University College London
Department of Economics

G023: Econometric Theory and Methods
Answers to Exercise 5

1. *Models for binary data.*

   (a) Writing the index as $\beta_0 + \beta_1 lfincpc$, the logit model estimates are as follows.

Coefficient	Estimate	Est. std. err.
β_0	-2.188	0.327
β_1	0.333	0.067

   The Wald test statistic for $H_0 : \beta_1 = 0$ is 25.05 and $P[\chi^2_{(1)} > 6.63] = 0.01$, so we reject H_0 using a test with approximate size 0.01. An approximate 95% confidence interval for the coefficient is $[0.20, 0.46]$. The coefficient is moderately well determined. The probability that a household with mean log family income per head (4.89) buys fruit juice is estimated to be 0.36. For households with 20% (100%) higher family income per head the estimated probability of fruit juice purchase is 0.38 (0.42).

   (b) The estimates for the probit model are as follows.

Coefficient	Estimate	Est. std. err.
β_0	-1.348	0.199
β_1	0.204	0.041

   The coefficient in the probit model is smaller - the ratio (logit:probit) is 1.63. The ratio of the two intercept estimates is 1.62. The probability that a household with mean log family income per head (4.89) buys fruit juice is estimated to be 0.36. For households with 20% (100%) higher family income per head the estimated probability of fruit juice purchase is 0.38 (0.42). These are, to the accuracy reported, exactly as in the logit estimation.

i. The Wald test statistic for the squared term when only it is included is 0.0004. We would not reject the null hypothesis that the coefficient on squared log income per head is zero at any reasonable test size (e.g. comparing with 2.71 for a size 0.10 test). The likelihood ratio test statistic for the joint hypothesis under which both the squared and cubed terms have zero coefficients is 0.437 and again we would not reject the hypothesis.

ii. With the index written as $\beta_0 + \beta_1 lfinc + \beta_2 \log(memhh)$ the results are as follows.

	Logit		Probit	
Coefficient	Estimate	Est. std. err.	Estimate	Est. std. err.
β_0	-3.899	0.381	-2.372	0.228
β_1	0.542	0.071	0.329	0.043
β_2	0.315	0.093	0.191	0.057

For the logit model the likelihood ratio test statistic for the hypothesis $H_0 : \beta_2 = -\beta_1$ is 25.48 and we reject the hypothesis using any reasonably sized test The ratios of the coefficients across model forms are 1.72 in both cases.

iii. Including region indicators we find that purchase probability is perhaps a little higher in London and the South East, but not especially low in Scotland and the North of England relative to other regions.

2. *Likelihood.*

(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

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

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) With $\lambda = \theta^{-1}$ the log likelihood function is

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

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

$$l_\theta(\theta; t) = \frac{n}{\lambda} - n\bar{t}$$

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

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

which gives $\hat{\theta} = 1/\bar{t}$. Note the MLE of the transformed parameter is what one obtains when the transformation is applied to the MLE of the original parameter. The second derivative of the log likelihood function is

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

so the information "matrix" (here again a scalar) is

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

and the limiting distribution of the MLE, with λ_0 denoting the true parameter value (note that is θ_0^{-1}), is therefore as follows.

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

3. *Likelihood for count data.*

(a) The log likelihood function is

$$l = \sum_{i=1}^{n} (y_i\,(x_i'\theta) - \exp(x_i'\theta) - \log(y_i!))$$

(b) The partial derivatives with respect to θ are as follows.

$$l_\theta = \sum_{i=1}^{n} (-y_i - \exp(x_i'\theta))\, x_i$$

Here x_i is a column vector of values of covariates and l_θ is a column vector of derivatives. Here are the first order conditions satisfied by the MLE $\hat{\theta}$.

$$\sum_{i=1}^{n} \left(-y_i - \exp(x_i'\hat{\theta})\right) x_i = 0$$

Here "0" denotes a column vector of zeros.

(c) The matrix of second derivatives of the log likelihood function is

$$l_{\theta\theta'} = -\sum_{i=1}^{n} \exp(x_i'\theta) x_i x_i'$$

which note does *not* depend on the values of the outcomes $y_1, \ldots, y_n$. When $\theta = \theta_0$ the information matrix is

$$I(\theta_0) = E\left(\sum_{i=1}^{n} \exp(x_i'\theta) x_i x_i'\right).$$

and

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \left(plim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \exp(x_i'\theta_0) x_i x_i'\right)^{-1}).$$

(d) Conduct a likelihood ratio test. Compute the value of the unrestricted maximised log likelihood function (l^U), that is with all 5 years of R&D expenditure included in x, and the value of the restricted maximised log likelihood function, that is with only one lagged value of R&D expenditure included. Then compute $2(l^U - l^R)$ which is a realisation (approximately) of a $\chi^2_{(4)}$ random variable when the null hypothesis is true. The 0.95 quantile of a $\chi^2_{(4)}$ random variable is 9.48, that is $P[\chi^2_{(4)} > 9.48] = 0.05$. So if the value obtained is greater than 9.48 the null hypothesis is rejected using a test with approximate size 0.05. A Wald or score test could be used instead.

