Problem Set 1: Solutions

1) Create the Dataset

Build a dataset containing per capita CO2 emissions and per capita real GDP for Japan from 1960-2010. Obtain the data:

- Japanese Real GDP per capita: from FRED (Federal Reserve Economic Data). Go to `http://research.stlouisfed.org/fred2/` and download the time series of "Real GDP per Capita in Japan".
- Japanese CO2 emissions per capita: from World Bank. Go to `http://data.worldbank.org` and download "CO2 emissions (metric tonnes per capita)".

Organise the data:

- Combine the series of real GDP per capita and CO2 emissions per capita for Japan into one Excel file. The format should be organised such that the first column indicates the years, the second column lists real GDP and the third column lists CO2 emissions per capita (see Table 1).

Table 1: Data

Year	**RGDP_pc**	**CO2_pc**
1960	6109	2.516538
1961	6772	2.981979
1962	7285	3.059736
1963	7846	3.359321
1964	8633	3.673035
1965	9022	3.912906
.	.	.
.	.	.
.	.	.
.	.	.

- Generate a PcGive .in7 database by importing (or copy pasting the data into PcGive). Make sure the starting date of the database matches the data you have obtained. Choose sensible variable names, e.g. rgdp_pc for real GDP per capita for Japan, and co2_pc for CO2 emissions per capita.

Solution:
The constructed dataset "japan_kuznets.in7" can be downloaded from: `www.felixpretis.org/teaching`

2) Summary Statistics and Overview of the Data

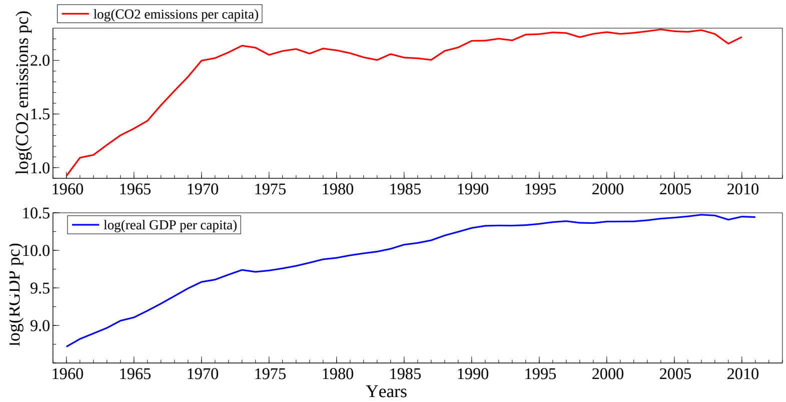
1. Using the *Algebra* editor construct the variables:
   - log(RGDP per Capita)
   - log(CO2 Emissions per Capita)
2. Why are log transformations of the data useful?
3. Plot log(RGDP per Capita) and log(CO2 Emissions per Capita) over time in separate plots
4. Plot log(CO2 Emissions per Capita) against log(RGDP per Capita) using a scatter plot
5. Given the time series and scatter plots of the data, what can you say about possible relationships between the series?
6. Provide a short table of summary statistics: using the *Model-Descriptive Statistics* menu, report the mean and standard deviations of all data series (log and non-log transformed) together with their unit of measurement.

Solution:

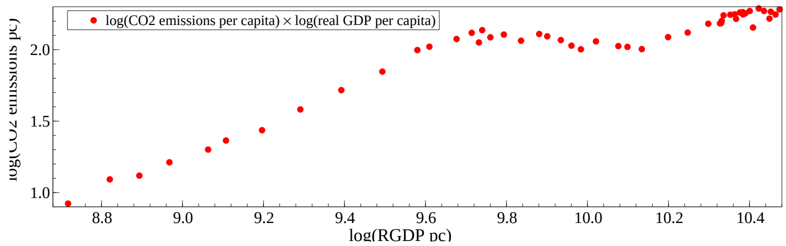
1. The variables can be constructed through the *Algebra* menu, alternatively the *Algebra* commands are:
   - `Lco2_pc = log(co2_pc);`
   - `Lrgdp_pc = log(rgdp_pc);`
2. Transforming the variables into logs allows us to interpret the relationships between the variables as approximately proportionally. Recall the result that $\frac{dlog(x)}{dx} = \frac{1}{x}$, the changes in logs are relative to their level. Re-writing this result we have that for small changes in x it holds that $100\Delta log(x) \approx \%\Delta x$. In the equation:

$log(y)_t = \beta_1 \log(x)_t$, the parameter β_1 denotes the approximate percentage change in y_t given a 1% change in x_t. In other words, the parameter β_1 is the elasticity of y_t with respect to x_t. Often econometric relationships are better approximated by proportional (log) models, rather than models in levels. For investigating the environmental Kuznets curve, transforming both variables using logs thus allows us to interpret percentage changes in CO2 emissions for given percentage changes in real per capita GDP.

3. Time Series Plots of log(CO2 Emissions per Capita) and log(RGDP per Capita)



4. Scatter Plot of log(CO2 Emissions per Capita) against log(RGDP per Capita)



5. Based on the time series plot both series are slightly increasing over time with the highest rate of increase between the 1960s and 1970s. The scatter plot suggests a potential positive relationship between CO2 emissions per capita and real GDP per capita. The relationship appears close to linear for low values of GDP and flattens out for higher values of GDP. The curved pattern matches the theoretical prediction of the environmental Kuznets curve up to the turning point.

