MAXIMUM LIKELIHOOD AND OPTIMISATION EXERCISES

PAWEL ZABCZYK

Centre for Central Banking Studies
Bank of England

ABSTRACT. In these exercises we will practice maximum likelihood and optimisation.

1. A POISSON EXAMPLE

In *Econometric Analysis*, Greene constructs a *Poisson distribution* example where the density for each independent observation x_i is given by

$$p(x_i, \ \theta^*) = \begin{cases} \frac{e^{-\theta^*}(\theta^*)^{x_i}}{x_i!} & \text{for } x_i \geq 0 \\ 0 & \text{for } x_i < 0 \end{cases}$$

with the underlying parameter $\theta^* > 0$. In reality, things modelled as random Poisson processes include: numbers of road traffic accidents, telephone calls etc..

Assume that we have observed the following sequence of x_i's

$$5, \ 0, \ 1, \ 1, \ 0, \ 3, \ 2, \ 3, \ 4, \ 1$$

Q1. Write down the likelihood function and then plot it in Excel.

Q2. Find the maximum of the likelihood function by calculus. What is it? Could you have guessed it faster, knowing that for a Poisson distributed random variable $E[X] = \theta^*$? How would you have found it in Excel?

2. MAXIMUM LIKELIHOOD LINEAR REGRESSIONS IN EVIEWS

The goal of this exercise is to estimate the following equation via maximum likelihood in Eviews and compare the estimates to those obtained using OLS:

$$y_t = c_1 + c_2 x_t^1 + c_3 x_t^2 + \nu_t$$

This is a simple linear model and we shall generate data using $c_1 = 0.5$, $c_2 = 2, c_3 = -3$ and $\sigma = 1$ where σ is the standard deviation of the normally distributed ν.

As this is our first Maximum Likelihood model in Eviews, we will work through it together.

Q3. The x^1, x^2 series (x1, x2 in Eviews) along with ν (nu) in Eviews have been saved in the file ml_exercise.wf1. Start by creating your y series.

To do that we would go to Quick → Generate Series and type in y = 0.5 + 2*x1 -3*x2 + nu.

**Q4.** Start by estimating c_1, c_2 and c_3 by OLS and note the results (save the output as OLS).

To do that go to Quick → Estimate Equation and type in y = c(1) + c(2)*x1 + c(3)*x2. You should get something like this:

Dependent Variable: Y
Method: Least Squares
Date: 04/11/16   Time: 11:42
Sample: 1 320
Included observations: 320
Y = C(1) + C(2)*X1 + C(3)*X2

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.448013	2.266104	0.197702	0.8434
C(2)	2.005604	0.007465	268.6681	0.0000
C(3)	-3.165334	1.655853	-1.911604	0.0568

R-squared	0.997636	Mean dependent var	93.34237
Adjusted R-squared	0.997621	S.D. dependent var	20.85305
S.E. of regression	1.017183	Akaike info criterion	2.881282
Sum squared resid	327.9876	Schwarz criterion	2.916610
Log likelihood	-458.0051	Hannan-Quinn criter.	2.895389
F-statistic	66876.50	Durbin-Watson stat	1.999732
Prob(F-statistic)	0.000000		

Q5. We now start by estimating the same equation by Maximum Likelihood. To do that go to Object → New Object → Logl and call the object likelihood. A blank window then opens, which is where we'll need to input our Likelihood object:
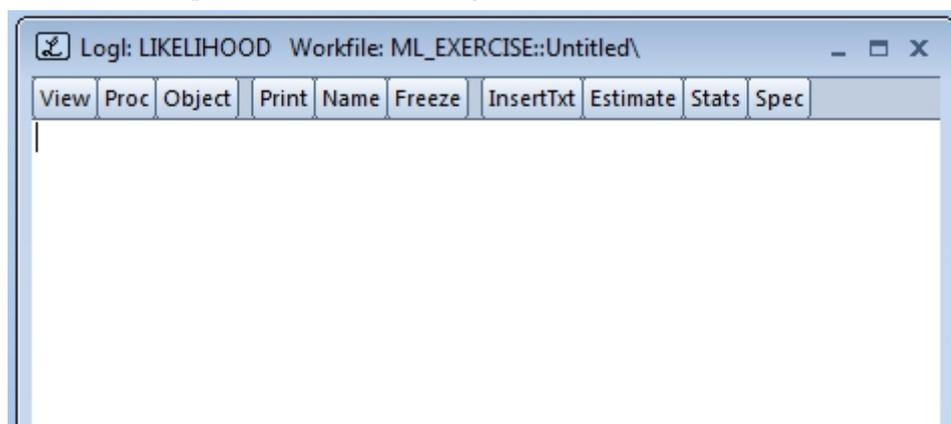
We now need to do several things:

