Empirical Finance: Study Questions Week 4
Imperial College London
Hormoz Ramian

Question 1 Suppose that de-meaned outcome y_i and explanatory variables are linearly modelled, for $i = 1, \ldots, N$ such that $u_i | x_{1,i},\ x_{2,i} \sim \mathcal{N}(0, \sigma^2)$ and:

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

(1.1) Show the implications of mutually orthogonal regressors condition on efficiency of least squares estimator for β_2 when $\boldsymbol{x}_1$ is unobservable and excluded from equation (1).

(1.2) Under what condition the variance of regression in equation (1) (strictly) decreases relative to the case without x_{2i}?

Question 2: Consider the system of two simultaneous equations,

$$y_{i1} = y_{i2}\alpha_1 + x_i\beta_1 + u_{i1} \tag{2}$$

$$y_{i2} = y_{i1}\alpha_1 + w_i\beta_1 + u_{i2} \tag{3}$$

where y_{i1} and y_{i2} are endogenous variables, and x_i and w_i are two exogenous regressors. There are four scalar structural parameters α_1, α_2, β_1 and β_2. The reduced form equations are,

$$y_{i1} = x_i\pi_{11} + w_i\pi_{21} + \epsilon_{i1} \tag{4}$$

$$y_{i2} = x_i\pi_{12} + w_i\pi_{22} + \epsilon_{i2} \tag{5}$$

where π_{11}, π_{12}, π_{21} and π_{22} are the reduced form parameters.

(2.1) Assume that $\alpha_1\alpha_2 \neq 1$. Find expressions for the reduced form parameters in terms of the structural parameters.

(2.2) Assume that $\pi_{11} \neq 0$ and $\pi_{22} \neq 0$. Show that all structural parameters are identified. Find expressions for the structural parameters in terms of the reduced form parameters.

(2.3) Consider estimation of α_1 and β_1 by applying 2SLS to the first structural equation,using w_i as an instrument for y_{i2}. Why is the condition $\pi_{22} \neq 0$ important for this 2SLS estimation? Can the parameters α_1 and β_1 be consistently estimated when π_{22}?

Question 3 Suppose that the dependent variable y_t is a proportion, so that $0 < y_t < 1$, $t = 1, \ldots, T$. An appropriate model for such a dependent variable is:

$$\log\left(\frac{y_t}{1 - y_t}\right) = \boldsymbol{X}_t\boldsymbol{\beta} + u_t \tag{6}$$

where $\boldsymbol{X}_t$ and $\boldsymbol{\beta}$ are k-dimensional vectors of exogenous variables and regression parameters, respectively. Write down the loglikelihood function for this model under the assumption $u_t \sim i.i.d.\mathcal{N}(0, \sigma^2)$.

Question 4 Consider the following a binary response model

$$y^* = \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e \tag{7}$$

where e is a residual, assumed uncorrelated with x (i.e. x is not endogenous). While we do not observe y, we do observe the discrete choice made by the individual, according to the following choice rule:

$$y = \begin{cases} 1 & if \quad y^* > 0 \\ 0 & if \quad y^* \leq 0 \end{cases} \tag{8}$$

assume

$$\begin{aligned} \mathbb{P}(y = 1|\boldsymbol{x}) &= G(\beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k) \\ &= G(\boldsymbol{x\beta}) \end{aligned} \tag{9}$$

where $G(.)$ can be a logit or probit function and derive the first order condition w.r.t. to $\boldsymbol{\beta}$ and interpret each parameter when the explanatory variables are continuous and when they are discrete.

Answers

Question 1:

(1.1) When $\boldsymbol{x}_1$ is omitted we have,

$$\begin{aligned} \mathbb{E}[\widehat{\boldsymbol{\beta}}_2] &= \mathbb{E}\left\{\mathbb{E}\left[(\boldsymbol{x}_2'\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2'\boldsymbol{y}|\boldsymbol{x}\right]\right\} \\ &= \mathbb{E}\left\{\mathbb{E}\left[(\boldsymbol{x}_2'\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2'\left\{\boldsymbol{x}_1\boldsymbol{\beta}_1 + \boldsymbol{x}_2\boldsymbol{\beta}_2 + \boldsymbol{u}\right\}|\boldsymbol{x}\right]\right\} \\ &= \boldsymbol{\beta}_2 + \mathbb{E}\left\{\mathbb{E}\left[(\boldsymbol{x}_2'\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2'\boldsymbol{x}_1\boldsymbol{\beta}_1|\boldsymbol{x}\right]\right\} \end{aligned}$$

which is unbiased[1] iff $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthogonal (or if $\boldsymbol{\beta}_1 = \boldsymbol{0}$). However, even in the case of orthogonality, efficiency is not guaranteed. Denote the vector of residuals from the regression with omitted variable with $\boldsymbol{\varepsilon}$,

