Computer Programming in Econometrics

Exercise

October 2009, Tinbergen Institute

Charles Bos

`cbos@feweb.vu.nl`

VU University Amsterdam

Tinbergen Institute

Setting exercise

We look at a simplified duration model, with a Weibull hazard rate. The data is generated, but imagine that the data considers unemployment duration. On $n = 145$ persons we observe

- Duration of unemployment (d_i in remainder)
- Constant
- Gender (0/1)
- City (0=country-side, 1=small village, 2=city)
- Age (in years)
- Log(Income)

We look for a link between d_i and the explanatory variables in X_i.
The dataset `duration.mat` is available on the website,
`http://www.tinbergen.nl/~cbos/courses/cpectr09.html`

Duration model

A simplified likelihood for such a case is the Weibull likelihood:

$$f(d_i; \lambda_i) = \gamma \lambda_i^\gamma d_i^{\gamma-1} \exp(-\lambda_i^\gamma d_i^\gamma)$$
$$\lambda_i \equiv \exp(X_i \beta)$$

The λ_i measures the individual ‘hazard’ to leave unemployment, and depends on the individual characteristics.
In the model, γ has to be positive else a negative density could result.
The total likelihood, of the full data set, is then

$$\mathcal{L}(d; \beta, \gamma) = \prod_{i=1}^{n} f(d_i; \lambda_i)$$

Exam question

For the exam, analyse this model and data set. Estimate the parameters, restrict γ if possible, compute standard errors, use the theory.

Output will be a set of programs, and a swift description of the output: E.g. create a word file, comment for each program what the intention is, copy/paste the important output of the program into the word file, and make clear what you did.

The report and programs will be evaluated on the basis of

1. Structure of solution (relating to analysis of problem) [20%]
2. Readability of programs/comments [20%]
3. Correctness of programming [20%]
4. Robustness of programming [20%]
5. Choice of descriptive statistics/graphs [10%]
6. Report, relating to structure of solution [10%]

Between brackets the approximative weight of each part in the final mark.

Deadline

Hand in both report and the programs together in a zip-file, before Monday 26/10 9.00h at Blackboard (`http://www.eur.edu/`).

For this exercise, you may work in groups of two students maximum. Working by yourself instead of with a groupmate is appreciated.

NB: If you miss the deadline, you flunked both this course and Math II... (yes, this is not logical; it is my intention to change this...)

Remainder pages give you a guideline on working out the exercise. It might be advantageous to follow the order.

Exercise

1. Take a piece of paper, analyse the exercise taking into account the questions below. What input/outputs will you have, what sizes, what routines, where will you stop to debug, when will you reanalyse on paper etc?
2. Read the data (check out `loadmat()`), calculate some first statistics, get a feel for the data. Maybe make some plots (see last pages) as well, e.g. of the durations against an index, durations against log-income, etc. Save program as `dur1.ox`.
3. Continue preparing a likelihood function (think of robustness, precision), and think of some starting values for the parameters. What is the value of the likelihood at the starting values? Is it reasonable? Save as `dur2.ox`
4. ...

Exercise (cont.)

4. Optimise likelihood over the parameters. What are the optimal parameters, how far off are they from the starting values? What is the gain in likelihood? (`dur3.ox`)
5. Implement an automatic restriction in the optimization to ensure that $\gamma > 0$ (`dur3b.ox`)
6. Estimate standard errors for the parameters. (`dur4.ox`)
7. Estimate a restricted model, where $\gamma \equiv 1$. Compute the likelihood ratio statistic, LR $= 2(\log \mathcal{L}(t; \hat{\beta}, \hat{\gamma}) - \log \mathcal{L}(t; \hat{\beta}, 1))$. A value of LR $> \chi^2_1(\alpha)$ = `quanchi(1-alpha, df)` would reject the null hypothesis. Conclusion? (`dur5.ox`)