ECOM40006/ECOM90013 Econometrics 3
Department of Economics
University of Melbourne

Week 10 Tutorial Exercise

Semester 1, 2025

1. Ask any questions that you may have about the lectures, etc. If there is still time then please attempt the following questions.

2. Consider the multiple linear regression model

$$y = X\beta + u$$

where $u \mid X \sim \mathrm{N}(0, \sigma^2 I_n)$. Let us partition $X = [x_1, X_2]$ and $\beta = [\beta_1, \beta_2']'$ conformably so that x_1 is comprised of a single column of X and X_2 contains any remaining explanatory variables including the constant $\mathbf{1} = [1, 1, \ldots, 1]'$. X is assumed to have full column rank. We have n observations on all observables (y, X) and β contains $k+1 < n$ elements. Thus, the relationship for the typical observation might be written

$$y_i = \beta_1 x_{1i} + x_{2i}'\beta_2 + u_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where $[x_{1i}, x_{2i}']$ is the i-th row of X. Under our assumptions the joint density function for the data is

$$f(y; \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \right\}.$$

If one wishes to test hypotheses in this model then one possible approach is based on the Lagrange Multiplier test, where the test statistic takes the form

$$LM = \mathcal{S}_n(\tilde{\theta}_n)'(\mathcal{I}_n(\tilde{\theta}_n))^{-1}\mathcal{S}_n(\tilde{\theta}_n) = \underbrace{\frac{\mathcal{S}_n(\tilde{\theta}_n)'}{\sqrt{n}}}_{\to \mathrm{N}(0, \imath_\theta)} \underbrace{(n^{-1}\mathcal{I}_n(\tilde{\theta}_n))}_{\xrightarrow{p} \imath_\theta}{}^{-1} \underbrace{\frac{\mathcal{S}_n(\tilde{\theta}_n)}{\sqrt{n}}}_{\to \mathrm{N}(0, \imath_\theta)} \overset{H_0}{\underset{a}{\sim}} \chi^2_j,$$

where $\mathcal{S}_n(\theta) = \sum_{i=1}^n \partial \ln \mathcal{L}_n(\theta; y_i)/\partial\theta$, and

$$\imath_\theta = -\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \mathrm{E}\left[\frac{\partial^2 \ln \mathcal{L}_n(\theta; y_i)}{\partial\theta\partial\theta'} \right] = -\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \mathrm{E}\left[\mathcal{H}(\theta; y_i) \right].$$

In this case, $\theta = [\beta', \sigma^2]$ and the notation indicates that the score and information matrix from the unrestricted model should be evaluated at the constrained or restricted estimator, $\tilde{\theta}_n$, which is that obtained when the restrictions of the null hypothesis are imposed during estimation.

We can obtain the unrestricted maximum likelihood estimators (mles) of β and σ^2 as follows.

$$\ln \mathcal{L}(\beta, \sigma^2; y) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

Hence,

$$\begin{aligned}
\frac{\partial \ln \mathcal{L}(\beta, \sigma^2; y)}{\partial \beta} &= -\frac{1}{2\sigma^2} \times \frac{\partial (y - X\beta)'}{\partial \beta} \times \frac{\partial (y - X\beta)'(y - X\beta)}{\partial (y - X\beta)} \\
&= -\frac{1}{2\sigma^2} \times (-X') \times 2(y - X\beta) = \frac{1}{\sigma^2} X'(y - X\beta), \\
\frac{\partial \ln \mathcal{L}(\beta, \sigma^2; y)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)'(y - X\beta),
\end{aligned}$$

giving,

$$\mathcal{S}_n(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} X'(y - X\beta) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)'(y - X\beta) \end{bmatrix},$$

$$\mathcal{H}_n(\theta) = \begin{bmatrix} -\frac{1}{\sigma^2} X'X & -\frac{1}{\sigma^4} X'(y - X\beta) \\ -\frac{1}{\sigma^4}(y - X\beta)'X & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}(y - X\beta)'(y - X\beta) \end{bmatrix}$$

