data=df2)
b0 <- coef(ols2)[1]
b1 <- coef(ols2)[2]
# get Malta trade share
x_obs <- df[df$country == "Malta", "tradeshare"]
y_hat <- b0 + b1 * x_obs
y_hat
```

```
##   tradeshare
## 1     4.3068
```

```
# or use the built-in predict:
predict(ols2, newdata=x_obs)
```

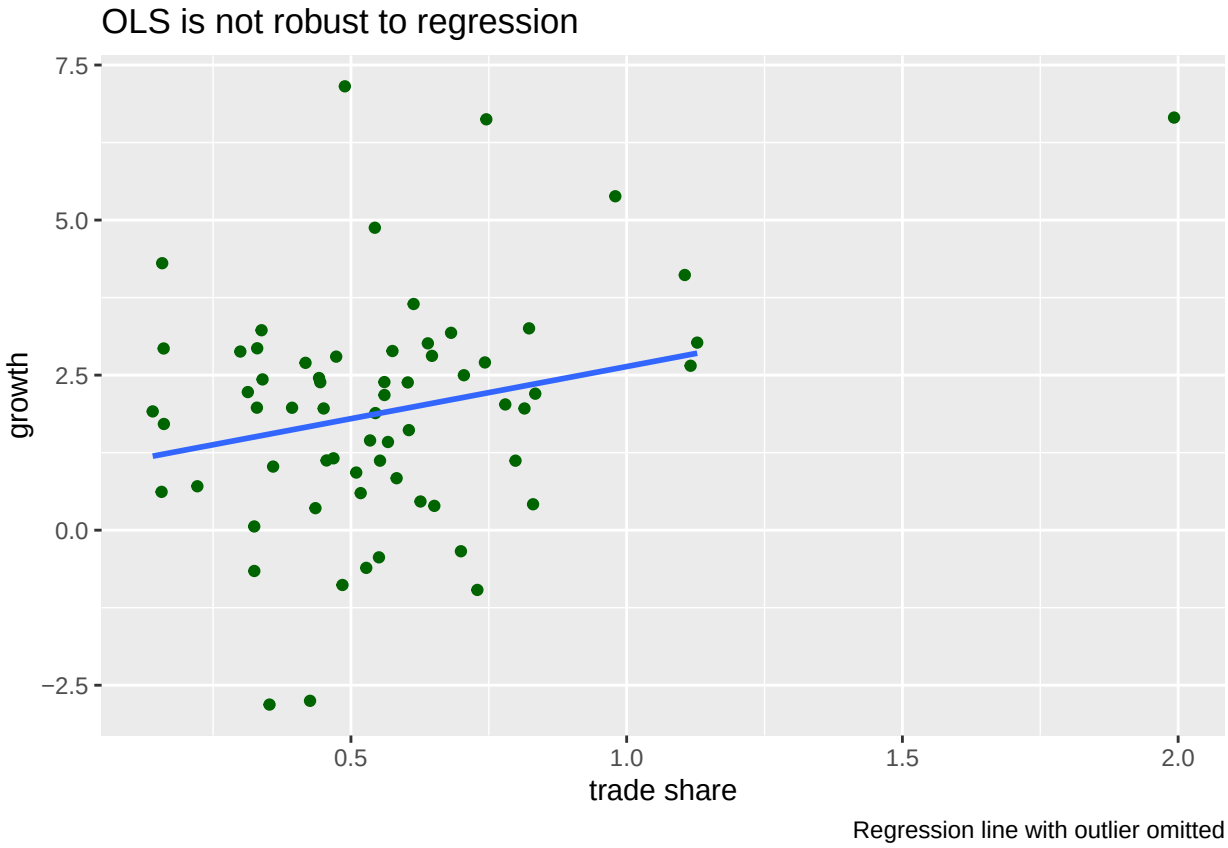
```
##      1
## 4.3068
```

(e)

Plot the estimated regression functions from (c) and (d). Using the scatterplot in (a), explain why the regression function that includes Malta is steeper than the regression function that excludes Malta.

```
ggplot(data=df, aes(x=tradeshare, y=growth)) +
    geom_point(col="darkgreen") +
    geom_smooth(method = "lm", se=FALSE, data=df2) +
    labs(title = 'OLS is not robust to regression',
         caption="Regression line with outlier omitted",
         x="trade share")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



(f)

Where is Malta? Why is the Malta trade share so large? Should Malta be included or excluded from the analysis?

Malta is a Mediterranean island off the coast of Sicily and not far from Tunisia. It is by necessity a very open economy. In a larger dataset with other islands, it would make sense to keep it, but here it makes it harder to make reliable predictions for economies that are much less open.