University College London
Department of Economics

G023: Econometric Theory and Methods*
Answers to Exercise 4

1. *Heteroskedasticity.*

   (a) The formula given is just $(X'X)^{-1}X'y$ for the case in which X and y are vectors. The variance of the OLS estimator is given by the formula $(X'X)^{-1}X'\Omega X(X'X)^{-1}$ where Ω is a diagonal matrix with ith diagonal element equal to $\omega^2\lambda(x_i)$. For the case in which X is a vector we have
   $$Var[\hat{\beta}|x] = \omega^2 \frac{\sum_{i=1}^n x_i^2 \lambda(x_i)}{\left(\sum_{i=1}^n x_i^2\right)^2}.$$

   (b) The formula given is just $(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ for the case in which X and y are vectors and Ω is as above. This is called the weighted least squares estimator because it can be obtained as
   $$\hat{\beta}_{GLS} = \arg\min_{\beta}(y - X\beta)'\Omega^{-1}(y - X\beta)$$
   which in this special case is
   $$\hat{\beta}_{GLS} = \arg\min_{\beta} \omega^{-2} \sum_{i=1}^n \frac{(y_i - \beta x_i)^2}{\lambda(x_i)}.$$
   In a model with heteroskedasticity the GLS estimator minimises a weighted sum of the squared residuals where the weights are inversely proportional to the conditional variance of Y given x.

   (c) Both estimators are unbiased - they are both special cases of the estimator $\hat{\beta}_H = (H'X)^{-1}H'y$ introduced earlier in the course. You

*

can show unbiasedness for each case directly. For example, for $\hat{\beta}_A$,

$$
\begin{aligned}
E[\hat{\beta}_A|x] &= \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i} E[\frac{1}{n}\sum_{i=1}^n y_i|x] \\
&= \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i} E[\frac{1}{n}\sum_{i=1}^n (\beta x_i + \varepsilon_i)|x] \\
&= \beta + \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i} E[\frac{1}{n}\sum_{i=1}^n \varepsilon_i|x] \\
&= \beta.
\end{aligned}
$$

To answer the second part of the question we must find specifications for $\lambda(x_i)$ which make the GLS estimator take the given forms. The answer is, for $\hat{\beta}_A$, $\lambda(x_i) = x_i$, and for $\hat{\beta}_B$, $\lambda(x_i) = x_i^2$. The OLS estimator is BLU when $\lambda(x_i) = 1$ for all x_i.

(d) The variance of the three estimators are as follows. Plug in the appropriate forms for $\lambda(x_i)$ to get the nine requested answers.

$$
\begin{aligned}
Var[\hat{\beta}_{OLS}|x] &= \omega^2 \frac{\sum_{i=1}^n x_i^2 \lambda(x_i)}{\left(\sum_{i=1}^n x_i^2\right)^2} \\
Var[\hat{\beta}_A|x] &= \omega^2 \frac{\sum_{i=1}^n \lambda(x_i)}{\left(\sum_{i=1}^n x_i\right)^2} \\
Var[\hat{\beta}_B|x] &= \frac{\omega^2}{n^2}\sum_{i=1}^n \frac{\lambda(x_i)}{x_i^2}
\end{aligned}
$$

Check that when each estimator is BLU, the result is what the GLS variance formula suggests. To get the result for, e.g., $\hat{\beta}_B$, we work as follows.

$$\hat{\beta}_B = \beta + \frac{1}{n}\sum_{i=1}^n \frac{\varepsilon_i}{x_i}$$

so,

$$
\begin{aligned}
Var[\hat{\beta}_B|x] &= Var\left[\frac{1}{n}\sum_{i=1}^n \frac{\varepsilon_i}{x_i}\right] \\
&= \frac{1}{n^2}E\left[\left(\sum_{i=1}^n \frac{\varepsilon_i}{x_i}\right)^2\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n E\left[\left(\frac{\varepsilon_i}{x_i}\right)^2\right] + \frac{1}{n^2}\sum_{i=1}^n \sum_{\substack{j=1 \\ i\neq j}}^n \frac{\varepsilon_i \varepsilon_j}{x_i x_j} \\
&= \frac{\omega^2}{n^2}\sum_{i=1}^n \frac{\lambda(x_i)}{x_i^2}
\end{aligned}
$$

the second line following because $E[\varepsilon_i|x] = 0$ and the last line because of the assumptions concerning lack of serial correlation.

