Econometrics 1 - Case Study 1

2022-10-20

0 Import data

```
data = read.csv("ELCONS_GDP.csv")
head(data)
```

```
##   COUNTRY TOTALCONS      GDP
## 1     AUS    179852 537613.6
## 2     AUT     52553 235450.2
## 3     BEL     79166 284950.5
## 4     CAN    503403 901029.6
## 5     CHL     37141 146574.2
## 6     CZE     52292 166568.3
```

1 Data Analysis

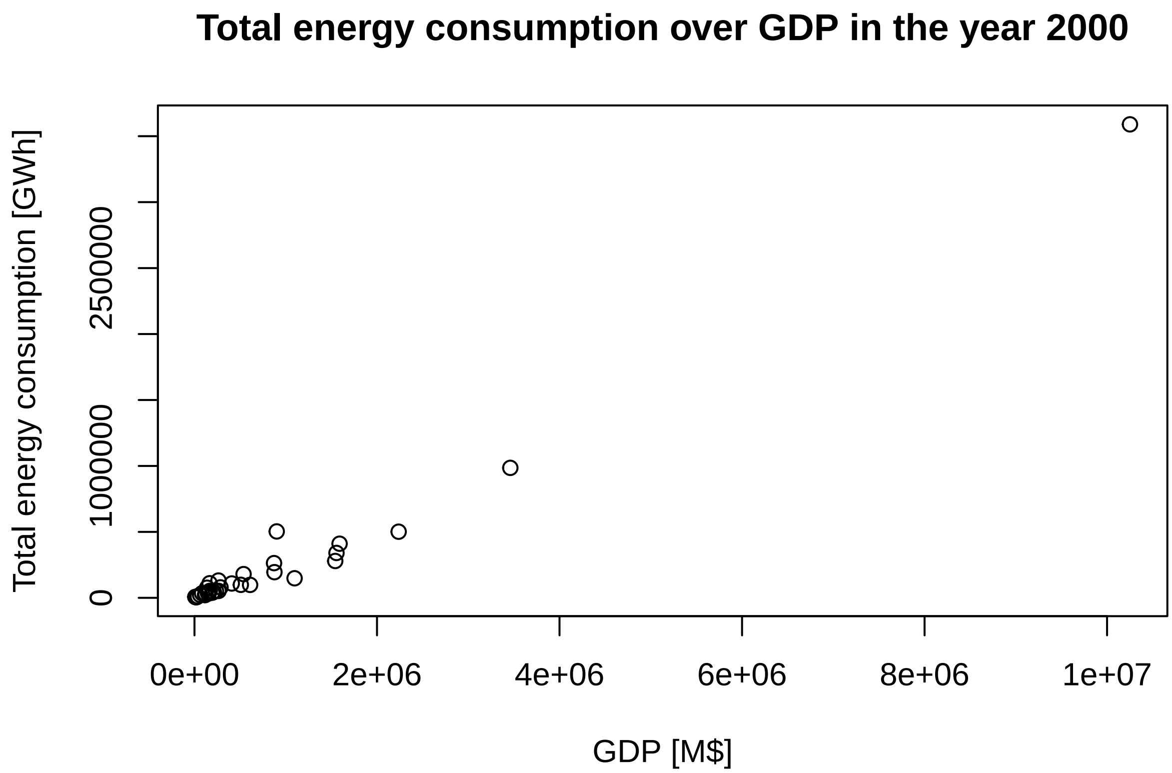
```
summary(data)
```

```
##    COUNTRY            TOTALCONS            GDP
##  Length:35          Min.   :   4484   Min.   :    8378
##  Class :character   1st Qu.:  33843   1st Qu.:  129894
##  Mode  :character   Median :  76468   Median :  235450
##                     Mean   : 245447   Mean   :  837337
##                     3rd Qu.: 187280   3rd Qu.:  873755
##                     Max.   :3589779   Max.   :10250952
```

The mean, median, minimum and maximum of the total energy consumption ($TOTALCONS$) and the GDP can be found in the cell's output above.