and, finally,

$$\begin{aligned}
\mathcal{I}_n(\theta) = -\operatorname{E}[\mathcal{H}_n(\theta)] &= \begin{bmatrix} \frac{1}{\sigma^2} X'X & \frac{1}{\sigma^4} X' \operatorname{E}[(y - X\beta)] \\ \frac{1}{\sigma^4} \operatorname{E}[(y - X\beta)'X] & -\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \operatorname{E}[(y - X\beta)'(y - X\beta)] \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma^2} X'X & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix},
\end{aligned}$$

as $\operatorname{E}[(y - X\beta)] = 0$ and $\operatorname{E}[(y - X\beta)'(y - X\beta)] = n\sigma^2$. If we choose to solve the first-order conditions then we obtain $\hat{\beta}_n = (X'X)^{-1}X'y$ and $\hat{\sigma}_n^2 = n^{-1}(y - X\hat{\beta}_n)'(y - X\hat{\beta}_n)$ as the mles. We note that the coefficient estimator is the same as OLS here but the variance estimator differs from the unbiased estimator s_n^2, which is not surprising as OLS really has nothing to say about variance estimation.

Suppose that in (1) you wish to test $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 \neq 0$. You can impose the restriction during estimation by simply dropping x_1 from the model when you fit it. Thus, the constrained log-likelihood is

$$\ln \mathcal{L}(\beta, \sigma^2; y) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(y - X[0, \beta_2']')'(y - X[0, \beta_2']'),$$

and, by symmetry with our earlier derivations, we have for the constrained model

$$\mathcal{S}_n(\theta) = \begin{bmatrix} 0 \\ \frac{1}{\sigma^2} X_2'(y - X_2\beta_2) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X_2\beta_2)'(y - X_2\beta_2) \end{bmatrix}$$

where the leading zero comes from the fact that β_1 does not appear in the constrained log-likelihood. Should we choose to solve the first-order conditions we obtain $\tilde{\beta}_n = [0, \tilde{\beta}'_{2,n}]'$, where $\tilde{\beta}_{2,n} = (X'_2X_2)^{-1}X'_2y$, and $\tilde{\sigma}^2_n = n^{-1}(y - X\tilde{\beta}_n)'(y - X\tilde{\beta}_n)$, where a tilde has been used to distinguish the constrained estimator. In order to calculate the LM or score test we need to evaluate its various components at the constrained, or restricted, mles. Consequently, defining $e_0 = y - X\tilde{\beta}_n = y - X_2\tilde{\beta}_{2,n} = M_{X_2}y$, we see that

$$\mathcal{S}_n(\tilde{\theta}_n) = \begin{bmatrix} \dfrac{1}{\tilde{\sigma}^2_n}X'e_0 \\ -\dfrac{n}{2\tilde{\sigma}^2_n} + \dfrac{1}{2\tilde{\sigma}^4_n}e'_0e_0 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\tilde{\sigma}^2_n}x'_1e_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where all the zeros are consequences of the first-order conditions for the constrained estimators, and

$$\mathcal{I}_n(\tilde{\theta}_n) = \begin{bmatrix} \dfrac{1}{\tilde{\sigma}^2_n}X'X & 0 \\ 0 & \dfrac{n}{2\tilde{\sigma}^4_n} \end{bmatrix},$$

We can now evaluate the LM test statistic, which you may recall is

$$LM = \mathcal{S}_n(\tilde{\theta}_n)'(\mathcal{I}_n(\tilde{\theta}_n))^{-1}\mathcal{S}_n(\tilde{\theta}_n),$$

in a variety of different ways. As a first approach, recognizing that the first-order conditions for the restricted mle imply that $X'_2e_0 = 0$, we can replace those zeros in the score and write

$$\begin{aligned} LM &= \begin{bmatrix} \dfrac{1}{\tilde{\sigma}^2_n}X'e_0 \\ 0 \end{bmatrix}' \begin{bmatrix} \tilde{\sigma}^2_n(X'X)^{-1} & 0 \\ 0 & \dfrac{2\tilde{\sigma}^4_n}{n} \end{bmatrix} \begin{bmatrix} \dfrac{1}{\tilde{\sigma}^2_n}X'e_0 \\ 0 \end{bmatrix} \\ &= \frac{e'_0X(X'X)^{-1}X'e_0}{\tilde{\sigma}^2_n} = n\frac{e'_0X(X'X)^{-1}X'e_0}{e'_0e_0} \end{aligned}$$

