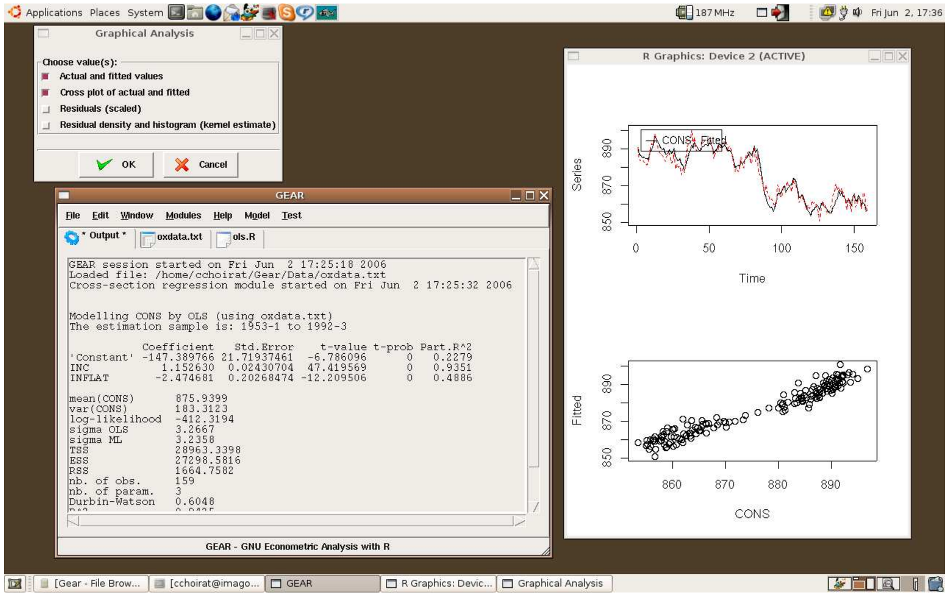
GEAR: GNU Econometric Analysis with R

Christine Choirat (Varese, Italy and Pamplona, Spain)
Paolo Paruolo (Varese, Italy)
Raffaello Seri (Varese, Italy)

useR! 2006, Vienna, 2006-06-17

A first glance



Choose value(s):

☑ Actual and fitted values
☑ Cross plot of actual and fitted
☐ Residuals (scaled)
☐ Residual density and histogram (kernel estimate)

```
GEAR session started on Fri Jun  2 17:25:18 2006
Loaded file: /home/cchoirat/Gear/Data/oxdata.txt
Cross-section regression module started on Fri Jun  2 17:25:32 2006


Modelling CONS by OLS (using oxdata.txt)
The estimation sample is: 1953-1 to 1992-3

           Coefficient   Std.Error    t-value t-prob Part.R^2
'Constant' -147.389766 21.71937461  -6.786096      0   0.2279
INC           1.152630  0.02430704  47.419569      0   0.9351
INFLAT       -2.474681  0.20268474 -12.209506      0   0.4886

mean(CONS)       875.9399
var(CONS)        183.3123
log-likelihood   -412.3194
sigma OLS        3.2667
sigma ML         3.2358
TSS              28963.3398
ESS              27298.5816
RSS              1664.7582
nb. of obs.      159
nb. of param.    3
Durbin-Watson    0.6048
```

Why another project?

We started the GEAR project from the observation that the econometric community and more generally the applied economics community lack a free, advanced and extensible software.

The available software for performing econometric analysis can be roughly divided in two classes:

1. Languages. Gauss and Ox are two common choices. Both are commercial products. They are high-level matrix languages and provide libraries oriented towards Econometrics.

2. Environments. Most econometric packages are commercial, the most popular at the moment being GiveWin (renamed OxMetrics) and its modules (PcGive in particular), and EViews .
   \begin{controversy}
   Gretl is free, cross-platform and released under the GPL. It is ready for undergraduate use, but a lot remains to be done for more advanced users and extensions are hard to implement.
   \end{controversy}

What R already provides

- Obviously, R can already be used for econometrics (for a review, see Cribari-Neto and Zarkos, 1999, Racine and Hyndman, 2002, Farnsworth, 2006 and A. Zeileis' CRAN task view for computational econometrics).