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{x}_2\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{M}_2\boldsymbol{y} \tag{10}$$

Computing regression (omitted $\boldsymbol{x}_1$) variance term $\widetilde{s}^2$:

$$\begin{aligned} \widetilde{s}^2 &= \frac{1}{N-k}[\boldsymbol{M}_2\boldsymbol{y}]'[\boldsymbol{M}_2\boldsymbol{y}] \\ &= \frac{1}{N-k}(\boldsymbol{M}_2\boldsymbol{x}_1\boldsymbol{\beta}_1 + \boldsymbol{M}_2\boldsymbol{u})'(\boldsymbol{M}_2\boldsymbol{x}_1\boldsymbol{\beta}_1 + \boldsymbol{M}_2\boldsymbol{u}) \\ &= \frac{1}{N-k}\left(\boldsymbol{u}'\boldsymbol{M}_2\boldsymbol{u} + \boldsymbol{\beta}_1'\boldsymbol{x}_1'\boldsymbol{M}_2\boldsymbol{x}_1\boldsymbol{\beta}_1\right) \end{aligned}$$

noting that cross-term $2\boldsymbol{u}'\boldsymbol{M}_2\boldsymbol{x}_1\boldsymbol{\beta}_1$ is zero and that $\boldsymbol{u}'\boldsymbol{M}_2\boldsymbol{u}$ is the RSS from the correctly specified model:

$$\widetilde{s}^2 = \sigma^2 + \frac{1}{N-k}\boldsymbol{\beta}_1'\boldsymbol{x}_1'\boldsymbol{M}_2\boldsymbol{x}_1\boldsymbol{\beta}_1$$

when regressors are orthogonal $\boldsymbol{x}_1'\boldsymbol{M}_2\boldsymbol{x}_1 = \boldsymbol{x}_1'(\boldsymbol{I} - \boldsymbol{x}_2(\boldsymbol{x}_2'\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2')\boldsymbol{x}_1 = \boldsymbol{x}_1'\boldsymbol{x}_1$:

$$\widetilde{s}^2 = \sigma^2 + \frac{1}{N-k}\boldsymbol{\beta}_1'\boldsymbol{x}_1'\boldsymbol{x}_1\boldsymbol{\beta}_1$$

where the second term is strictly positive if $\boldsymbol{\beta}_1 \neq 0$ requiring stronger conditions relative to unbiasedness.

(1.2) Denote the vector of residuals from the regression with one regressor with $\boldsymbol{e}_1$ and that of the

[1] Although this is in general biased, but knowledge about $\boldsymbol{x}_2'\boldsymbol{x}_1$ and $\boldsymbol{\beta}_1$ can be useful information to determine the direction of the bias, e.g. if unobservable $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are positively (negatively) correlated and $\boldsymbol{x}_1$ and $\boldsymbol{y}$ are positively (negatively) correlated then $\boldsymbol{\beta}_2$ is overestimated. Alternatively, if the relationships between regressors and the unobservable and the outcome variable are of the opposite directions, then the estimator for $\boldsymbol{\beta}_2$ is underestimated.

full regression with $\boldsymbol{e}_2$, and respective variances with s_1^2 and s_2^2:

$$\begin{aligned}\boldsymbol{e}_2 &= \boldsymbol{y} - \boldsymbol{x}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{x}_2\widehat{\boldsymbol{\beta}}_2 &(11)\\ \boldsymbol{M}_1\boldsymbol{e}_2 &= \boldsymbol{M}_1(\boldsymbol{y} - \boldsymbol{x}_2\widehat{\boldsymbol{\beta}}_2)\\ &= \boldsymbol{M}_1(\boldsymbol{I} - \boldsymbol{x}_2(\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2'\boldsymbol{M}_1)\boldsymbol{y} &(12)\end{aligned}$$

then,

$$\begin{aligned}\boldsymbol{e}_2'\boldsymbol{e}_2 &= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{x}_2(\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2'\boldsymbol{M}_1)'\boldsymbol{M}_1(\boldsymbol{I} - \boldsymbol{x}_2(\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2'\boldsymbol{M}_1)\boldsymbol{y}\\ &= \boldsymbol{y}'\boldsymbol{M}_1\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{M}_1\boldsymbol{x}_2(\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{y}\\ &= \boldsymbol{e}_1'\boldsymbol{e}_1 - \boldsymbol{y}'\boldsymbol{M}_1\boldsymbol{x}_2(\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{y} &(13)\end{aligned}$$