A couple of observations are now appropriate. First, provided that X_2 contains an intercept, as was assumed in the model definition, then the first-order conditions imply that

$$\mathbf{1}'e_0 = \sum_{i=1}^{n} e_{0,i} = \sum_{i=1}^{n}(y_i - x_{2,i}\tilde{\beta}_{2,n}) = 0 \implies \bar{e}_0 = 0$$

$$e'_0e_0 = (e_0 - \bar{e}_0\mathbf{1})'(e_0 - \bar{e}_0\mathbf{1}) = \sum_{i=1}^{n}(\bar{e}_{0,i} - \bar{e}_0)^2$$

$$\mathbf{1}'e_0 = \sum_{i=1}^{n} e_{0,i} = \sum_{i=1}^{n}(y_i - x_{2,i}\tilde{\beta}_{2,n}) = 0 \implies \bar{e}_0 = 0 \implies e'_0e_0 = (e_0 - \bar{e}_0\mathbf{1})'(e_0 - \bar{e}_0\mathbf{1}),$$

and so we can write the LM statistic as

$$LM = n\frac{e_0'X(X'X)^{-1}X'e_0}{\sum_{i=1}^{n}(e_{0,i}-\bar{e}_0)^2} = nR^2,$$

where R^2 denotes the coefficient of determination from the artificial or auxiliary regression of e_0 on X; that is,

$$e_0 = X\delta + \text{error}. \tag{2}$$

Note, this auxiliary regression has no interpretation, it is simply a device for getting our OLS software to numerically evaluate a test statistic for us, to save us from doing it by hand. Furthermore, even though we know that $X_2'e_0 = 0$, it is crucial that we include X_2 as regressors in the auxiliary regression otherwise our calculations will be wrong.

This is such an important result that I shall write it out once more just to be clear.

$$LM_1 = nR^2 \overset{H_0}{\underset{a}{\sim}} \chi_j^2,$$

where, in this case, $j = 1$ and R^2 comes from the auxiliary regression of the restricted residuals, e_0, on all of the explanatory variables in the original equation, X. Note that I have added a 1 subscript to the LM symbol to highlight that it is just the first of our possibilities.

A second approach that might be adopted is to leave all the zeros in the score, rather than replacing most of them with $X_2'e_0$, so that the statistic takes the form

$$LM_2 = \begin{bmatrix} \frac{1}{\tilde{\sigma}_n^2}x_1'e_0 \\ 0 \end{bmatrix}' \begin{bmatrix} \tilde{\sigma}_n^2(X'X)^{-1} & 0 \\ 0 & \frac{2\tilde{\sigma}_n^4}{n} \end{bmatrix} \begin{bmatrix} \frac{1}{\tilde{\sigma}_n^2}x_1'e_0 \\ 0 \end{bmatrix} = \frac{e_0'x_1(X'X)^{11}x_1'e_0}{\tilde{\sigma}_n^2}$$

$$= \frac{e_0'x_1(x_1'M_{X_2}x_1)^{-1}x_1'e_0}{\tilde{\sigma}_n^2} = n\frac{e_0'x_1(x_1'M_{X_2}x_1)^{-1}x_1'e_0}{e_0'e_0},$$

where $(X'X)^{11}$ denotes the 1,1 element of $(X'X)^{-1}$, an expression for which we obtained from Equation (4.3.6) of the Matrices handout. Noting that $e_0 = M_{X_2}y$, we see that