- Give a name to the object that will hold values of the likelihood; this is done using the @logl command, so to save the results in an object called LL1 we would type @logl LL1

- Define the parameters that we will be maximising over; this is done using the `param` command. We also need to give an initial value to all the parameters - this is where the optimiser will start. In our case we have four parameters and we shall initialise the three regression coefficients in 0, and the fourth (corresponding to the log of the variance) in one. Accordingly we would type `param c(1) 0 c(2) 0 c(3) 0 c(4) 1`

- We then need to define the value of the (log)likelihood in every period. Before we do that, to cut on notation we define three auxiliary series / constants:

`res1 = y - c(1) - c(2)*x1 - c(3)*x2`

`pi = @acos(-1)`

`sigma = @exp(c(4))`

   - The first defines a series of residuals corresponding to `c(1) c(2) c(3)` and the data in `y` and `x`; how is this different from ν (`nu`)?
   - The second defines a series equal to π (which happens to equal *arccosine* of -1).
   - What does the third line do, and why?

- Armed with these definitions we can define the *contribution to the likelihood* in any period as

`LL1 = - ((1/2)*log(2*(pi))) - ((1/2)*log(sigma)) - 1/2*((res1^2)/sigma)`

   Where does this formula come from?

Once we've defined the value of the likelihood we can then estimate the parameters `c(1) c(2) c(3)` and `c(4)`. To do that we go to Estimate select the Berndt-Hall-Hall-Hausman algorithm and press ok. What are the reported coefficients? What has happened – why are they so different from the OLS ones? How could we eliminate the source of the problem?

Once you've corrected the problem you should obtain a table like this:

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	0.448010	2.351148	0.190549	0.8489
C(2)	2.005604	0.007889	254.2326	0.0000
C(3)	-3.165332	1.728255	-1.831519	0.0670
C(4)	0.024654	0.070698	0.348726	0.7273

|			
Log likelihood	-458.0051	Akaike info criterion	2.887532
Avg. log likelihood	-1.431266	Schwarz criterion	2.934636
Number of Coefs.	4	Hannan-Quinn criter.	2.906341

How do the coefficients compare with the OLS ones? Using the fact that $e^{0.02465} = 1.02496$ comment on how the variances compare. Why are they not identical under both approaches? Why does the value of the likelihood appear to be identical?

To ensure that the log-likelihood function has been properly maximised we can click on View → Gradients → Summary. What would you expect for an algorithm that has converged? Is this the case?

3. Estimating a fat-tailed regression by ML in Eviews (adapted from Dave Giles)

The previous example may lead you to wonder: what's the point of maximum likelihood if the results of this and OLS are so similar? In this exercise we will focus on a case in which OLS does not give good estimates - i.e. a setup with fat-tailed residuals (the example will also illustrate a few other things about maximum likelihood and OLS).

A fat-tailed distribution is one in which the probability of outliers is higher than in the normal distribution (this may be a sensible assumption when modelling financial returns, for example). The Student-t distribution has a density function with this property if the associated degrees of freedom are relatively small. Setting v the degrees of freedom equal to 3 we can plot the corresponding pdfs:

