A first regression in OxMetrics-PcGive

Ragnar Nymoen, August 22., 2011

Last update: August 15, 2013

1 Downloading data

Download the zip-file **KonsDataSim.zip** from the internet page (press ctrl and click **here**) . "Unzip"all files and save them in one of your own archives. In total there are 4 files in the folder: **konsum_sim.in7**, **konsum_sim.bn7**, **konsum_sim.xls** and **sim_konsumOx4.fl**)

The file pair **konsum_sim.in7** and **konsum_sim.bn7** contains the data set in Pc-Give/OxMetrics format. The reason why there are two files is because (with much larger data sets than here!) this keeps the file size down. The "bn7"file is a binary file which is called by the "in7"file with the same name. So, when you open a data set in PcGive format, you do that by loading an "in7"file into PcGive. The data numbers, which are in the associated "bn7"file, are then loaded automatically. You can also load the data set, in *in*7 format, from outside the program, simply by double clicking on the file name of the "in7"in Windows Explorer.

konsum_sim.xls is the same data set in Microsoft Excel format. PcGive can also read and write this format. In practice, unless your data set is very large (in which case you could use a comma separated file, "csv"), you may want to stick to this format in practice, since it is then "seamless"to switch between PcGive and Excel, as well as many other econometric software packages such as STATA, RATS or EViews. Sometimes you may want to organize your data in one program and do the econometric analysis in another program. In that case, it may be both confusing and take a lot of memory on your computer if you save the same data set in several different file formats.

sim_konsumOx4.fl is a batch file with the PcGive/OxMetrics code that have been used to generate the data set. It is a simple example of programming in PcGive! This file can be opened in any text editor, but you need to save the content as plain text (not in Word's "doc "or "docx"format), and with the "fl"extension. We will not comment on the content of this batch file specifically at this stage (but it is almost self explanatory: What is the consumption function written into the program?). You will see (and create) many, more advanced, batch files later in the course!

2 Loading data into PcGive and doing a first regression

To help you do your first regression in PcGive, we include two "step-by-step"instructions: The first for the newest versions of the program (relevant for us) and the other for an

older version (PcGive 10). The differences are not large, and the outputs are practically identical. All this and more is of course in the PcGive/Oxmetrics manuals—and in the built in Help system.

2.1 PcGive 11, 12 and 13 (OxMetrics 4,5 og 6)

Start OxMetrics. Click File-Open in the main menu of OxMetrics and choose **konsum_sim.in7** by two clicks (alternatively do the same for **konsum_sim.xls**). An illustration is shown in Figure 1 below:

Figure 1: Opening data
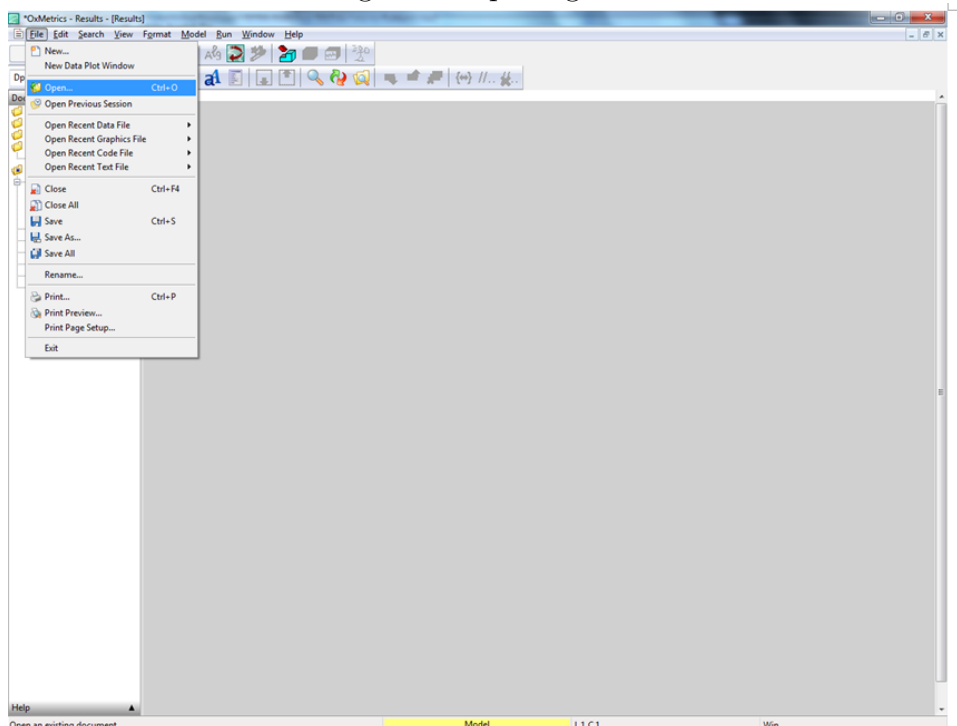
2.1.1 Estimating a cross section model

Click Model-Model in the main menu of OxMetrics. You then get a dialogue where you can use different model "types"for different data sets. In this course we will mainly consider "Models for Cross-section data"and "Models for time-series data". This artificial data set is organized as a time-series data set, but we will use it to estimate static models, and therefore the difference between cross-section and time-series is only a formality. Since cross-section regression gives the simplest output, choose first"Models for Cross-section data".