$$LM_2 = n\frac{y'M_{X_2}x_1(x_1'M_{X_2}x_1)^{-1}x_1'M_{X_2}y}{y'M_{X_2}y} = nR^2$$

where now the R^2 is the coefficient of determination from the regression of $M_{X_2}y$ on $M_{X_2}x_1$. Noting that the $M_{X_2}y$ are the OLS residuals from the regression of y on X_2, we see that $M_{X_2}x_1$ are the residuals of the regression of x_1 on X_2. That is, $LM_2 = nR^2 \overset{H_0}{\underset{a}{\sim}} \chi_1^2$, where R^2 comes from the auxiliary regression

$$e_0 = M_{X_2}x_1\phi + \text{error} \tag{3}$$

Returning to the expression

$$LM_2 = \frac{e_0'x_1(x_1'M_{X_2}x_1)^{-1}x_1'e_0}{\tilde{\sigma}_n^2}$$

observe that I can write

$$LM_2 = \frac{e_0' x_1 (x_1' M_{X_2} x_1)^{-1} (x_1' M_{X_2} x_1)(x_1' M_{X_2} x_1)^{-1} x_1' e_0}{\tilde{\sigma}_n^2} = \frac{\hat{\phi}^2}{\tilde{\sigma}_n^2 (x_1' M_{X_2} x_1)^{-1}},$$

where $\hat{\phi} = (x_1' M_{X_2} x_1)^{-1} x_1' e_0$ is the OLS estimator for ϕ in (3), and $\tilde{\sigma}_n^2 (x_1' M_{X_2} x_1)^{-1}$ is an estimate of the variance of $\hat{\phi}$ using $\tilde{\sigma}_n^2$ as an estimate of σ^2.[1] Now, the ratio of $\hat{\phi}$ to its estimated standard error is just a t-statistic that one would use to test the null hypothesis $H_0 : \phi = 0$. Given our assumptions, in this case you might reasonably compare this statistic against critical values from a t distribution with $n - k_2 - 1$ degrees of freedom.[2] However, if you are uncomfortable with that then simply compare the statistic with critical values from a standard Normal distribution. Note that LM_2 is the square of a statistic which we can treat as being $\mathrm{N}(0, 1)$. It is little wonder then that LM_2 is χ_1^2 under the null, at least asymptotically. Defining

$$LM_3 = t = \sqrt{LM_2}$$

to be the t statistic used to test $H_0 = \delta$ against a two-sided alternative in (3), using critical values from a standard Normal distribution, we see that taking the square root of LM_2 does not cause any outcome to change whether it belongs to either the acceptance or rejection region region and so LM_3 is the same test as LM_2.

Finally, we note that LM_3 is a measure of the partial correlation between y and x_1, given X_2, which will be invariant to which side of the equals sign the two variables lie. That is, if we consider the reverse regression

$$M_{X_2} x_1 = e_0 \varphi + \text{error} \tag{4}$$

then the t-ratio used to test $H_0 : \varphi = 0$ in (4) will be numerically equal to that used to test $H_0 : \phi = 0$ in (3) or, indeed, that in (2). (Some of this is proved in the Solutions.) So this gives us a fourth variant of the LM test: LM_4 is the t-statistic used to test $H_0 : \varphi = 0$ against a two-sided alternative in (4).

We note in passing that the coefficient of determination from (4) is the same as that from (3), and so yet another variant of the test is $LM_5 = nR^2 \overset{H_0}{\underset{a}{\sim}} \chi_1^2$ from (4).

