Exam *Advanced Programming in Quantitative Economics*

15–19 August 2011

DGPE MSc course
University of Aarhus

The exam below can be solved over the course of the next weeks, in groups of a maximum of three students (smaller groups or individual work is permitted, and even preferred). Only one set of answers per group of students is necessary. The final results have to be handed in by email before Friday, Oct. 21 2011, 9.00h to `c.s.bos@vu.nl`.

1 Setting

The topic of the exam is a data set on transactions on IBM stocks. The file `ibm1012.csv` (available through the course web site) contains all transactions over the month December 2010 on this stock.

The exam centers not on the prices of the stocks, but only on the timing of the transactions and the durations between them. The intention is that you analyse these durations using a so-called Autoregressive Conditional Duration model. The question is how much of a correlation between these durations exists, and if they can be modelled in a satisfactory way.

2 The model

The Exponential Autoregressive Conditional Duration ACD(1,1) model is written as

$$y_i = \psi_i \varepsilon_i, \qquad y_i \sim \text{Exp}(\frac{1}{\psi_i}) \tag{1}$$

$$E(\varepsilon_i) \equiv 1, \qquad E(y_i) = \psi_i \tag{2}$$

$$\psi_i = \omega_0 + \gamma_1 y_{i-1} + \omega_1 \psi_{i-1} \tag{3}$$

$$\mathcal{L}(y_i; \theta) = \frac{1}{\psi_i} \exp\left(-\frac{y_i}{\psi_i}\right) \tag{4}$$

Notice that in this manner of writing the model ψ_i is the expected conditional duration of y_i (given the past), and it evolves in a deterministic format given the past duration and the past expected duration.

The likelihood for a certain set of parameters $(\omega_0, \gamma_1, \omega_1)$ can be constructed if the expected duration is known. This latter duration would have to be constructed in a loop:

1. Set $i = 0$, initialise $\psi_0 = \omega_0/(1 - \gamma_1 - \omega_1)$
2. While $i < N$, then calculate the expected duration for ψ_{i+1}, increase i, and repeat.

Given the expected duration, the individual loglikelihoods are easily calculated.

As the (expected) durations should all be positive, there are some restrictions on the parameters:

$$\omega_0 > 0, \qquad 0 < \gamma_1 < 1, \qquad 0 < \omega_1 < 1$$
$$\gamma_1 + \omega_1 < 1$$

It is rather hard to implement officially the restriction that $\gamma_1 + \omega_1 < 1$; you could probably get away with only checking within the likelihood function that this restriction is not broken and returning a 0 otherwise.

3 The data

This model assumes that the unconditional expected duration $\text{E}(y_i) = \omega_0/(1 - \gamma_1 - \omega_1)$ is constant, so it cannot react to the time of day. On the other hand, during opening and closing times durations tend to be shorter, at lunch time quite a bit larger. Therefore, you will have to treat the data before estimating the model.

As the file is rather large at 0.5 million observations, you should begin by trimming the file to retain only 1 out of each 100 subsequent observations. For these observations, get hold of the time (in seconds) of each transaction, and the difference between them.

Then, compute the average duration of a transaction for each period of 15 minutes within the day. That is to say, find all transactions which start between e.g. 9:15 and 9:30, and calculate their average. A good way to do this could be to realize that a day contains 24*4 15 minute intervals. Hence create a matrix `mCount` of size 96 x 2, for each transaction you could compute the interval to which it belongs (e.g. a transaction at 9:18 belongs to interval `j= floor((9+18/60)*4)`, increase `mCount[0][j]` by 1 and add the transaction duration to `mCount[1][j]`. After covering all transactions, you can easily calculate the average durations in all blocks. As a hint, your average durations could look like the ones in Figure 1.

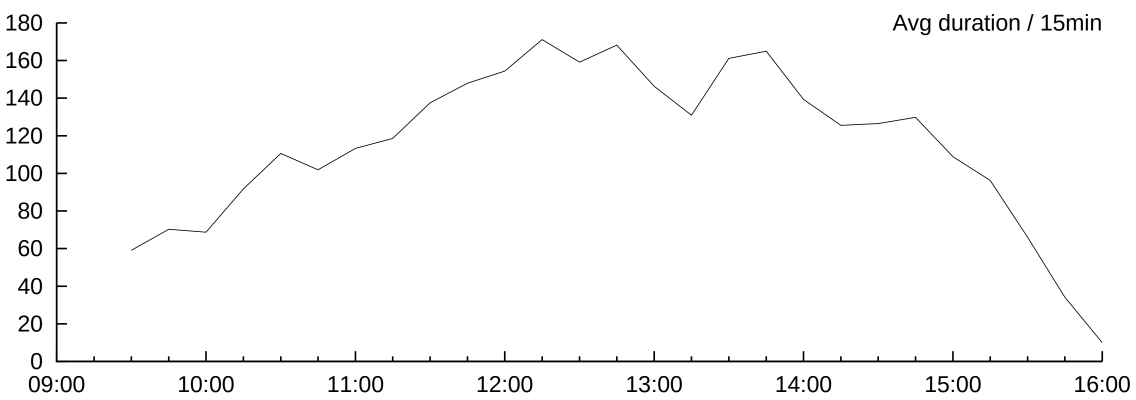

Figure 1: Average durations for IBM data

Now create your adapted durations data: Calculate

$$y_i = \frac{y_i^*}{\overline{y}_{j(i)}}$$

That is, your adapted duration y_i is the original duration y_i^* divided by the average duration in this period of the day $\overline{y}_{j(i)}$.

Further information on average conditional duration models you could find in Financial Econometric textbooks like Tsay (2010), though you should not really need more background information for the tasks at hand.

4 Questions

Read the data-set, trim it to the smaller size, and compute the average durations. Use these to find the adjusted durations y_i. These latter adjusted durations are the input for your estimation; find the parameters of the ACD model, using the theory on the approach of advanced programming. An incomplete list of things you might consider in your report is

- indication of the problem (no need to extensively repeat the exam)
- data description including peculiarities, how you transformed, what average durations you found etc
- parameter estimates of the model, including final log-likelihood and standard errors

Such a list is not exhaustive; own imagination is appreciated.

Write a report, maximum 5 pages excluding graphs/tables, on the analysis of the problem and the questions you answered. Hand in by email a zip-file with the report and the programs 'ready-to-run', including data files.

5 Help

This project is relatively large, especially if you don't start it in a structured fashion. Check clearly, before you start, how to split it up in more manageable subtasks. Furthermore, as a bit of help, the webpage contains a zip file with `genreacd.ox`: A program generating data from the EACD model, such that you can check your estimation program first on some simulated data. Also, it contains a small version of the data file, which you could use in testing your program to read the data.

Make sure that, while working on the project, you do not get stuck. If you do not know how to continue, also after sleeping for a good night, get in touch either with Manuel (`mlukas@creates.au.dk`), Kasper (`kolesen@creates.au.dk`) or me (`c.s.bos@vu.nl`), explaining clearly what the problem seems to be.

6 Evaluation of results

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.

References

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed.). New Jersey: John Wiley & Sons.