This means that your screen looks like the screen-capture in Figure 2 (differences may occur because of small variations in the interface between older/newer program versions, and differences in configuration. For example you may not have the GRCH and STAMP icons):

Figure 2: Cross section regression
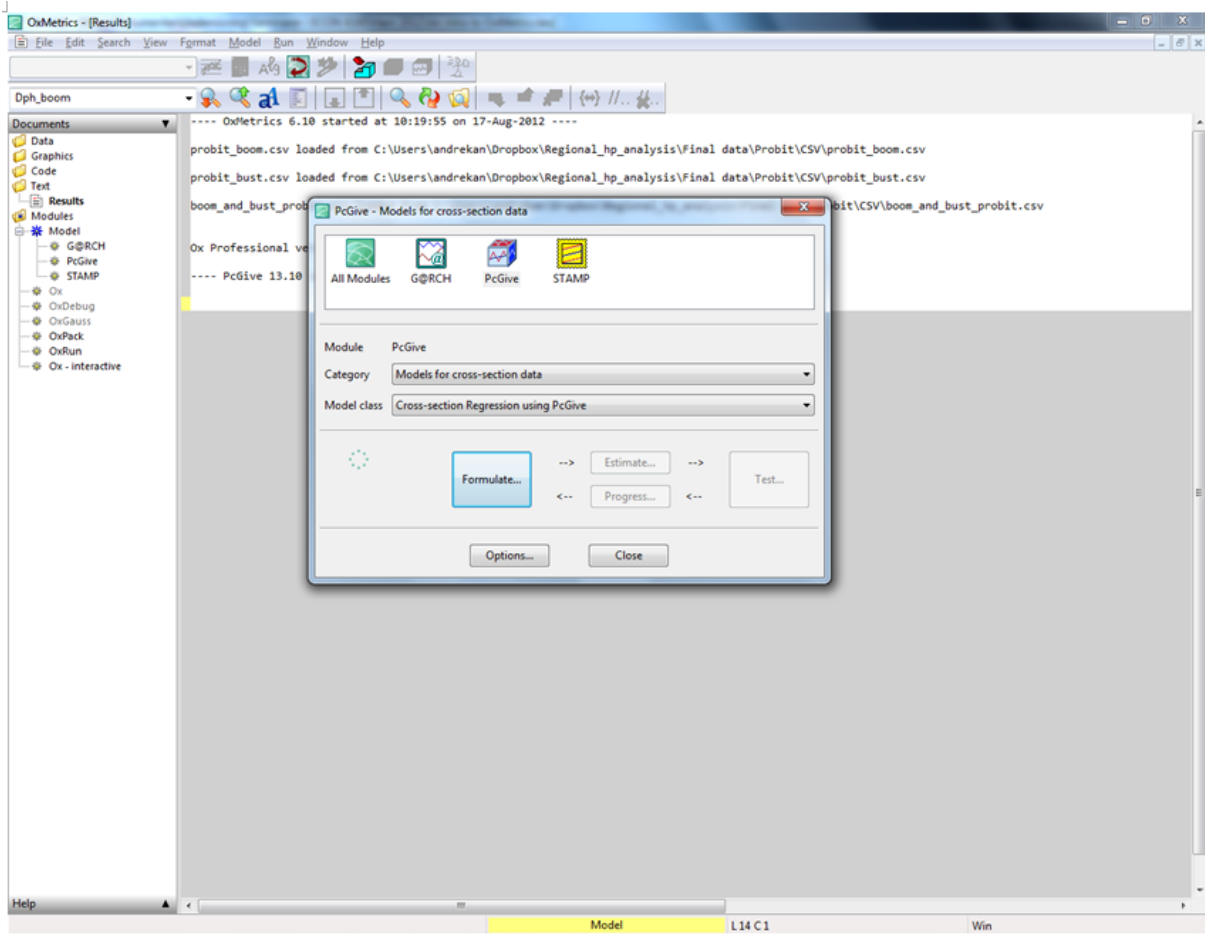
In the dialogue box called PcGive-Models for cross-section data: Click the "Formulate..."button. You will then see the screen capture illustrated in Figure 3 below.

Figure 3: Choosing variables in cross section regression
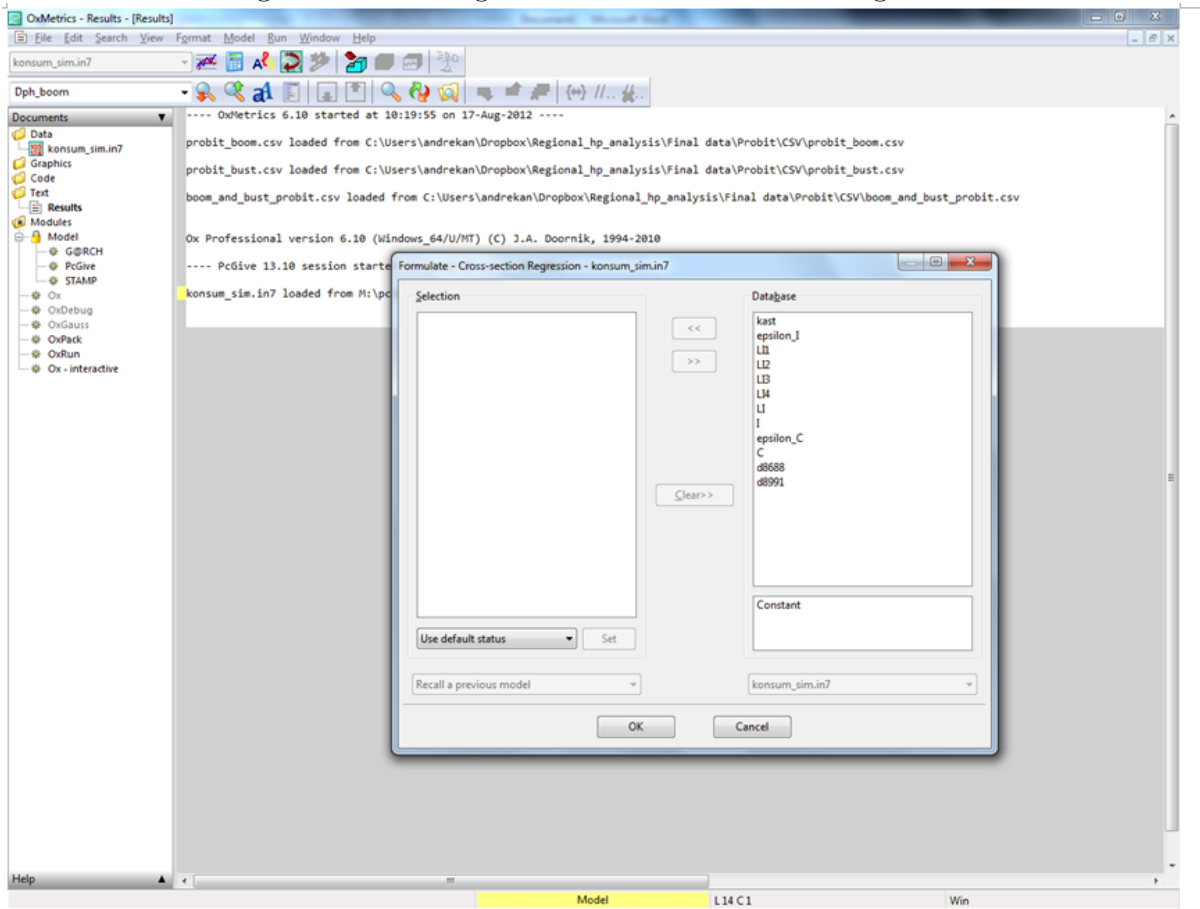
Then, choose the variable C (consumption) as the dependent (endogenous) variable by placing the cursor on C in the Database part of the dialogue and clicking once to choose (highlight) and then clicking the $<<$ button (alternatively, you can double click on C). C now appears in the Selection part of the dialogue, with the mark "Y" to the left (indicating that we want C as the endogenous variable in the regression). Note that also a Constant appears in the Selection part (below C). This means that the program puts in a constant in the regression by default (which we normally would prefer anyway, why?).