The file `wages.csv` contains observations on the yearly wage of 474 bank employees along with some of their characteristics. Specifically in the first column of the file is an individual identifier, the second column contains information on the individual's yearly salary ($SALARY$), the third column the natural logarithm of

[1] Here is one for those who are really on the case. Because M_{X_2} is of rank $n - k_2 < n$, were I just to get (3) by pre-multiplying (1) by M_{X_2}, then I could generate a singular covariance matrix for $M_{X_2} u$, which is at odds with what I have done so far. However, a full rank decomposition — see Section 7.2 of the Matrices handout, or Section 7.3.2. of the same document, as the spectral decomposition will do the same job here — would allow me to write $M_{X_2} = CC'$, where C is an $n \times (n - k_2)$ matrix of full column rank such that $C'C = I_{n-k_2}$ and $C'X_2 = 0$. Then pre-multiplying (1) by C will give $C'y = C'x_1 + C'u$, and we note that $(C'x_1)'C'x_1 = x_1' M_{X_2} x_1$, $(C'x_1)'C'y = x_1' M_{X_2} y$, with $\mathrm{E}[u \mid X] = 0$ and $\mathrm{Var}[u \mid X] = \sigma^2 CC' = \sigma^2 I_{n-k_2}$, meaning that σ^2 is still the right parameter to be estimating and all the preceding analysis goes through largely untouched.

[2] The $n - k_2$ part is our effective sample size here (see Footnote 1 for where this comes from) and we lose a further degree of freedom because of the need to estimate the scalar parameter δ to completely estimate the fitted model (or conditional mean, which is another name for a regression function).

salary ($LOGSAL$), column 4 records the individual's number of completed years of schooling ($EDUC$), columns 5 and 6 contain the individual's starting salary ($SALBEGIN$) and the natural logarithm of this starting salary ($LOGSALBEGIN$), while the remaining 3 columns contain information on the employee's gender ($GENDER$: 0 for females, 1 for males) , whether or not they belong to a minority group ($MINORITY$: 0 for non-minority, 1 for minorities), and a categorical variable indicating the nature of the position in which the individual is employed ($JOBCAT$: 1 for administrative jobs, 2 for custodial jobs, and 3 for management jobs). We shall be interested in testing hypotheses in the model

$$LOGSAL_i = \beta_0 + \beta_1 EDUC_i + \beta_2 GENDER_i + \beta_3 MINORITY_i + \beta_4 JOBCAT_i + u_i,$$

where $i = 1, 2, \ldots, 474$.

(a) For each of the tests LM_1–LM_4 test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$.
Hint: When you run the auxiliary regression (3), you will find that R bases its estimate of σ^2 on the residuals from that equation rather than on e_0. Consequently, its reported t statistics are not the statistics that you need to calculate either LM_3 or LM_4. You need to construct the statistics by hand, using the residuals from the regression of $LOGSAL$ on an intercept, $GENDER$, $MINORITY$, and $JOBCAT$.

(b) Use $LM1$ to test $H_0 : \beta_3 = \beta_4 = 0$.

(c) Use $LM2$ to test $H_0 : \beta_3 + \beta_4 = 0$.

As a postscript, while the opg form of the information matrix makes it very convenient to calculate a form of the LM test statistic, it is the case that this particular form tends to be over-sized, so the actual size of the test can far exceed that which you thought you were getting when you chose your critical value(s). That said, tests based on auxiliary regressions are used a lot in practice.

3. Most of the maximum likelihood examples that we have encountered to date have been characterised by the fact that they have nice neat closed form solutions. In practice that is a very rare occurence and one must be able to find the roots of the first-order conditions numerically. The problem of solving (systems of) equations numerically has been a very active part of both computer science and applied mathematics for a very long time and there are very many approaches available that typically work better or worse depending on the circumstance. One technique for finding the root of a single equation is the bi-section method. Suppose that you have a continuous function that has opposite signs at either end of some closed interval $[a, b]$, $a < b$. Then it must be the case that the function is equal to zero at at least one point in the interval. The idea of the bi-section method is to repeatedly divide your interval into two parts, keeping the part for which the function takes different signs when evaluated at the boundary points and discarding the other. Eventually, remaining interval will become so narrow that you will be able to find an (approximate) solution to yur problem. The method pretty much always works but can be very slow because it is not really taking any account of the shape of the function.