Comparing variance requires the following conjecture. Suppose, $s_1^2 > s_2^2$ then:

$$\frac{1}{N-k}\boldsymbol{e}_1'\boldsymbol{e}_1 > \frac{1}{N-k-1}\boldsymbol{e}_2'\boldsymbol{e}_2 \qquad (14)$$

substituting for $\boldsymbol{e}_1'\boldsymbol{e}_1$ from equation (13),

$$\begin{aligned}\frac{N-k-1}{N-k}\frac{\boldsymbol{e}_2'\boldsymbol{e}_2 + \boldsymbol{y}'\boldsymbol{M}_1\boldsymbol{x}_2(\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x}_2)^{-1}\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{y}}{\boldsymbol{e}_2'\boldsymbol{e}_2} &> 1\\ \frac{N-k-1}{N-k}\frac{\boldsymbol{e}_2'\boldsymbol{e}_2 + \widehat{\boldsymbol{\beta}}_2'\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x}_2\widehat{\boldsymbol{\beta}}_2}{\boldsymbol{e}_2'\boldsymbol{e}_2} &> 1\\ \frac{\boldsymbol{e}_2'\boldsymbol{e}_2 + \widehat{\boldsymbol{\beta}}_2'\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x}_2\widehat{\boldsymbol{\beta}}_2}{\boldsymbol{e}_2'\boldsymbol{e}_2 + \boldsymbol{e}_2'\boldsymbol{e}_2/(N-k-1)} &> 1\\ \frac{\boldsymbol{e}_2'\boldsymbol{e}_2 + \widehat{\boldsymbol{\beta}}_2'\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x}_2\widehat{\boldsymbol{\beta}}_2}{\boldsymbol{e}_2'\boldsymbol{e}_2 + s_2^2} &> 1\end{aligned}$$

For simplicity suppose $\widehat{\boldsymbol{\beta}}_2$ is a scalar then,

$$\frac{\widehat{\beta}_2^2}{s_2^2/([\boldsymbol{M}_1\boldsymbol{x}_2]'[\boldsymbol{M}_1\boldsymbol{x}_2])} > 1$$

indicating that the associated (squared) t-ratio of the additional regressor has to be above one in order to have lower variance (of extended regression).

Question 2:

(2.1) Vectorizing the structural and reduced form equations, respectively, gives:

$$\begin{aligned}\begin{bmatrix}1 & -\alpha_1\\ -\alpha_2 & 1\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} &= \begin{bmatrix}\beta_1 & 0\\ 0 & \beta_2\end{bmatrix}\begin{bmatrix}x\\ w\end{bmatrix} + \begin{bmatrix}u_1\\ u_2\end{bmatrix} &(15)\\ \boldsymbol{A}\boldsymbol{y} &= \boldsymbol{B}\boldsymbol{z} + \boldsymbol{u} &(16)\end{aligned}$$

and

$$\begin{aligned}\begin{bmatrix}y_1\\ y_2\end{bmatrix} &= \begin{bmatrix}\pi_{11} & \pi_{21}\\ \pi_{12} & \pi_{22}\end{bmatrix}\begin{bmatrix}x\\ w\end{bmatrix} + \begin{bmatrix}\epsilon_1\\ \epsilon_2\end{bmatrix} &(17)\\ \boldsymbol{y} &= \boldsymbol{\Pi}\boldsymbol{z} + \boldsymbol{\epsilon} &(18)\end{aligned}$$

re-arranging equations (16) and (18) requires $\boldsymbol{\Pi}' = \boldsymbol{B}'(\boldsymbol{A}^{-1})'$:

$$\boldsymbol{\Pi}' = \begin{bmatrix}\beta_1 & 0\\ 0 & \beta_2\end{bmatrix} \times \frac{1}{1-\alpha_1\alpha_2}\begin{bmatrix}1 & \alpha_1\\ \alpha_2 & 1\end{bmatrix}' \qquad (19)$$

thus

$$
\begin{aligned}
\pi_{11} &= (1-\alpha_1\alpha_2)^{-1}\beta_1 \\
\pi_{12} &= (1-\alpha_1\alpha_2)^{-1}\alpha_2\beta_1 \\
\pi_{21} &= (1-\alpha_1\alpha_2)^{-1}\alpha_1\beta_2 \\
\pi_{22} &= (1-\alpha_1\alpha_2)^{-1}\beta_2
\end{aligned}
$$