(e) If Z has a Poisson distribution with parameter λ so that

$$P[Z = z] = \frac{\lambda^z \exp(-\lambda)}{z!}$$

then $E[Z] = Var[Z] = \lambda$. This is most easily shown by first deriving the moment generating function:

$$\begin{aligned}
E[\exp(tZ)] &= \sum_{z=0}^{\infty} \frac{e^{tz}\lambda^z \exp(-\lambda)}{z!} \\
&= \sum_{z=0}^{\infty} \frac{(\lambda e^t)^z \exp(-\lambda)}{z!} \\
&= \exp(-\lambda)\exp(\lambda e^t) \sum_{z=0}^{\infty} \frac{(\lambda e^t)^z \exp(-\lambda e^t)}{z!} \\
&= \exp(\lambda\left(e^t - 1\right))
\end{aligned}$$

and then differentiating once and then twice and setting $t = 0$ to give $E[Z]$ and then $E[Z^2]$, finally using $Var[Z] = E[Z^2] - E[Z]^2$. At the last line above I have used $e^a = (1 + a + a^2/2! + \dots)$. When $\lambda(x) = x'\theta$

$$Y = E[Y|x] + \varepsilon = x'\theta + \varepsilon$$

where $E[\varepsilon|x] = 0$ and $Var[\varepsilon|x] = x'\theta$. So OLS of y on x produces an estimate of θ, but the ε's are heteroskedastic so GLS would produce a more accurate estimator if θ were known. Of course θ is not known but we can obtain a "first round" estimate using OLS and the apply the GLS formula with $\hat{\theta}$ used in place of θ. The linear model could lead to estimates of $E[Y|x]$ which are negative but Y is by definition positive.

4. *Likelihood ratio Wald and score tests.* As in Question 2, suppose unemployment durations, T, have an exponential distribution with probability density function

$$f_T(t;\theta) = \theta^{-1}\exp(t/\theta), \qquad t, \theta > 0.$$

You have independent realisations (perhaps from a survey of those leaving unemployment) of completed unemployment durations, $t_1, \dots, t_n$.

(a) The MLE is $\hat{\theta} = \bar{t}$ and the log likelihood function is

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

Substituting the MLE, $\bar{t}$ for θ gives the result. The restricted log likelihood is just

$$l(\theta^*, t) = -n\log\theta^* - n\bar{t}/\theta^*$$

and so the likelihood ratio statistic is

$$2\left(l(\hat{\theta}, t) - l(\theta^*, t)\right) = 2n\left(-\log\left(\bar{t}/\theta^*\right) + \bar{t}/\theta^* - 1\right)$$

that is, twice the difference between the maximised unrestricted and restricted log likelihood functions, which gives the desired answer on substituting $q = \bar{t}/\theta^*$.

(b) The information "matrix" for this problem is $I(\theta) = n/\theta^2$, so we have, with θ_0 denoting the data generating parameter value,

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

The Wald statistic is therefore

$$S_W = \frac{n(\hat{\theta} - \theta^*)^2}{\hat{\theta}^2} = n(1 - q^{-1})^2$$

where we have used $\hat{\theta} = \bar{t}$ and $q = \bar{t}/\theta^*$. Note that we have estimated the variance of the limiting distribution (θ_0^2) using the MLE of θ. The gradient of the log likelihood function evaluated at θ^* (the restricted "estimate") is

$$l_\theta(\theta^*, y) = -\frac{n}{\theta^*} + \frac{n\bar{t}}{(\theta^*)^2}.$$

We have $\bar{I}(\theta) = \theta^{-2}$. Therefore the score statistic is

$$\begin{aligned} S_S &= n^{-1} \frac{\left(-\frac{n}{\theta^*} + \frac{n\bar{t}}{(\theta^*)^2}\right)^2}{(\theta^*)^{-2}} \\ &= n\left(-1 + \frac{\bar{t}}{\theta^*}\right)^2 \\ &= n\,(1 - q)^2 \end{aligned}$$

where as before, $q = \bar{t}/\theta^*$.