There is potentially much gain available if one does take account of the shape of the function. One approach that does so is Newton's method. Suppose that we take a linear (or first-order Taylor) approximation to a continuous function $f(x)$ about some point x_0. Then we have

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

Simple re-arrangement yields

$$x \approx x_0 + \frac{f(x) - f(x_0)}{f'(x_0)}.$$

Recalling that we are trying to find a value for x such that $f(x) = 0$, we see that at such a value for x

$$x \approx x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Such a result lends itself to the creation of an iterative schem that requires knowledge of of the function f, its derivative with respect to x, f', and a starting value for x, x_0. (Note that there are ways of estuimating derivatives numberically if you don't have analytical derivatives.) The updating formula used in the iterative scheme is then

$$x_k \approx x_{k-1} - \frac{f(x_{k-1})}{f'(x_{k-1})}, \qquad k = 1, 2, 3, \ldots, K.$$

The idea is to keep iterating until some convergence criteria is satisfied. Typically, there is a number of such criteria in play at any given time. First, you would usually stop when $|x_k - x_{k-1}| < \epsilon$, for some ϵ sufficiently small. Alternatively, you may compare values of the function itself, e.g. $|f(x_k) - f(x_{k-1})| < \delta$, say. (Those of you really on the ball might see some similarities here to our classical testing approaches. In case you haven't seen it yet, another obvious thing that you might do is see wjether or not $f'(x_{k)} \approx 0$, as should be the case when you optimize a function.

Just in case none of the above happen happen there will be some number K, the maximum number of iterations, chosen such that, if a solution hasn't otherwise been reached then you just stop the process and go back and think about what is going wrong. Otherwise you run the risk of the computer being caught in an infinite loop. There are obvious potential problems. For example, if $f'(x_{k-1}) \approx 0$ then the procedure is likely to become very unstable. That is to say, such approaches don't work very well on flat or nearly flat surfaces. Similarly, if there are multiple roots to an equation, such methods may be able to find them but won't be able to distinguish between them. You will need to be able to think your way through such problems.

What to do if things go pear-shaped? Well one obvious thing to do is try a different starting value to see if there is a better path to the solution from there. Allternatively you may need to try a different method for finding the root of an equation. Note, optimatization problems involving continuous function reduce to solving a first-order condition, which is done by finding the root of an equation.

What to do if things go well? Well one obvious thing to do is try a different starting value to see if all roads lead to Rome or if they can lead to different roots of the equation. The latter situation is, of course, bad and something that will require further thought.

As mentioned above, we are typically interested in finding the roots of equations as part of some optimization process. So the functions that I have called f above are themselves the derivative(s) of some criterion function. And the derivatives f' are in fact going to be second derivatives. Specifically, if $Q(x)$ is the function that you are trying to optimize and it is twice differentiable then a second-order Taylor polynomial approximation about some point x_0 yields

$$Q(x) = Q(x_0) + Q'(x_0)(x - x_0) + \frac{1}{2}Q''(x_0)(x - x_0)^2.$$

Differentiating the right-hand side of this equation with respect to x yields a FOC of the form

$$Q'(x_0)x + Q''(x_0)(x - x_0) = 0$$

leading to our updating formula

$$x = x_0 - \frac{Q'(x_0)}{Q''(x_0)}.$$

We note that, when evaluated at x_0, the derivatives are not functions of x and so are constants for the purposes of differentiation. This formula readily extends to the multivariate case, where x is a vector, in which case $Q''(x_0)$ is a hessian matrix and we have

$$x = x_0 - [Q''(x_0)]^{-1}Q'(x_0)$$

In this form we see that any circumstance under which $Q''(x_0)$ is either singular or near singular is going to result in numerical instability, because it is the matrix equivalent of dividing by zero. As the hession can be thought of as a measure of curvature of a surface, we agin see that flat surfaces are hard to search across.