Choose I (income) as the explanatory variable (regressor) by placing the cursor on I in the Database part of the dialogue and clicking once to choose (highlight) and then clicking the $<<$ button (again, an option is to double click on C). I now appears in the Selection part of the dialogue and the screen should be identical to the picture in Figure 4.
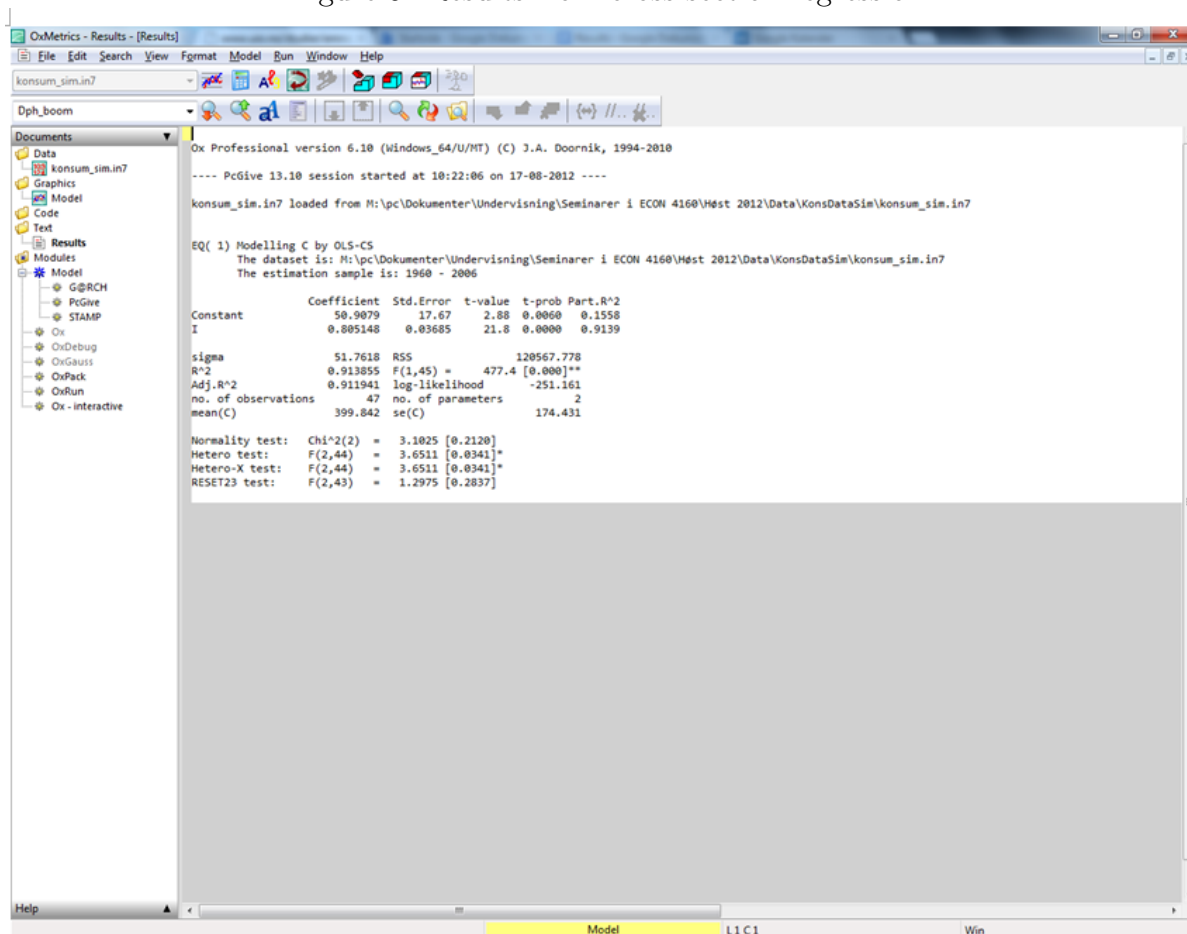
Figure 4: The chosen model
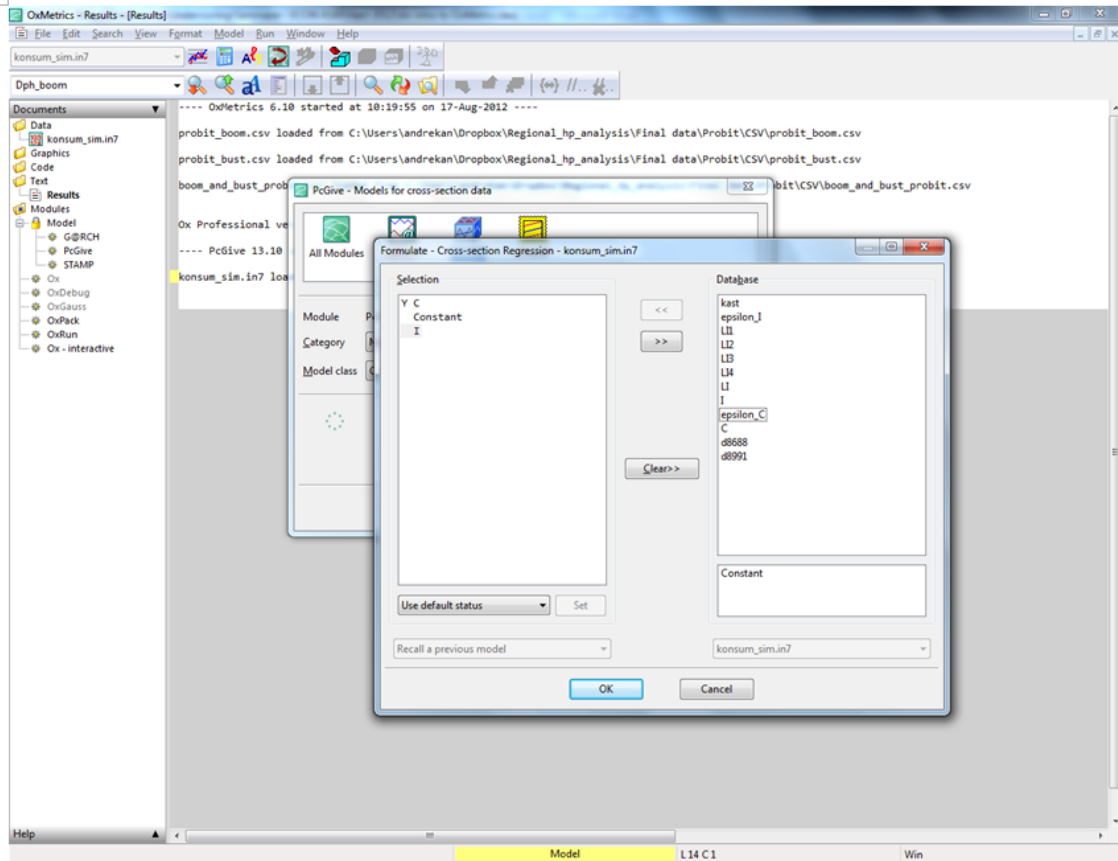
Click the OK button in the Formulate dialogue. If a dialogue named Choose the Autometrics Options appears, just leave the Automatic model selection box unchecked and click on the OK button.