- However, the econometric models are implemented in a non-systematic way (moreover by independent authors so that many econometric methods are lacking whilst others are redundant).

Our requirements

GEAR (standing for "GNU Econometric Analysis with R") is meant to be both:

- a program with a graphical user interface (when performing standard tasks),

- a set of libraries (for more advanced analysis). It would ideally become a free and open-source alternative to EViews and GiveWin but also to Gauss and Ox.

It is really cross-platform (MS Windows, Mac OSX with X11 and Aqua, several Linux and Unix flavors).

What has been done yet?

Implementation choices

- Underlying language. We come from distinct econometric backgrounds (respectively 1. Ox, 2. Gauss and EViews, 3. RATS). R was chosen after some discussion (C++ and Python were other serious challengers) in late 2002 when we started the project. GEAR is entirely written in R (except for computer-intensive tasks which are coded as DLL's).

- GUI library. The choice was very hard and is still object of many discussions. At the time we started:
  - a free (in the sense of beer) version of Qt was only available for Linux,

  - GTK was rudimentary on Windows and not available on Mac OS X,
  - wxWidgets (which was called wxWindows in these days...) could have been an excellent choice, but the interface with R was far from being easy (even with the Python step).
  - Not convinced by any type of C or C++ GUI library, we decided to use the good old Tcl/Tk and the work of Peter Dalgaard on the package `tcltk` (`Rcmdr` by John Fox was a proof that a useful and rich GUI application could be developed). We found out that more sophisticated widgets were needed, in particular the classical extensions BWidgets and TkTable.
  - It might change in the future, especially in favor of wxWidgets.

A modular structure

- The basic tools provided by GEAR take the form of a standard R package `gear-main` (except from the requirement that the additional Tcl/Tk packages have to be installed for example using the ActiveTcl bundle and that an environment variable has to be defined under Windows).

- Specialized tools also take the form of standard R packages (*e.g.* `gear-coint` for cointegration analysis or `gear-panel` for panel data analysis) that depend on `gear-main`.

Libraries: Model-driven approach

- An econometric model is represented as an R environment.

```
eModel$sFile <- ‘‘/home/user/data.csv’’ # data name on disk
eModel$bModified <- FALSE # has data been modified?
eModel$sMethod <- ‘‘OLS’’ # estimation method
eModel$asX <- c(‘‘INC’’,‘‘INFLAT’’) # explanatory variable names
eModel$sY <- ‘‘CONS’’ # dependent variable name
eModel$bConstant <- TRUE # is there an intercept?

eFit <- fnEstimate.OLS(eModel) # estimation
fnPlot.OLS(eFit) # plots
```

- We have written utilities to save and load models in a human-readable way (using the functions `dput` and `dget`)

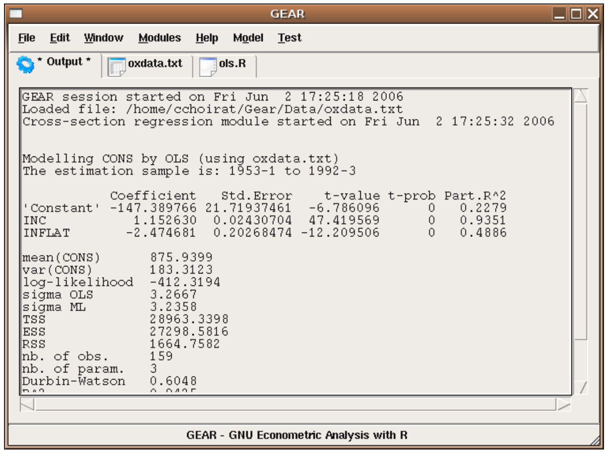
- The benefits of using the `R.oo` package will be further investigated.

GUI: Tcl/Tk interface

- Many unfruitful attempts. In particular IWidgets (problems with portability across different Linux distributions) and the MDI proposed by mkWidgets.