5. (a) The results of the five OLS estimates are as follows:

```
. reg w1 lrp1-lrp5 lrfepc

  Source |       SS       df       MS                  Number of obs =     312
---------+------------------------------               F(  6,   305) =  142.99
   Model |  .10407053      6  .017345088               Prob > F      =  0.0000
Residual | .036996232    305  .000121299               R-squared     =  0.7377
---------+------------------------------               Adj R-squared =  0.7326
   Total | .141066763    311  .000453591               Root MSE      =  .01101

------------------------------------------------------------------------------
      w1 |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lrp1 |   .1076639   .0090212     11.935   0.000       .0899123    .1254155
    lrp2 |  -.0070751   .0086653     -0.816   0.415      -.0241264    .0099762
    lrp3 |   .0490431   .0073664      6.658   0.000       .0345478    .0635385
    lrp4 |  -.0406829   .0146007     -2.786   0.006      -.0694139    -.011952
    lrp5 |   .0259303   .0100763      2.573   0.011       .0061026    .0457581
  lrfepc |  -.0317888    .013986     -2.273   0.024        -.05931   -.0042675
   _cons |   .9168859   .0789228     11.618   0.000       .7615839    1.072188
------------------------------------------------------------------------------
```

```
. reg w2 lrp1-lrp5 lrfepc

  Source |       SS       df       MS                  Number of obs =     312
---------+------------------------------               F(  6,   305) =  143.79
   Model |  .016345932     6  .002724322               Prob > F      =  0.0000
Residual |  .005778866   305  .000018947               R-squared     =  0.7388
---------+------------------------------               Adj R-squared =  0.7337
   Total |  .022124798   311  .000071141               Root MSE      =  .00435

------------------------------------------------------------------------------
      w2 |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lrp1 |   .0444591   .0035654     12.470   0.000       .0374432    .0514749
    lrp2 |  -.0153481   .0034247     -4.482   0.000      -.0220872    -.008609
    lrp3 |   .0135582   .0029114      4.657   0.000       .0078293    .0192871
    lrp4 |   .0270844   .0057706      4.694   0.000       .0157293    .0384395
    lrp5 |  -.0149917   .0039824     -3.765   0.000      -.0228281   -.0071553
  lrfepc |  -.0057722   .0055276     -1.044   0.297      -.0166492    .0051049
   _cons |   .3978725   .0311921     12.756   0.000       .3364935    .4592515
------------------------------------------------------------------------------

. reg w3 lrp1-lrp5 lrfepc

  Source |       SS       df       MS                  Number of obs =     312
---------+------------------------------               F(  6,   305) =   73.06
   Model |  .007156408     6  .001192735               Prob > F      =  0.0000
Residual |  .004979314   305  .000016326               R-squared     =  0.5897
---------+------------------------------               Adj R-squared =  0.5816
   Total |  .012135722   311  .000039022               Root MSE      =  .00404

------------------------------------------------------------------------------
      w3 |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lrp1 |    .039102   .0033096     11.815   0.000       .0325895    .0456144
    lrp2 |   .0001432    .003179      0.045   0.964      -.0061123    .0063987
    lrp3 |   .0065459   .0027025      2.422   0.016       .0012281    .0118638
    lrp4 |    -.01486   .0053565     -2.774   0.006      -.0254003   -.0043196
    lrp5 |  -.0054198   .0036966     -1.466   0.144      -.0126939    .0018543
  lrfepc |  -.0142447    .005131     -2.776   0.006      -.0243412   -.0041481
   _cons |   .1638071    .028954      5.657   0.000       .1068323     .220782
------------------------------------------------------------------------------

. reg w4 lrp1-lrp5 lrfepc

  Source |       SS       df       MS                  Number of obs =     312
---------+------------------------------               F(  6,   305) =    7.94
   Model |  .002100569     6  .000350095               Prob > F      =  0.0000
Residual |  .013453346   305  .000044109               R-squared     =  0.1351
---------+------------------------------               Adj R-squared =  0.1180
   Total |  .015553915   311  .000050013               Root MSE      =  .00664
```

```
------------------------------------------------------------------------------
      w4 |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lrp1 |   .0131749     .00544      2.422   0.016        .0024702    .0238796
    lrp2 |  -.0047368   .0052254     -0.906   0.365       -.0150192    .0055456
    lrp3 |   .0022634   .0044421      0.510   0.611       -.0064777    .0110045
    lrp4 |   .0182984   .0088046      2.078   0.039        .0009729    .0356239
    lrp5 |  -.0057454   .0060762     -0.946   0.345       -.0177021    .0062113
  lrfepc |  -.0388643   .0084339     -4.608   0.000       -.0554603   -.0222682
   _cons |   .2458552   .0475925      5.166   0.000        .1522039    .3395064
------------------------------------------------------------------------------

. reg w5 lrp1-lrp5 lrfepc

  Source |       SS       df       MS                  Number of obs =     312
---------+------------------------------               F(  6,   305) =   42.82
   Model |  .003616578     6  .000602763               Prob > F      =  0.0000
Residual |  .004293029   305  .000014076               R-squared     =  0.4572
---------+------------------------------               Adj R-squared =  0.4466
   Total |  .007909608   311  .000025433               Root MSE      =  .00375

------------------------------------------------------------------------------
      w5 |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    lrp1 |   .0016169    .003073      0.526   0.599       -.0044301    .0076639
    lrp2 |  -.0027713   .0029518     -0.939   0.349       -.0085797    .0030372
    lrp3 |  -.0077061   .0025093     -3.071   0.002       -.0126439   -.0027683
    lrp4 |  -.0185639   .0049737     -3.732   0.000       -.0283509   -.0087768
    lrp5 |   .0329317   .0034324      9.594   0.000        .0261775     .039686
  lrfepc |  -.0314725   .0047643     -6.606   0.000       -.0408475   -.0220975
   _cons |   .0060247   .0268847      0.224   0.823       -.0468783    .0589277
------------------------------------------------------------------------------
```