In the following exercise you will be asked to use an R package that is designed to allow you to solve systems of nonlinear equations. It is far more powerful than we really need now but it rovides you an opportunity to see what it can do and provides you will experience in loading a different sort of package. There is some work in setting up the information that it requires to work. In order to figure out what it requires enter help("nleqslv") into the console in RStudio. This will bring up the Help panel which will contain the following information:

```
nleqslv(x, fn, jac=NULL, ...,
        method = c("Broyden", "Newton"),
        global = c("dbldog", "pwldog",
                   "cline", "qline", "gline", "hook", "none"),
        xscalm = c("fixed","auto"),
        jacobian=FALSE,
        control = list()
)
```

Everything after the first and last lines here refers to optional arguments with which you need not concern yourself. As for the first line: x refers to a starting value for your algorithm, fn is a function that you are going to have to write for yourself, jac=NULL is an opportunity for you to provide an initial value for the hession matrix. By entering NULL you are telling the package to calculate a hessian numerically. This value will be updated as the packages iterates towards a solution. Finally the three dots is a place holder where you can supply any information required by your function. Specifically, it is here that you include your data. Simply enter the variable names in the same order as they appear in your function. For the rest, it is seeking information on things like the ϵ, δ, and K referred to above. There is scope to provide much more information as well. My advice is to simply run with the package defaults, which means do add any of this information yourselves. So, in my code, I have the function call

nleqslv(x, fn, jac=NULL, x, y)

where x and y are my two variables (described below). Good luck with it.

The file Stocks.xls (in the Exercises folder on the LMS) contains 240 monthly observations over the period January 1980 to December 1999. The variable RENDNCCO (y) contains excess returns on an index of 104 stocks in the sector of non-cyclical consumer goods and the variable RENDMARK (x) contains excess returns on an overall stock market index of the total market in the UK. We seek to model sectoral excess returns as a function of excess returns in the market overall, via a simple linear regression model

$$y_i = \alpha + \beta x_i + u_i, \qquad i = 1, \ldots, n.$$

Conditional on the x_i, the disturbances u_i are assumed to be iid, from either a Normal distribution $\mathrm{N}(0, \sigma^2)$ or a Cauchy distribution with density $f(u_i) = \left(\pi\left(1 + u_i^2\right)\right)^{-1}$.

(a) Determine the log-likelihood for the case of Cauchy disturbances. Show that the maximum likelihood estimates for α and β are obtained from the two conditions

$$\sum_{i=1}^{n} e_i \left(1 + e_i^2\right)^{-1} = 0$$
$$\sum_{i=1}^{n} e_i x_i \left(1 + e_i^2\right)^{-1} = 0,$$

where $e_i = y_i - \widehat{\alpha} - \hat{\beta} x_i$.

(b) Use the R package `nleqslv` to obtain mles for α and β based on the Cauchy distribution. Determine also the (asymptotic) standard errors of these estimates.

Hint: You probably want to use an opg form of the information matrix here.

Bigger Hint: The R package `nleqslv` needs you to give it a function to find the zero of. Below is the code that I used to define such a function. It may save you some time.

```
##
# Create a function to construct the score.
# This is needed by nleqslv.
# - theta are values for the parameters
# - x and y are your regression variables
```

```
# (in an obvious notation)
##
cauchy.score <- function(theta,x,y){
  alpha <- theta[1]
  beta <- theta[2]
  n <- length(y)
  e <- rep(0,n)
  for(i in 1:n)
    {
      temp=y[i]-alpha-beta*x[i]
      e[i]=temp/(1+temp*temp)
    }
  score.alpha <- sum(e)
  score.beta <- t(x) %*% e
  score=c(score.alpha,score.beta)
  return(score)
}
```

(c) Estimate α and β by maximum likelihood, based on the Normal distribution. Compute also the standard errors of these estimates. Compare the results with those obtained by OLS.

(d) Test the hypothesis that $\alpha = 0$ using the results in 3(b). Test this result also using the results in 3(c). Use 5% significance level.

(e) Construct histograms of the residuals from the models estimated in 3(b) and 3(c). On the basis of this information, which of the two models do you prefer. Justify your answer.