- The GUI is based on the NoteBook provided by BWidgets.

Output tab


GEAR
File Edit Window Modules Help Model Test
* Output *
oxdata.txt
ols.R
GEAR - GNU Econometric Analysis with R

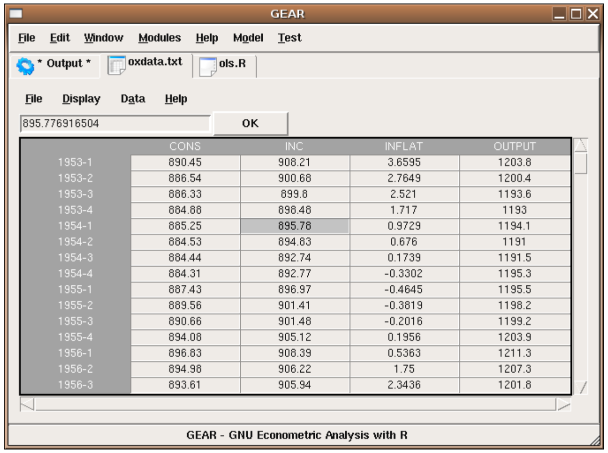
```
GEAR session started on Fri Jun  2 17:25:18 2006
Loaded file: /home/cchoirat/Gear/Data/oxdata.txt
Cross-section regression module started on Fri Jun  2 17:25:32 2006


Modelling CONS by OLS (using oxdata.txt)
The estimation sample is: 1953-1 to 1992-3

            Coefficient   Std.Error    t-value t-prob Part.R^2
'Constant' -147.389766 21.71937461  -6.786096      0   0.2279
INC           1.152630  0.02430704  47.419569      0   0.9351
INFLAT       -2.474681  0.20268474 -12.209506      0   0.4886

mean(CONS)      875.9399
var(CONS)       183.3123
log-likelihood  -412.3194
sigma OLS       3.2667
sigma ML        3.2358
TSS             28963.3398
ESS             27298.5816
RSS             1664.7582
nb. of obs.     159
nb. of param.   3
Durbin-Watson   0.6048
```

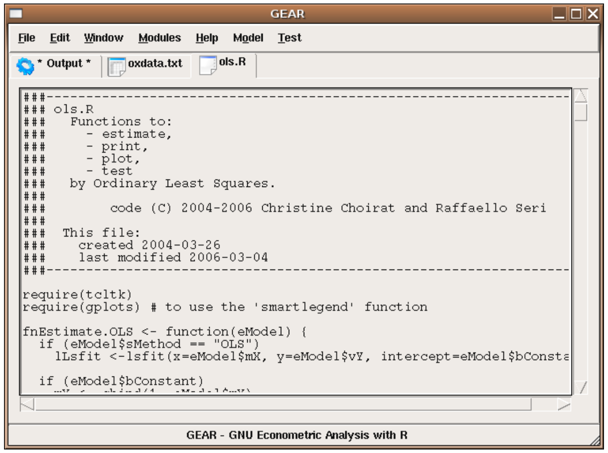
Spreadsheet tab


GEAR
File Edit Window Modules Help Model Test
* Output *
oxdata.txt
ols.R
File Display Data Help
OK
GEAR - GNU Econometric Analysis with R

895.776916504

	CONS	INC	INFLAT	OUTPUT
1953-1	890.45	908.21	3.6595	1203.8
1953-2	886.54	900.68	2.7649	1200.4
1953-3	886.33	899.8	2.521	1193.6
1953-4	884.88	898.48	1.717	1193
1954-1	885.25	895.78	0.9729	1194.1
1954-2	884.53	894.83	0.676	1191
1954-3	884.44	892.74	0.1739	1191.5
1954-4	884.31	892.77	-0.3302	1195.3
1955-1	887.43	896.97	-0.4645	1195.5
1955-2	889.56	901.41	-0.3819	1198.2
1955-3	890.66	901.48	-0.2016	1199.2
1955-4	894.08	905.12	0.1956	1203.9
1956-1	896.83	908.39	0.5363	1211.3
1956-2	894.98	906.22	1.75	1207.3
1956-3	893.61	905.94	2.3436	1201.8