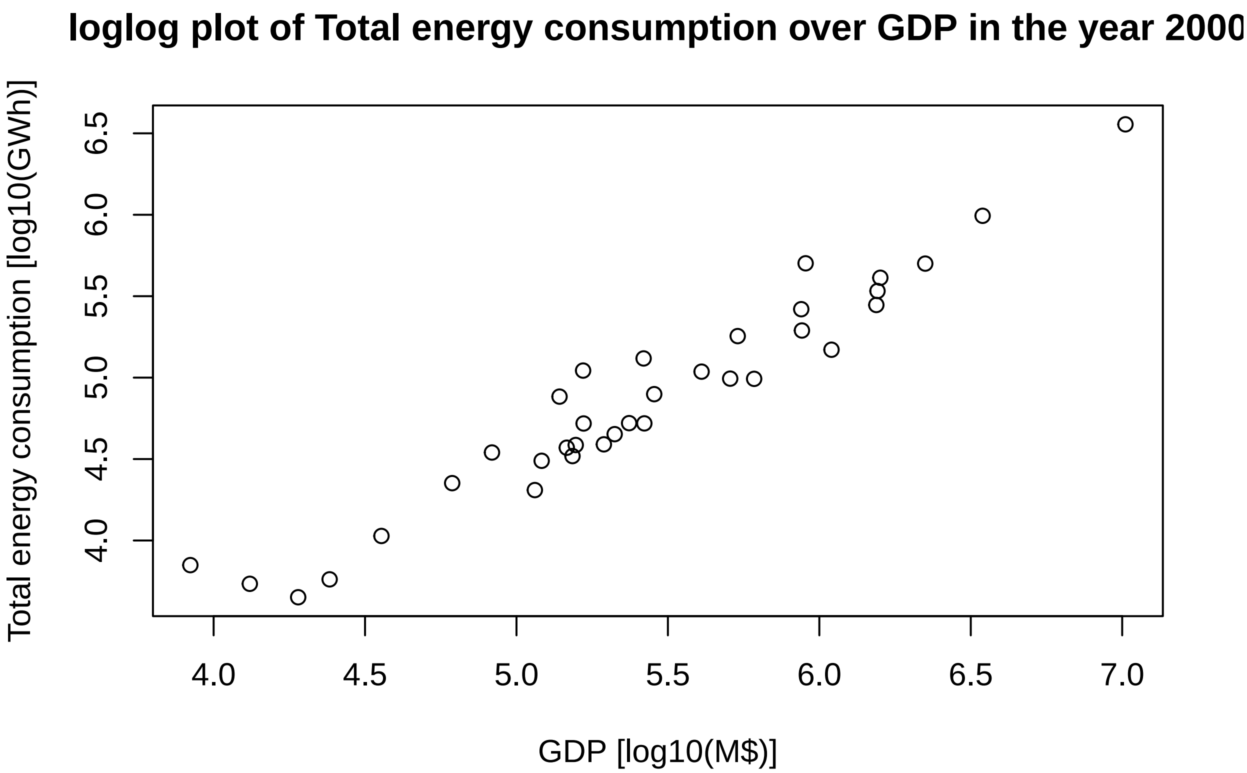
```
#define plot labels
xlabel="GDP [M$]"
ylabel="Total energy consumption [GWh]"
xloglabel="GDP [log10(M$)]"
yloglabel="Total energy consumption [log10(GWh)]"
```

```
plot(data$GDP, data$TOTALCONS,
     main="Total energy consumption over GDP in the year 2000",
     xlab=xlabel,
     ylab=ylabel)
```



Not much is visible, as this is the wrong way to plot multiplicative growth processes. One can see an extreme outlier on the top right. With this visualization no certain judgement can be made about the relationship of variables.

```
plot(log10(data$GDP), log10(data$TOTALCONS),
     main="loglog plot of Total energy consumption over GDP in the year 2000",
     xlab=xloglabel,
     ylab=yloglabel)
```



This is the log-log-plot, where the logarithm to the basis 10 is plotted for both variables/axes.

The outlier is still there of course, but lies on the imaginary eyeballed regression line. Because a clear line can be drawn by hand which incorporates most data points, a linear relationship with a high correlation and dependence can be assumed. We will investigate this assumption now by calculating it.

1.2 Simple linear regression

A simple linear regression with a vanishing error term ($E(u) = 0$) is easily possible with built-in commands:

```
ols = lm(
         TOTALCONS ~ GDP,
         data = data
)
summary(ols)
```

```
##
## Call:
## lm(formula = TOTALCONS ~ GDP, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -217087  -17446   20405   37307  236430
##
## Coefficients:
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.754e+04  1.791e+04  -2.096   0.0438 *
## GDP          3.380e-01  9.134e-03  37.001   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 95810 on 33 degrees of freedom
## Multiple R-squared:  0.9765, Adjusted R-squared:  0.9758
## F-statistic:  1369 on 1 and 33 DF,  p-value: < 2.2e-16
```

```
ols$coefficients
```

```
##   (Intercept)           GDP
## -3.754232e+04  3.379637e-01
```

The OLS estimates for $\hat{\beta}_0$ is $-3754\ GWh$ and for $\hat{\beta}_1$ is 0.3380 $\frac{GWh}{M\$} = 3380\frac{GWh}{G\$} = 3380\frac{Wh}{\$}$. $\hat{\beta}_1$ is the coefficient associated with an increase of GDP. This means that, in the model, an increase of 1 $M\$$ results in an increase of 0.338 GWh.

```
summary(ols)
```

```
##
## Call:
## lm(formula = TOTALCONS ~ GDP, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -217087  -17446   20405   37307  236430
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.754e+04  1.791e+04  -2.096   0.0438 *
## GDP          3.380e-01  9.134e-03  37.001   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 95810 on 33 degrees of freedom
## Multiple R-squared:  0.9765, Adjusted R-squared:  0.9758
## F-statistic:  1369 on 1 and 33 DF,  p-value: < 2.2e-16
```