We now get to the Estimate Cross-section Regression dialogue. By default PcGive suggests the largest sample available (1958–2007 in our case). It is easy to change the sample. To do so, change Estimation starts at from 1958 to 1960 and Estimation ends at from 2007 to 2006 by using the dialogue. Then click OK.

Depending on the configuration of your installation of the program you now get both numerical and graphical output. The numerical output, i.e., the estimated model is reported in the OxMetrics - Results window, and is shown in the screen-capture in Figure 5.

Figure 5: Results from cross section regression


```
OxMetrics - Results - [Results]

File  Edit  Search  View  Format  Model  Run  Window  Help

konsum_sim.in7

Dph_boom

Documents
  Data
    konsum_sim.in7
  Graphics
    Model
  Code
  Text
    Results
  Modules
  Model
    G@RCH
    PcGive
    STAMP
    Ox
    OxDebug
    OxGauss
    OxPack
    OxRun
    Ox - interactive

Help
```

```
Ox Professional version 6.10 (Windows_64/U/MT) (C) J.A. Doornik, 1994-2010

---- PcGive 13.10 session started at 10:22:06 on 17-08-2012 ----

konsum_sim.in7 loaded from M:\pc\Dokumenter\Undervisning\Seminarer i ECON 4160\Høst 2012\Data\KonsDataSim\konsum_sim.in7


EQ( 1) Modelling C by OLS-CS
       The dataset is: M:\pc\Dokumenter\Undervisning\Seminarer i ECON 4160\Høst 2012\Data\KonsDataSim\konsum_sim.in7
       The estimation sample is: 1960 - 2006

                   Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant              50.9079      17.67      2.88   0.0060   0.1558
I                    0.805148    0.03685      21.8   0.0000   0.9139

sigma                 51.7618  RSS               120567.778
R^2                  0.913855  F(1,45) =       477.4 [0.000]**
Adj.R^2              0.911941  log-likelihood      -251.161
no. of observations        47  no. of parameters           2
mean(C)               399.842  se(C)             174.431

Normality test:   Chi^2(2)  =   3.1025 [0.2120]
Hetero test:      F(2,44)   =   3.6511 [0.0341]*
Hetero-X test:    F(2,44)   =   3.6511 [0.0341]*
RESET23 test:     F(2,43)   =   1.2975 [0.2837]
```

```
Model          L1 C1                    Win
```

This output will be explained in terms of the tests from elementary econometrics and Greene's book during the first computer class.

We will focus in particular on the last part of the screen-capture which contains mis-specification tests that are (particularly) relevant for cross-section data.

### 2.1.2 Estimating a time series model

We also want to comment on mis-specification tests for time-series data models. The fastest route is to go back to Model-Model in the main menu of OxMetrics, and then change Models for Cross-section data and Models for time-series data - Single-equation Dynamic Modelling using PcGive, see Figure 6.