Viewer tab


GEAR
File Edit Window Modules Help Model Test
* Output *
oxdata.txt
ols.R
GEAR - GNU Econometric Analysis with R

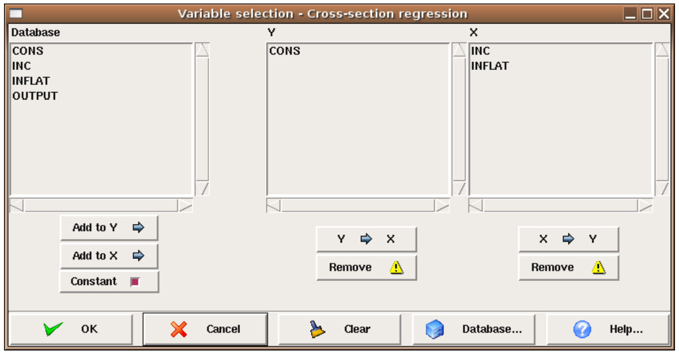
```
###----------------------------------------------------------------------
### ols.R
###   Functions to:
###     - estimate,
###     - print,
###     - plot,
###     - test
###   by Ordinary Least Squares.
###
###          code (C) 2004-2006 Christine Choirat and Raffaello Seri
###
###  This file:
###    created 2004-03-26
###    last modified 2006-03-04
###----------------------------------------------------------------------

require(tcltk)
require(gplots) # to use the 'smartlegend' function

fnEstimate.OLS <- function(eModel) {
  if (eModel$sMethod == "OLS")
    lLsfit <-lsfit(x=eModel$mX, y=eModel$vY, intercept=eModel$bConsta

  if (eModel$bConstant)
```

Variable selection (cross-section regression)

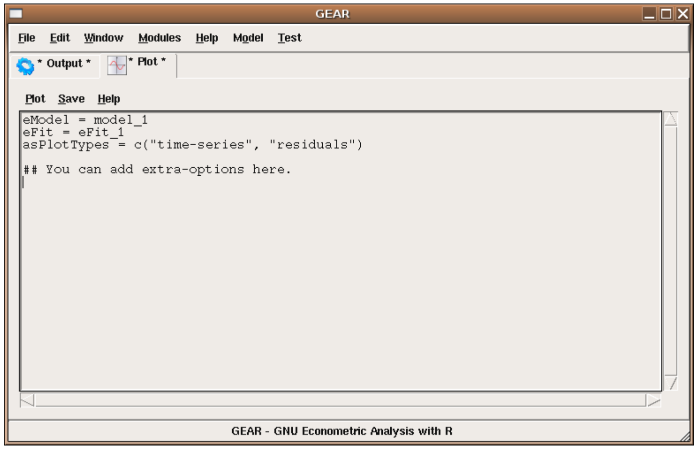


Database
CONS
INC
INFLAT
OUTPUT

Y
CONS

X
INC
INFLAT

A word on interactive graphics


GEAR
File Edit Window Modules Help Model Test
* Output *
* Plot *
Plot Save Help
GEAR - GNU Econometric Analysis with R

```
eModel = model_1
eFit = eFit_1
asPlotTypes = c("time-series", "residuals")

## You can add extra-options here.
```

What still has to be done?

A realistic answer is: “a lot...”! First, more modules need to be completed (of course borrowing a lot from what is already available but not limiting ourselves to writing wrappers):

- univariate time series,
- simultaneous equations,
- database interface,
- multivariate time series,

- financial econometrics,
- panel data.

More generally...

1. From a structural viewpoint
   - More OO paradigm.
   - Website, specific documentation.
   - Translations.
2. From a cosmetic viewpoint
   - Much polishing is needed (in particular extensive testing under MS Windows).
   - Artistic material (*e.g.* specific icons) is lacking.
   - Tk extras: ctext for syntax highlighting, more keyboard shorcuts.
   - Export output to HTML or LaTeX.