Problem Set 2: Solutions

1) Diagnostic Tests of the Model

Recall the model you estimated yesterday:

$$\log(\text{CO2})_t = \hat{\beta}_1 + \hat{\beta}_2 \log(\text{RGDP})_t \quad (1)$$

1. Comment on the diagnostic tests of the model - is the model well specified?

Solution:

1. The results of the diagnostic tests are reported directly by PCGive:

```
AR 1-2 test:       F(2,47)    =   104.46 [0.0000]**
ARCH 1-1 test:     F(1,49)    =   124.17 [0.0000]**
Normality test:    Chi^2(2)   =   4.8922 [0.0866]
Hetero test:       F(2,48)    =   12.014 [0.0001]**
Hetero-X test:     F(2,48)    =   12.014 [0.0001]**
RESET23 test:      F(2,47)    =   55.101 [0.0000]**
```

   The null-hypothesis for each of the above diagnostic tests is that the condition being tested for is not rejected. For example, in the case of normality, under the null-hypothesis the model residuals are normally distributed. Equally, under the null hypothesis the residuals do not exhibit auto-correlation (correlation or dependence over time). Based on the statistical tests reported we can therefore conclude that all but the normality test are rejected at the 1% level. This suggests that the model is highly mis-specified. The model exhibits residual auto-correlatio (AR test), the variance appears to be correlated over time (ARCH test), as well as not constant for different levels of the explanatory variable (Hetero tests), and the functional form is likely mis-specified (RESET test).

---

2) Expanding the Model

The environmental Kuzents curve suggests that pollution follows a U-pattern as income increases. This is not well captured by our simple linear model. Consider the following extension of the model where we include a squared term of real GDP per capita.

$$\log(\text{CO2})_t = \beta_1 + \beta_2 \log(\text{RGDP})_t + \beta_3 \log(\text{RGDP})_t^2 + \epsilon_t \quad (2)$$

1. What is now the marginal effect of log(RGDP per capita) on log(CO2 per capita)? What if β_2 is positive and β_3 is negative? (Hint: differentiate equation (2) with respect to log(RGDP)).

2. Use the *Algebra* menu to construct the variable $(\log(\text{RGDP}))^2$

3. Estimate model (2) and comment on the results. Report the estimated results in a table showing the coefficients (with standard errors in parentheses), together with the number of observations (T), R^2, and diagnostic tests (see Table 1 for an example of how results should be presented). In particular comment on the signs of $\hat{\beta}_2$, $\hat{\beta}_3$ and what this means in the context of the environmental Kuznets curve.

Table 1: Example Regression Results

Variable	**Estimated Coefficient**
Constant	0.5 (0.05)**
L(RGDP)	...
L(RGDP)^2	...
R^2	0.86
T	46
AR 1-2 Test	104.3 [0.00]**
ARCH 1-1 Test	10.5 [0.00]**
Normality Test	...
Hetero Test	...
RESET Test	...

4. Are $\log(\text{RGDP})_t$ and $\log(\text{RGDP})_t^2$ individually significant? Are they jointly significant? (hint: construct an F-test for joint-significance through the *Test* menu by selecting *Exclusion Restrictions*)

Solution:

1. We are interested in the marginal effect of log(RGDP) on log(CO2), in other words: given a one-unit change in log(RGDP), what is the expected change in log(CO2). To determine this marginal effect we differentiate equation (2) with respect to log(RGDP):
$$\frac{\partial \log(\text{CO2})_t}{\partial \log(\text{RGDP})} = \beta_2 + 2\beta_3 \log(\text{RGDP})_t \tag{3}$$
In the previous simple model the marginal effect was simply equal to β_2, constant for all levels of log(RGDP). Now the marginal effect varies with the level of log(RGDP). If β_2 is positive and large while β_3 is negative and small, then at first the marginal effect of log(RGDP) is positive but declining with the level of log(RGDP). As log(RGDP) gets large, the marginal effect becomes negative, leading to an inverse U-shape effect. The turning point from when onwards the marginal effect is negative can easily be calculated by setting the above equation equal to zero and solving for log(RGDP):
$$\frac{\partial \log(\text{CO2})_t}{\partial \log(\text{RGDP})} = \beta_2 + 2\beta_3 \log(\text{RGDP})_t = 0 \tag{4}$$
Solving for log(RGDP) to find the level of log(RGDP) from when onwards the marginal effect is negative:
$$\log(\text{RGDP})_t = -\frac{\beta_2}{2\beta_3} \tag{5}$$

2. The squared variable can be generated using the *Calculator* menu or equivalently the *Algebra* code is:

   - `Lrgdp_pc_sq = Lrgdp_pc^2;`

3. The regression model is estimated as:
$$\log(\text{CO2}) = -\underset{(4.1)}{46} + \underset{(0.86)}{9.3} \log(\text{RGDP})_t - \underset{(0.044)}{0.45} \log(\text{RGDP})_t^2$$