Figure 6: Time series regression
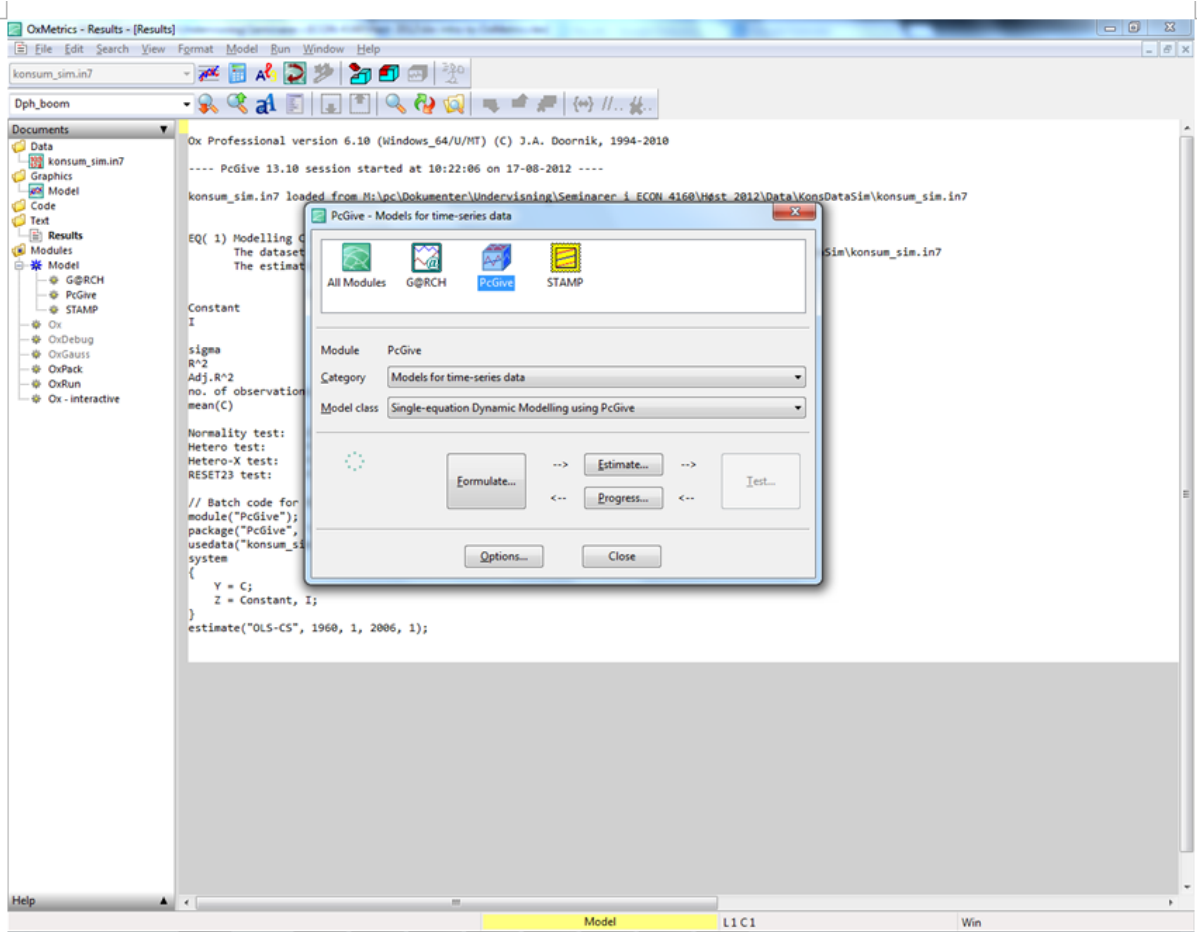
Then, again use the dialogue *Formulate...* button to specify the same regression as before. You will discover that since we now have told the program that the data set consists of time series the program will suggest that you formulate a dynamic model that includes lags of C and I, they are denoted C_1 and I_1, with the $\_1$ denoting the first lag (in general lag i is denoted $\_i$ in PcGive). However, for the moment we are only interested in replicating the results from our first (cross section) regression using the time series data environment, and therefore you can remove any lags form the model by first clicking on the lagged variable to be removed and then clicking the $>>$ button. An alternative too this is to reduce the lag length to zero prior to adding any variables to the Selection part of the dialogue. This is done by moving the cursor to the small white window in the $<<$ Lags part of the dialogue and changing from 1 to 0, see Figure 7 for an illustration.

Figure 7: Changing the lag order
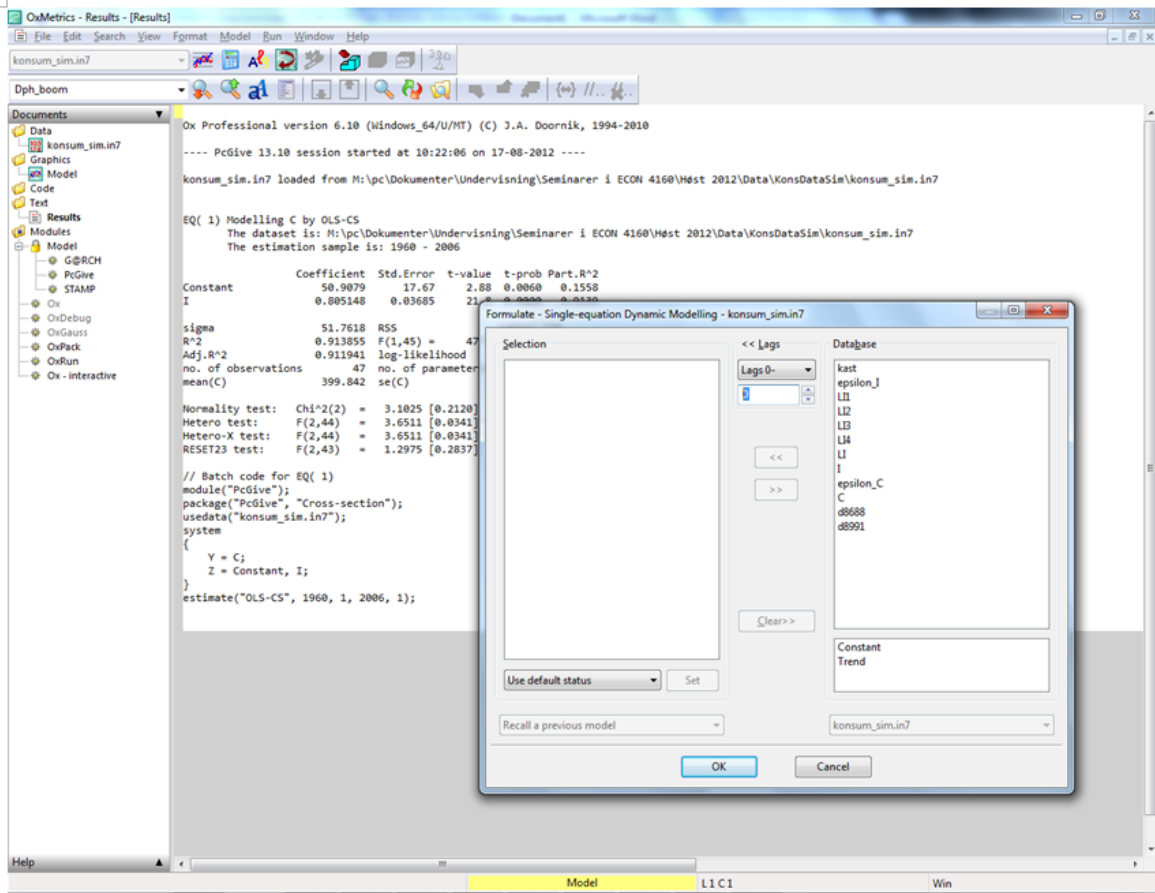
After having selected C as the dependent variable and I as the explanatory variable, you should see the picture displayed in Figure 8.