Standard Normal & Student t Densities
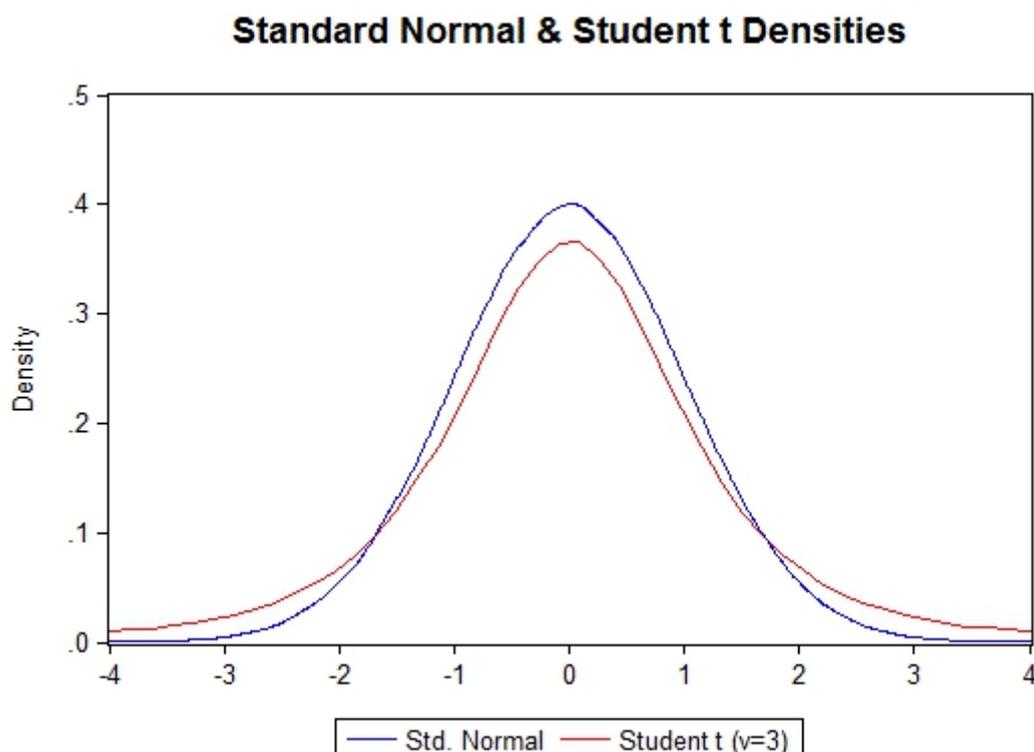
Q6. As in exercise 3, start by generating a y^{ft} series using residuals drawn from the Student-t distribution with 3 degrees of freedom ν^{ft} (nuft). Specifically, assume yft = 0.5 + 2*x1 -3*x2 + nuft.

Q7. Now estimate the coefficients c_1, c_2 and c_3 by OLS. What do you find (you can save the results as **OLS_FT**)? Comment on the R^2 and estimate precision. What is going on?

Q8. Now estimate the coefficients of c_1, c_2 and c_3 using Maximum Likelihood. To do that note that the density of a Student-t distributed random variable equals

$$f(x|v) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sqrt{v\pi}} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

Hint: You will need to use the fact that the $\Gamma(\cdot)$ function has been coded in Eviews as @*gamma* while the square root is @*sqrt*. Comment on the values of the coefficients and the value of the likelihood function?

4. The Steepest Ascent Method (adapted from Hamilton, Ch.5)

Q9. Assume that the likelihood function is as in the previous exercise, and let's pretend that the maximum of the likelihood is not immediately obvious. Suppose that your original guess for the parameters was $\theta^0 = (-1, 1)'$. Compute an improved guess for the maximum $\hat{\theta}$ using the method of steepest ascent. If you had to chose between a step size of 1 and 1/5, which would you select and why?

5. Hypothesis Testing (taken from Hamilton, Ch.5)

Q10. Suppose that we are interested in testing the hypothesis, at the 5% significance level, that

$$\theta_2 = \theta_1 + 1$$

where θ_1 and θ_2 are unknown parameters, and where the unrestricted log-likelihood is given by

$$\mathcal{L}(\theta) = -1.5\theta_1^2 - 2\theta_2^2.$$

Can you reject that hypothesis if the probability that a $\chi^2(1)$ variable exceeds 3.84 equals 0.05?
Hint: Use the Likelihood ratio test.