6. Summary statistics are provided in Table 2:

Table 2: Summary Statistics of the Data

Data	**Mean**	**Standard Dev.**	**Units**
CO2 Em. pc.	7.74	2.05	tons
RGDP pc.	23082	9342	USD (2011)
log(CO2 Em. pc.)	1.99	0.35	
log(RGDP pc.)	9.9	0.5	

---

3) Initial Econometric Investigation

The theory suggests a link between emissions of CO2 per capita and real income (GDP) per capita. To investigate this, use the *Model-Single Equation Time Series* menu to estimate the following regression model and report the results:

$$\log(\text{CO2})_t = \beta_1 + \beta_2 \log(\text{RGDP})_t + \epsilon_t \tag{1}$$

1. Report the regression output in equation format showing estimated coefficients and standard errors.
2. Plot the fitted values and residuals and comment on the results
3. What values do the estimates $\hat{\beta}_1, \hat{\beta}_2$ take and what does this suggest?
4. How much variation of log(CO2 Emissions per capita) is explained by variation in log(real GDP per capita)?

Solution:

1. The estimated regression model is reported by PCGive as:

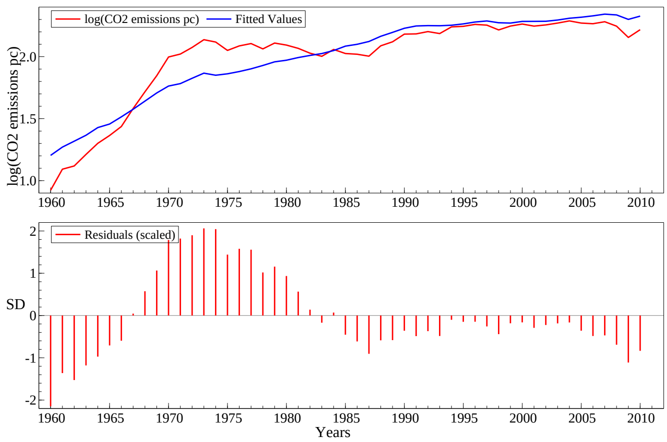
```
                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant             -4.44848     0.3652    -12.2  0.0000   0.7518
Lrgdp_pc             0.648412    0.03669     17.7  0.0000   0.8644

sigma                0.130945  RSS              0.840182195
R^2                  0.864377  F(1,49) =      312.3 [0.000]**
Adj.R^2               0.86161  log-likelihood       32.3362
no. of observations        51  no. of parameters          2
mean(Lco2_pc)          1.9969  se(Lco2_pc)         0.351994
When the log-likelihood constant is NOT included:
AIC                  -4.02753  SC                  -3.95177
HQ                   -3.99858  FPE                0.0178190
When the log-likelihood constant is included:
AIC                  -1.18965  SC                  -1.11390
HQ                   -1.16070  FPE                 0.304339
```

However, the results should be presented more neatly, for example in equation format with standard errors given in parentheses:

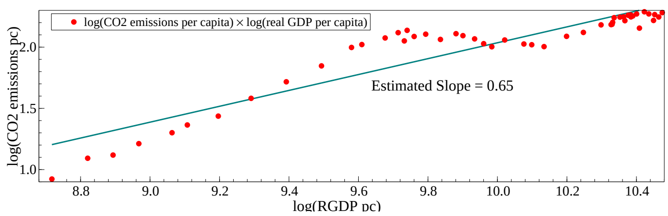
$$\log(\text{CO2}) = \underset{(0.37)}{-\;4.4} + \underset{(0.037)}{0.65}\;\log(\text{RGDP})_t$$

2. Fitted values and Residual Plot



The model fit shows that the overall trend in CO2 emissions per capita can be captured by real GDP per capita, however, the trend is under-estimated before 1984 and over-estimated after this point in time suggesting that the model is not a particularly good fit. The residual plot further emphasises this point, the residuals appear highly auto-correlated and are negative towards the latter half of the sample.

3. The estimates for β_1 and β_2 are given by the intercept $\hat{\beta}_1 = -4.4$ and the slope coefficient $\hat{\beta}_2 = 0.65$ respectively. If log(RGDP pc) were zero, according to the estimated model we could expect log(CO2 emissions per capita) to take a value of $\hat{\beta}_1 = -4.4$. For a one-unit increase in log(RGDP) we expect an approximate increase of 0.65 units in log(CO2 emissions). Using the properties of the log transformation we can say that the model estimate suggests an 0.65% increase in CO2 emissions per capita for a 1% increase in real GDP per capita. The fitted regression line is shown in the figure below:



4. The amount of variation in log(CO2 Emissions per capita) explained by variation in log(real GDP per capita) is measured through the R^2 of the regression. The R^2 of this model is given by 0.86, suggesting that approximately 86% of the variation in log(CO2 Emissions per capita) is explained by variation in log(real GDP per capita). This measure is only one of many measures of so-called "goodness of fit" and should not be over-interpreted.