Figure 8: The chosen model
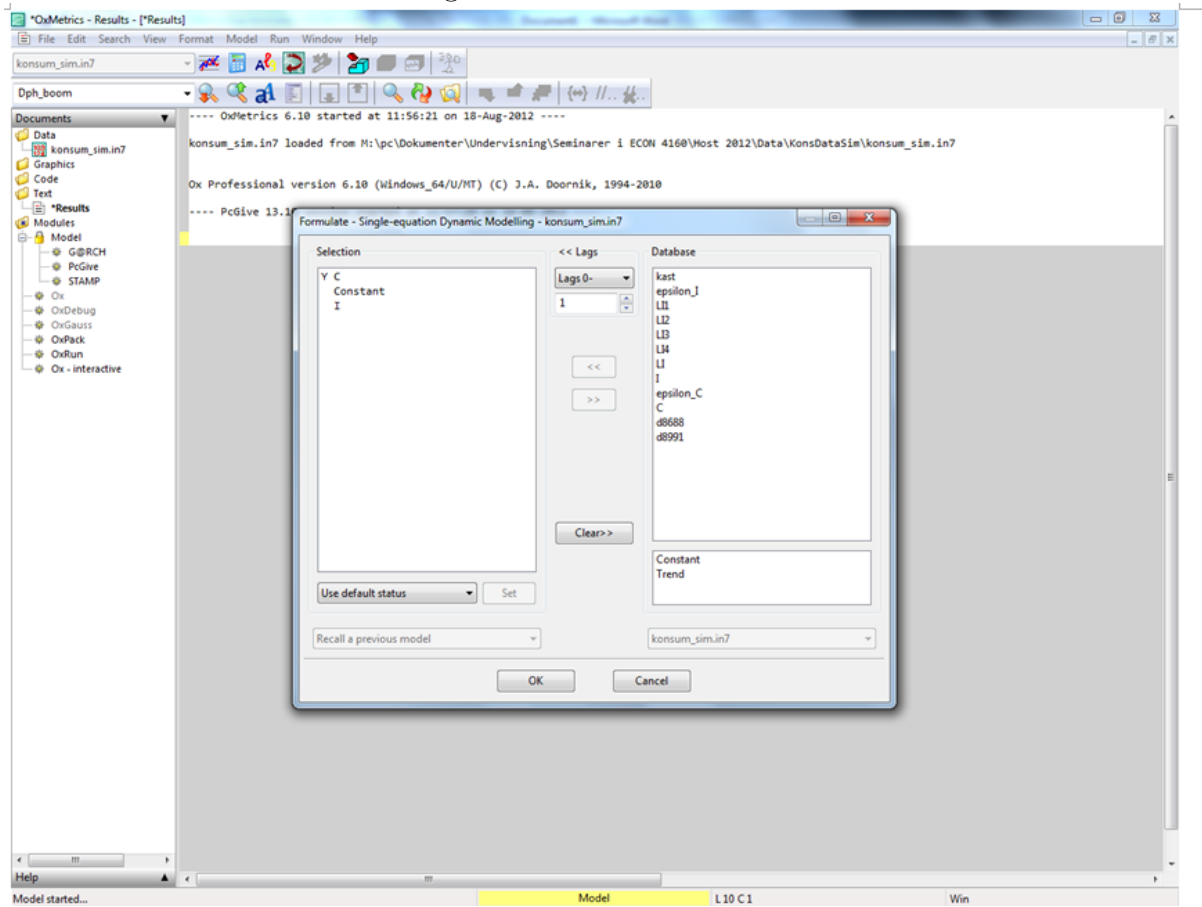
Press OK. The picture in Figure 10 will then appear on the screen. In the Model Settings-Single-equation Dynamic Modeling dialogue, accept Ordinary Least Squares, as chosen by the radio button in the screen capture below. Click the OK button.