2. MAFF used estimators of the type $\hat{\beta}_A$ in Question 2, with y_i interpreted as food expenditure in household i and x_i interpreted as number of household members in household i. This would be good (in a BLU sense) if the variance of household food expenditures increases linearly with the number of household members. Does it?

3. *Limiting distributions.* Issues of the sort introduced in this question dominate the literature on non-stationary time series.

   (a) This is just the formula $(X'X)^{-1}X'y$ for the special case in which $X = [1\vdots 2\vdots \ldots \vdots n]'$. We have

   $$\hat{\beta}_n = \beta + \frac{\sum_{t=1}^{n} t\varepsilon_t}{\sum_{i=1}^{n} t^2}$$

   Therefore $\hat{\beta}_n$ is a linear function of normal random variables, so it has a normal distribution. The expected value is

   $$E[\hat{\beta}_n] = \beta + \frac{\sum_{t=1}^{n} tE[\varepsilon_t]}{\sum_{i=1}^{n} t^2} = \beta.$$

   The variance is, using the $\sigma^2(X'X)^{-1}$ formula

   $$Var[\hat{\beta}_n] = \frac{\sigma^2}{\sum_{i=1}^{n} t^2}$$

   which gives the required result on using $\sum_{i=1}^{n} t^2 = n(n+1)(2n+1)/6$.

   (b) We have $E[Q_n] = 0$ and therefore, trivially, $\lim_{n\to\infty} E[Q_n] = 0$. Also, $Var[Q_n] = 6\sigma^2/((n+1)(2n+1))$ and $\lim_{n\to\infty} Var[Q_n] = 0$. Therefore $Q_n \overset{qm}{\to} 0$ which implies that $plim_{n\to\infty} Q_n = 0$.

   (c) Here $E[R_n] = 0$ and therefore, trivially, $\lim_{n\to\infty} E[R_n] = 0$. But

   $$\begin{aligned} Var[R_n] &= \frac{6n^3\sigma^2}{n(n+1)(2n+1)} \\ &= \frac{\sigma^2}{\frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}} \end{aligned}$$

   and $\lim_{n\to\infty} Var[R_n] = 3\sigma^2$.

   (d) The answer is $h(n) = n$. Why? Try finding $h(n)$ for the models

   $$\begin{aligned} y_t &= \beta t^{3/2} + \varepsilon_t \\ y_t &= \beta t^2 + \varepsilon_t. \end{aligned}$$

   Can you see a general result here?

4. *Likelihood, continuous outcomes.*

   (a) The log likelihood function is

   $$l(\theta; t) = -n\log\theta - n\bar{t}/\theta$$

where $\bar{t} = n^{-1} \sum_{i=1}^{n} t_i$. The gradient is

$$l_\theta(\theta; t) = -\frac{n}{\theta} + \frac{n\bar{t}}{\theta^2}$$

and the first order condition for $\hat{\theta}$, the MLE, is the solution to

$$l_\theta(\hat{\theta}; t) = -\frac{n}{\hat{\theta}} + \frac{n\bar{t}}{\hat{\theta}^2} = 0$$

which gives $\hat{\theta} = \bar{t}$.

(b) The second derivative of the log likelihood function is

$$l_{\theta\theta}(\theta; t) = \frac{n}{\theta^2} - 2\frac{n\bar{t}}{\theta^3}.$$

Written as a random variable, a function of random variables rather than realisations, we have

$$l_{\theta\theta}(\theta; T) = \frac{n}{\theta^2} - 2\frac{n\bar{T}}{\theta^3}.$$

The expected value of T is θ, so the expected second derivative is

$$\begin{aligned} E_{T_1 \dots T_n}[l_{\theta\theta}(\theta; T)] &= E_{T_1 \dots T_n}[\frac{n}{\theta^2} - 2\frac{n\bar{T}}{\theta^3}] \\ &= \frac{n}{\theta^2} - 2\frac{n}{\theta^2} \\ &= -\frac{n}{\theta^2}. \end{aligned}$$

The information "matrix" (here a scalar) is therefore

$$I(\theta) = \frac{n}{\theta^2}.$$

The limiting distribution of the MLE, with θ_0 denoting the true parameter value, is therefore as follows.

