

Econometrics: Computer-based Modelling

Problem Set

Felix Pretis*

Department of Economics, University of Oxford

Overview

Questions 1 and 2 require the use of the datasets: "income_consumption.xlsx" and "Loan Data.xlsx" which can be downloaded from: <http://www.felixpretis.org/teaching/university-of-oxford-econometrics-undergraduate/>. Question 3 requires you to construct your own dataset according to the instructions provided.

Question 1: Cross-Section Analysis

1) Create the Dataset

The file "income_consumption.xlsx" contains 4 pairs of income and consumption variables, where "consumption_A" is related to "income_A", "consumption_B" to "income_B", "consumption_C" to "income_C", and "consumption_D" to "income_D". Create a new dataset in OxMetrics using the Excel file and store the dataset as an ".in7" file.

2) Modelling

1. Investigate the relationships between consumption and income in for each pair: how is consumption A related to income A? (repeat this analysis for B, C, D).
2. How do regression models of consumption on income compare for A, B, C, D? What is the expected effect of one unit increase in income on consumption for A, B, C, D?
3. Present your model results in a table and graphs, and comment on your results.

*Contact: felix.pretis@nuffield.ox.ac.uk. Lecture slides and data available at: www.felixpretis.org

Question 2: Binary Dependent Variable & Micro-Econometrics

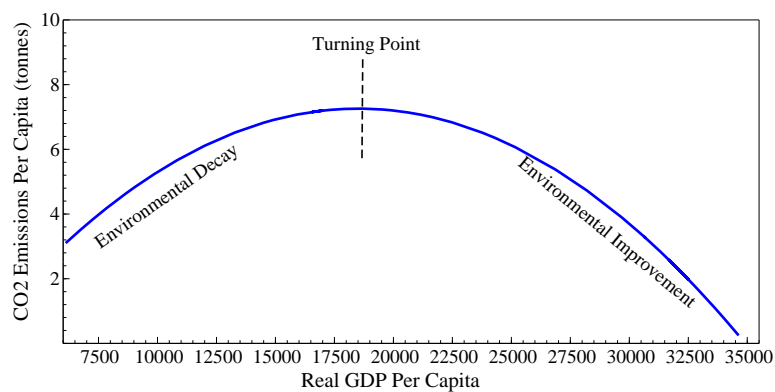
Use the dataset “Loan Data.xlsx” to determine if there is discrimination in the mortgage loan market (dataset from Wooldridge ‘Introductory Econometrics’). Make sure the dataset is *loaded correctly* into OxMetrics.

1. Estimate a *probit* model of the approval of mortgages (‘*approv*’) on a dummy variable whether the individual applying for the loan was white (‘*white*’). Find the estimated probabilities of loan approval for both whites and non-whites. How do the results compare to linear probability estimates?
2. Now expand the model by adding: housing expenditure as proportion of total income (‘*hrat*’), other obligations (‘*obrat*’), the ratio of the loan to the house price (‘*loanprc*’), the unemployment rate by industry (‘*unem*’), whether the individual was male (‘*male*’), ‘*married*’, the number of dependents (‘*dep*’), if the individual had more than 12 years of schooling (‘*sch*’), if there is a co-signer (‘*cosgin*’), if the account has been delinquent for more than 60 days (‘*chist*’), if the individual has filed for bankruptcy (‘*pubrec*’), if there were one or two (‘*mortlat1*’) or more than two (‘*mortlat2*’) late payments in the past. Is there statistical evidence for discrimination against non-whites? Is there statistical evidence for discrimination against men?
3. Conduct a Likelihood ratio test for excluding late payments from the model.

Question 3: Time-Series & Macro-Econometrics

The hypothesis of the environmental Kuznets curve suggests that, as economies grow, pollution and environmental degradation increases with income, however, beyond a certain level of income (a turning point) economic growth leads to a reduction in pollution (Stern 2004). This implies that pollution and environmental degradation are an approximate inverted U-shape function of income. The idea behind the theory is that initial growth is often driven by pollutant-intense industries, then, as economies become more developed, production shifts away from polluting industries. Equally, as incomes rise, people may be able to afford to care more about the environment.

Investigate this hypothesis in a simple setting by assessing the link between per-capita CO₂ emissions and per-capita real GDP for the UK using time series spanning the last 50 years.



1) Create the Dataset

Build a dataset containing per capita CO₂ emissions and per capita real GDP for the UK from 1960-2011. Obtain the data:

- UK Real GDP per capita: from FRED (Federal Reserve Economic Data). Go to <http://research.stlouisfed.org/fred2/> and download the time series of "Constant GDP per capita for the United Kingdom".
- UK 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 the UK 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	13478.9	11.15
1961	13721	11.15
1962	.	.
1963	.	.
1964	.	.
1965	.	.
.	.	.
.	.	.
.	.	.
.	.	.

- 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 the UK, and co2_pc for CO2 emissions per capita.

2) Summary Statistics and Overview of the Data

1. Using the *Algebra* editor construct the variables:
 - $\log(\text{RGDP per Capita})$
 - $\log(\text{CO2 Emissions per Capita})$
2. Why are log transformations of the data useful?
3. Plot $\log(\text{RGDP per Capita})$ and $\log(\text{CO2 Emissions per Capita})$ over time in separate plots
4. Plot $\log(\text{CO2 Emissions per Capita})$ against $\log(\text{RGDP per Capita})$ 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.

3) Econometric Investigation

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

$$\log(\text{CO2})_t = \beta_1 + \beta_2 \log(\text{RGDP})_t + \beta_3 \log(\text{RGDP})_t^2 + \epsilon_t \quad (1)$$

1. What is the marginal effect of $\log(\text{RGDP}$ per capita) on $\log(\text{CO2}$ per capita)? What if β_2 is positive and β_3 is negative?
2. Use the *Algebra* menu to construct the variable $(\log(\text{RGDP}))^2$
3. Estimate model (1) 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. 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.
4. Report the regression output in equation format showing estimated coefficients and standard errors.
5. How much variation of $\log(\text{CO2}$ Emissions per capita) is explained by variation in $\log(\text{real GDP per capita})$?
6. Are $\log(\text{RGDP})_t$ and $\log(\text{RGDP})_t^2$ individually significant? Are they jointly significant?
7. Are you confident in the conclusions of your model?

4) Time-Series Properties

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

1. Plot the Auto-correlation and Partial auto-correlation functions of $\log(\text{CO2})$ and $\log(\text{RGDP})$ and comment on the plots. What lag length would you suggest is optimal?
2. Estimate the model by adding a lagged value of $\log(\text{CO2})$ and comment on the results:

$$\log(\text{CO2})_t = \beta_1 + \beta_2 \log(\text{CO2})_{t-1} + \beta_3 \log(\text{RGDP})_t + \beta_4 \log(\text{RGDP})_t^2 + \epsilon_t \quad (2)$$

How could you improve the model?

References

Stern, D. I. (2004). The rise and fall of the environmental kuznets curve. *World development*, 32(8), 1419–1439.