Or given in table format which is more common and used in publications and reports:

Table 2: Modelled Variable: log(CO2 Emissions per capita)

Variable	**Estimated Coefficient**
Constant	-46.48 (4.15)**
L(RGDP)	9.32 (0.85)**
L(RGDP)^2	-0.45 (0.044)**
R^2	0.96
T	51
AR 1-2 Test	67.647 [0.00]**
ARCH 1-1 Test	48.077 [0.00]**
Normality Test	1.3641 [0.50]
Hetero Test	5.6484 [0.002]**
RESET Test	20.552 [0.00]**

Consistent with the theory of the environmental Kuzents curve, the coefficient on log(RGDP) is positive, individually significant and large, while the coefficient on the square of log(RGDP) is negative, significant and comparatively smaller. This suggests a positive marginal effect of log(RGDP per capita) on log(CO2) which falls with high levels of log(RGDP). In particular, the estimates show an expected increase of 9.3% in per capita CO2 emissions for a 1% increase in real per capita GDP, with a decreasing marginal effect of 0.45%.

The estimated turning point level of log(RGDP) is given by: $-\frac{\hat{\beta}_2}{2\hat{\beta}_3}$ which in the present application equals: $\frac{9.32}{2 \cdot 0.45} = 10.44$. Converting the log(RGDP) into RGDP per capita this corresponds to a value of 34490 USD (2011 prices). This value was reached in Japan in 2007. If the theory is correct, we should expect a decrease in per-capita emissions from then onwards given an increase in income. However, the model appears to be highly mis-specified judging from the range of misspecification tests.

4. Using t-tests the coefficients are individually significant. Using an F-test for joint significance we find the test-statistic for excluding both log(RGDP) and the square of log(RGDP) to be 532.16 which has an F distribution with (2, 48) degrees of freedom. The null hypothesis of the two variables being jointly insignificant is rejected at the 1% level.

3) Time series Properties of the Data

Now we turn to the time series properties of the data:

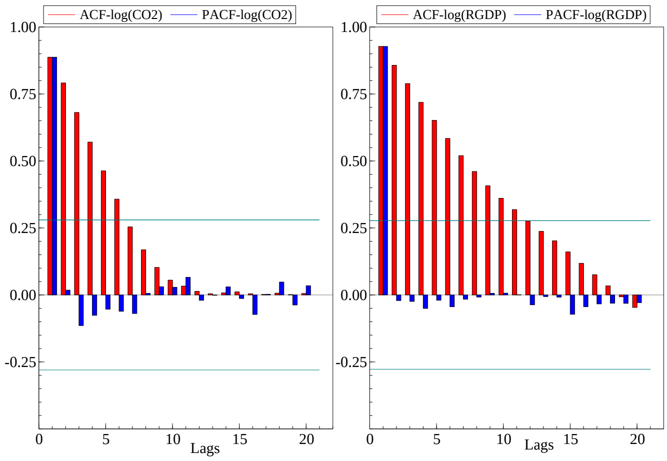
1. Plot the Auto-correlation and Partial auto-correlation functions of log(CO2) and log(RGDP) and comment on the plots. What lag length would you suggest is optimal?

2. Estimate an auto-regressive model with 1-lag (AR-1 models) for log(CO2) report the results in equation format and comment on the output.

$$\log(\text{CO2})_t = \alpha_0 + \alpha_1 \log(\text{CO2})_{t-1} + \epsilon_t \tag{6}$$

Solution:

1. Auto-correlation and partial auto-correlation functions



The auto-correlation for both time series decreases with the lag length, consistent with an auto-regressive process for both time series. The partial auto-correlation function for both series shows that, when controlling for all possible lags up to 20, only the first lag is significant. This suggests that an auto-regressive model of order one including one lag, (an AR(1) model), is appropriate. Failure to account for the apparent auto-correlation likely results in residual auto-correlation which leads to a violation of the assumption of independence of the error terms.

2. The estimated AR(1) model for log(CO2 per capita) is:

$$\log(\text{CO2}) = \underset{(0.018)}{0.895} \log(\text{CO2})_{t-1} + \underset{(0.036)}{0.234}$$

The first lag of the dependent variable, the auto-regressive variable is highly significant in the regression model. This suggests that per capita CO2 emissions are relatively persistent, emissions this year are highly dependent on emissions of the previous year. The model satisfies most diagnostic tests at the 1% level of significance. The model residuals show little to no pattern confirming this result. The model long-run equilibrium $\mu = \frac{\alpha_0}{1-\alpha_1}$ is estimated to be $\frac{0.234}{1-0.895} = 2.23$ log(CO2 emmissions per capita), which corresponds to 9.29kg CO2 per capita when inverting the log transformation.