Figure 9: Choosing estimation method for time series model
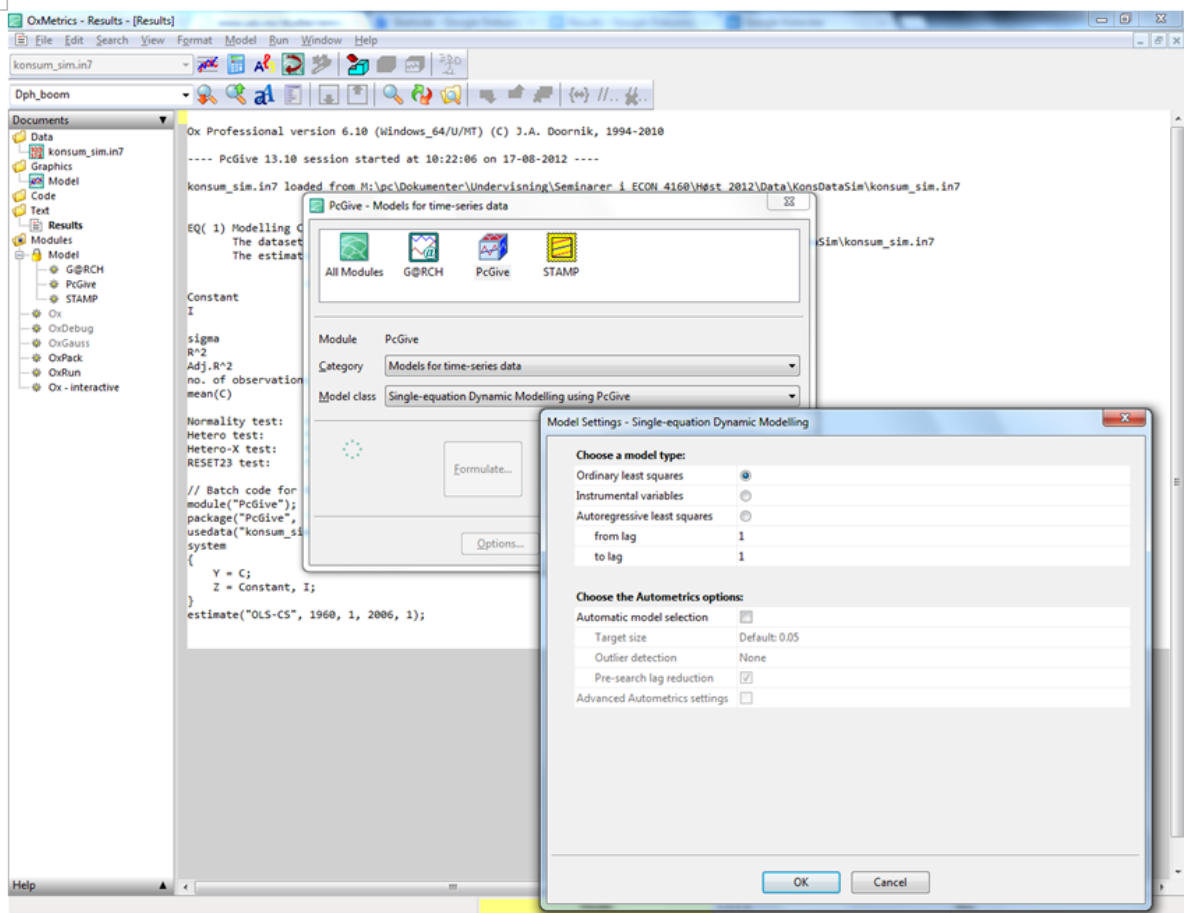
This leads to the Estimate Single Equation Dynamic Modelling dialogue shown in Figure 9. Here you can change the sample if you want, reserve some observations for forecasts (we will not do forecasting right now, so we keep the zero) and you can choose to do Recursive estimation by checking the box with that label. Recursive estimation means that the model is estimated on several shorter samples first. The initialization, which we set to 10 periods, will be the number of observations (as counted from the sample start) used in the first regression. After that, the model is estimated on a sample of 11 periods, an so on until the model is estimated using the full sample size. This generates much more information than a single (full sample) estimation, and PcGive lets us inspect that output graphically, as we shall see in the computer class and in the seminar.

Figure 10: Change sample, reserve observations for forecasting and recursive estimation
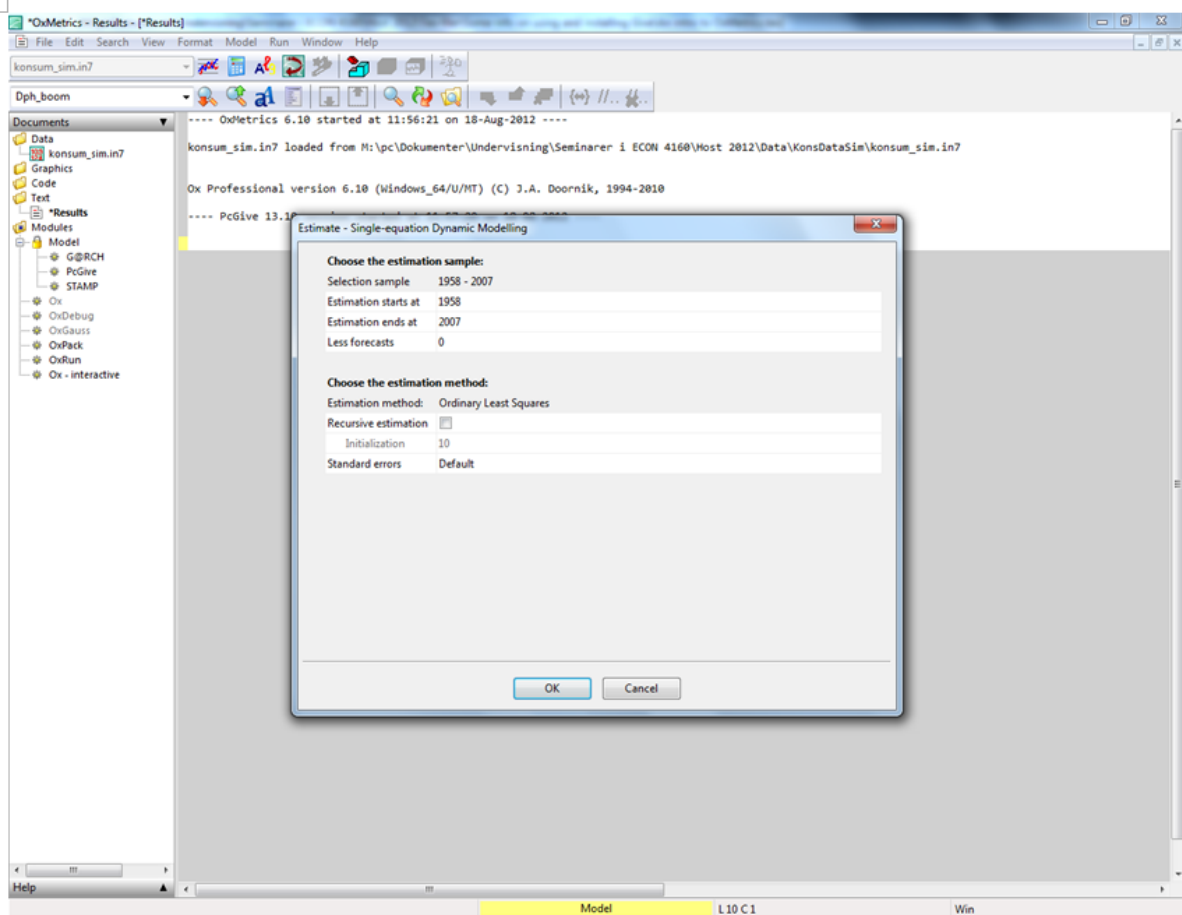
Note, however, that only the full sample estimation results are reported automatically in the results window. In our case, the results should be as displayed in Figure 11.