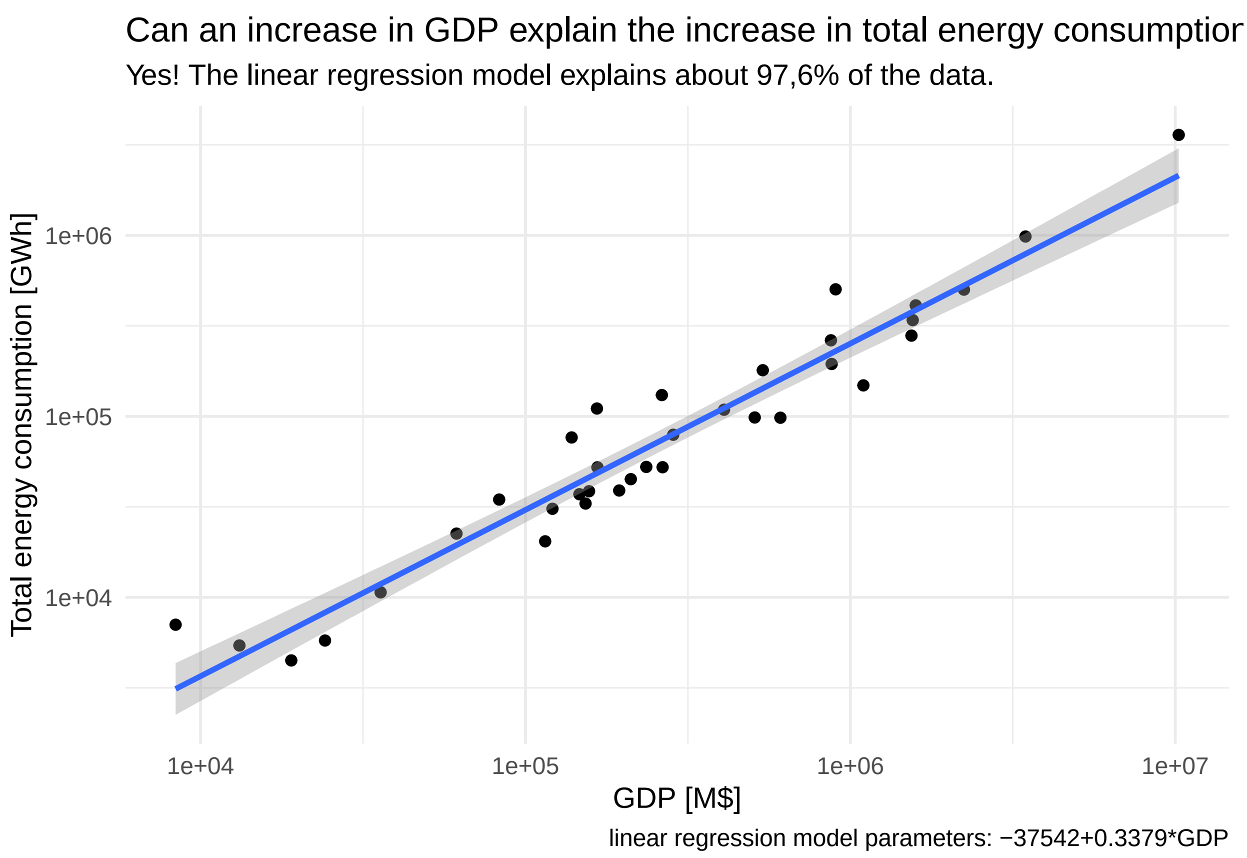
1.3 Plot

```
#install.packages("dplyr")
library(dplyr)
#install.packages("ggplot2")
library(ggplot2)
```

```
data %>% ggplot(mapping = aes(x = GDP, y = TOTALCONS)) +
  theme_minimal() +
  geom_point() +
```

```r
  #geom_smooth(method = "lm", se = FALSE) + #no standard error plotted
  geom_smooth(method = "lm") +
  labs(x=xlabel, y=ylabel,
       title="Can an increase in GDP explain the increase in total energy consumption?",
       subtitles="Yes! The linear regression model explains about 97,6% of the data.",
       caption="linear regression model parameters: -37542+0.3379*GDP") +
  scale_x_continuous(trans='log10') +
  scale_y_continuous(trans='log10')
```

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



We can see a very strong linear relationship between the GDP and total energy consumption, which is supported by the value of $R^2 = 97.6\%$ in the model's output.

1.4 Residual plot of Model 1

1.5 Log-log-linear Model

1.6 Plot of Model 2 (the log-log-linear model)

1.7 Numerical implications

a

b

c

1.8 Model generalization discussion

1.9 Bonusquestion

2 Theory