(b) The estimated standard error is 0.0032 using the conventional OLS formula and 0.0034 using the Eicker-White formula. The own price coefficinet for fish is 0.0016. Approximate 95% confidence intervals are: standard - $[-0.0047, 0.0079]$, robust - $[-0.0050, 0.0083]$. The results are very similar. If the errors in the equation were heteroskedastic then large differences could arise.

(c) Note that

$$w_i = \frac{p_i \cdot q_i}{x}$$

and so

$$\log q_i = \log w_i - \log p_i + \log x.$$

Taking the partial derivative with respect to $\log p_j$ there is

$$\begin{aligned}\frac{\partial \log q_i}{\partial \log p_j} &= \frac{\partial \log w_i}{\partial \log p_j} - \delta_{ij} \\ &= \frac{1}{w_i}\frac{\partial w_i}{\partial \log p_j} - \delta_{ij} \\ &= \frac{\gamma_{ij}}{w_i} - \delta_{ij}.\end{aligned}$$

The estimated cross elasticity has the form

$$\frac{\hat{\gamma}_{ij}}{w_i}$$

so the approximate standard error is the approximate standard error of $\hat{\gamma}_{ij}$ devided by w_i. The estimated own elasticity has the form

$$\frac{\hat{\gamma}_{ii}}{w_i} - 1$$

so the approximate standard error is the approximate standard error of $\hat{\gamma}_{ii}$ divided by w_i. The choice of food share can have a substantial effect on an estimated elasticity, unless of course the price coefficient is close to zreo. Note that this is a consequence of the choice of functional form.

(d) This is the endogeneity issue discussed in the lectures and course notes. An instrumental variable correlated with price and uncorrelated with demand would be useful. A variable measuring some aspect of producer's costs might be suitable.

(e) Again, note that

$$\log q_i = \log w_i - \log p_i + \log x.$$

The partial derivative with respect to $\log x$ is:

$$\begin{aligned}\frac{\partial \log q_i}{\partial \log x} &= \frac{\partial \log w_i}{\partial \log x} + 1 \\ &= \frac{1}{w_i}\frac{\partial w_i}{\partial \log x} + 1 \\ &= \frac{\beta_i}{w_i} + 1.\end{aligned}$$

The estimated total expenditure elasticity is

$$\frac{\hat{\beta}_i}{w_i} + 1$$

and hence the approximate standard error is the approximate standard error of $\hat{\beta}_i$ divided by w_i.

(f)

(i) To examine seasonality, one can include monthly indicator variables. The results indicate that the coefficients on the dummy variables are jointly significant (there is higher consumption during the winter than during the spring and summer) but that the maginitudes of the differenecs across months are very small..

(ii) Score tests for heteroskedasticity suggest there is heteroskedastic variation.

(iv) The "Chow test" gives a highly significant result - a test statistic of around 60 compared with a $\chi^2_{(7)}$ 0.95 quantile of 14.1.

(g) Standard errors are much larger with IV as is to be expected. IV and OLS estimates are different.