Figure 11: Results from time series regression
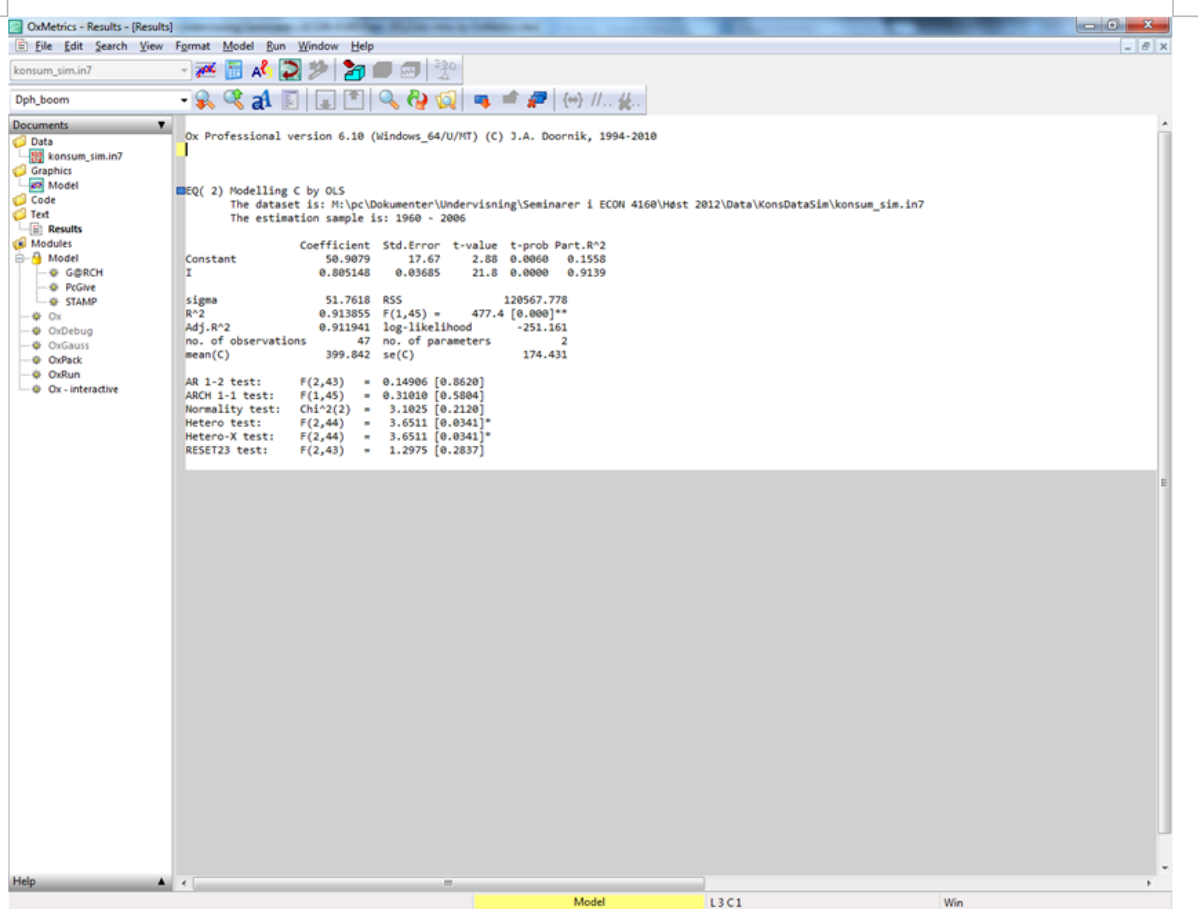
Most of the output is identical to the cross-section example, but there are two new mis-specification tests that we will comment on in the first computer class.

2.2 PcGive version 10

1. Start GiveWin

2. Click File-open and choose **konsum_sim.in7** by two clicks.

3. If PcGive has not already started, start PcGive, for example from the Modules menu in GiveWin.

4. From the Model menu in PcGive, choose 2. Single Equation Dynamic Modelling.

5. In the dialogue Data Selection choose C and click the $<<$ Add button. Delete C_1 if it appears in the Model part of the dialogue.

6. Follow the same procedure to choose I as explanatory variable.

7. In the Model Settings dialogue, accept Ordinary least squares by clicking OK.

8. In the Estimate Model dialogue, change 1958 to 1960 and 2007 to 2006.

9. Click OK. The estimated model is shown in the GiveWin Results window.

3 A small test/exercise to see what you have learnt

- Download the zip-file **US_hp_data.zip** from the internet page (press ctrl and click **here**)

- Unzip and save.

- You should then have the following 3 files:
  AnnualUSHP_data_1890_to_2009.xlsx, **AnnualUSHP_data_1890_to_2009.in7** and **AnnualUSHP_data_1890_to_2009.bn7**

- Load the data into PcGive

- Use the Time-Series model to regress log of real housing prices (hp) on the real interest rate (r) and real GDP (y) for the sample 1929 to 2009

- What does the results suggest?