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \theta_0^2)$$

(c) Let $D = 1$ for uncensored realisations, $D = 0$ otherwise and let t_i^r denote the uncensored realisations. Let n_{ob} be the number of uncensored realisations. The log likelihood function is then as follows.

$$l(\theta; t) = -n_{ob} \log \theta - \frac{1}{\theta} \sum_{i=1}^{n} d_i t_i^r - \frac{1}{\theta}(n - n_{ob})c$$

The gradient of the log likelihood function is

$$l_\theta(\theta; t) = -\frac{n_{ob}}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} d_i t_i^r + \frac{1}{\theta^2}(n - n_{ob})c$$

which leads to the following expression for the MLE.

$$\hat{\theta} = \frac{1}{n_{ob}} \left(\sum_{i=1}^{n} d_i t_i^r + (n - n_{ob})c \right)$$

Let $z_i = \min(t_i, c)$ then the MLE can be written as

$$\hat{\theta} = \frac{1}{n_{ob}} \sum_{i=1}^{n} z_i = \frac{n}{n_{ob}} \bar{z}.$$

The second derivative of the log likelihood function, written as a function of random variables rather than realisations is

$$l_{\theta\theta}(\theta; t) = \frac{N_{ob}}{\theta^2} - 2\frac{1}{\theta^3} \sum_{i=1}^{n} D_i T_i^r - 2\frac{1}{\theta^3}(n - N_{ob})c.$$

The expected value of the second derivative of the log likelihood function is[1]

$$E_{T_1 \ldots T_n}[l_{\theta\theta}(\theta; T)] = -\frac{n}{\theta^2}\left(1 - \exp(-c/\theta)\right)$$

and so the information "matrix" is

$$I(\theta) = \frac{n}{\theta^2}\left(1 - \exp(-c/\theta)\right).$$

The limiting distribution of the MLE , with θ_0 denoting the true parameter value, is

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \frac{\theta_0^2}{1 - \exp(-c/\theta_0)}).$$

Note that as c becomes small the variance of the limiting distribution increases without limit.

(d) Using the parameterisation invariance result discussed in the course notes, since $\lambda = \theta^{-1}$, the MLE $\hat{\lambda} = \hat{\theta}^{-1}$. λ is a continuous function of θ around any value $\theta_0 \neq 0$ and so by Slutsky's Theorem if $\theta_0 \neq 0$ then $\hat{\lambda}$ converges in distribution to a normal random variable. The variance can be found using the delta method. Let $\lambda_0 = \theta_0^{-1}$. There is:

$$\hat{\lambda} - \lambda_0 \simeq \left(\hat{\theta} - \theta_0\right)\left(-\frac{1}{\theta_0^2}\right)$$

and so, arguing informally,

$$Var\left(n^{1/2}\left(\hat{\lambda} - \lambda_0\right)\right) \simeq Var\left(n^{1/2}\left(\hat{\theta} - \theta_0\right)\right)\left(\frac{1}{\theta_0^4}\right) \simeq \frac{\theta_0^2}{\theta_0^4} = \frac{1}{\theta_0^2} = \lambda_0^2$$

and finally

$$n^{1/2}(\hat{\lambda} - \lambda_0) \xrightarrow{d} N(0, \lambda_0^2).$$

[1] Take expectation first with respect to T_i^r's given the D_i's (using the truncated distribution) and then with respect to D_i's.

5. *Likelihood, normal distribution.* Let $\theta = \{\mu, \eta\}$. The log likelihood function is

$$l(\theta) = -\frac{m}{2}\log(2\pi) - \frac{m}{2}\log(\eta) - \frac{1}{2\eta}\sum_{i=1}^{m}(Y_i - \mu)^2$$

with first derivatives:

$$l_\theta(\theta) = \begin{bmatrix} \frac{1}{\eta}\sum_{i=1}^{m}(Y_i - \mu) \\ -\frac{m}{2\eta} + \frac{1}{2\eta^2}\sum_{i=1}^{m}(Y_i - \mu)^2 \end{bmatrix}.$$

The solution to $l_\theta(\hat{\theta}) = 0$ gives the required answer. This is the approximate distribution.

$$n^{1/2}\begin{bmatrix} \hat{\mu} - \mu_0 \\ \hat{\eta} - \eta_0 \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \eta_0 & 0 \\ 0 & 2\eta_0^2 \end{bmatrix}\right)$$