(2.2) Solving for the structural parameters gives:

$$
\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{\pi_{21}}{\pi_{22}} \\ \frac{\pi_{12}}{\pi_{11}} \end{bmatrix}, \quad \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = (\pi_{11}\pi_{22} - \pi_{12}\pi_{21}) \begin{bmatrix} \frac{1}{\pi_{22}} \\ \frac{1}{\pi_{11}} \end{bmatrix} \tag{20}
$$

(2.3) The relevance condition (w vs. y) depends on $\pi_{22} \neq 0$ that is to say α_1 and β_1 are not identified if $\pi_{22} = 0$.

Question 3 If $u_t \sim \mathcal{N}$ with zero mean and variance σ^2, then the density of u_t is:

$$
f(u_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u_t^2}{2\sigma^2}} \tag{21}
$$

We need to construct the density of y_t to set up the ML objective function. However, note that y_t is a nonlinear function of u_t

$$
\log\left(\frac{y_t}{1-y_t}\right) = \ldots + u_t \tag{22}
$$

In this problem, first we substitute instead of u_t in the density of u_t:

$$
f\left[\log\left(\frac{y_t}{1-y_t}\right) - \boldsymbol{X}_t\boldsymbol{\beta}\right] = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left\{\log\left(\frac{y_t}{1-y_t}\right) - \boldsymbol{X}_t\boldsymbol{\beta}\right\}^2}{2\sigma^2}} \tag{23}
$$

Second, we need to work out an additional term called the Jacobian term:

$$
\frac{d}{dy_t}\log\left(\frac{y_t}{1-y_t}\right) = \frac{d}{dy_t}(\log(y_t) - \log(1-y_t)) = \frac{1}{y_t(1-y_t)} \tag{24}
$$

The density of y_t is constructed using:

$$
\begin{aligned}
\text{density}(y_t) &= f\left[\log\left(\frac{y_t}{1-y_t}\right) - \boldsymbol{X}_t\boldsymbol{\beta}\right] \times \left|\frac{d}{dy_t}\log\left(\frac{y_t}{1-y_t}\right)\right| && (25) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left\{\log\left(\frac{y_t}{1-y_t}\right) - \boldsymbol{X}_t\boldsymbol{\beta}\right\}^2}{2\sigma^2}} \times \frac{1}{y_t(1-y_t)} && (26)
\end{aligned}
$$

where |...| is the absolute value operator. We can now take log of the last expression and setup the log-likelihood function $\ell(\boldsymbol{y};\boldsymbol{\beta},\sigma)$ or $\ell(\boldsymbol{y};\boldsymbol{\beta},\sigma^2)$. Hence the log-likelihood function is:

$$
\ell(\boldsymbol{y},\boldsymbol{\beta},\sigma) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \sum_{t\in T}\log y_t - \sum_{t\in T}\log(1-y_t) - \frac{1}{2\sigma^2}\sum_{t\in T}(\log(y_t/(1-y_t))\boldsymbol{X}_t\boldsymbol{\beta})^2
$$

Noting that $-\frac{n}{2}\log(2\pi)$ is just a constant and inconsequential through the maximization, then:

$$
\ell(\boldsymbol{y},\boldsymbol{\beta},\sigma) \propto -\frac{n}{2}\log\sigma^2 - \sum_{t\in T}\log y_t - \sum_{t\in T}\log(1-y_t) - \frac{1}{2\sigma^2}\sum_{t\in T}(\log(y_t/(1-y_t))\boldsymbol{X}_t\boldsymbol{\beta})^2
$$

Question 4

(4.1) Consider a binary response model of the form

$$
\begin{aligned}
\mathbb{P}(y=1|\boldsymbol{x}) &= G(\beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k) \\
&= G(\boldsymbol{x}\boldsymbol{\beta}) && (27)
\end{aligned}
$$

where G is a function taking on values strictly between zero and one $0 < G(.) < 1$ for all real numbers to ensure that the estimated response probabilities are strictly between zero and one. G is sometimes chosen to be a cumulative density function (cdf), monotonically increasing in the arguments. A corollary to this statement is that $G(.)$ must be a non-linear function, and hence we cannot use the OLS. A logic model assume the following expression:

$$G(\boldsymbol{x\beta}) = \frac{\exp(\boldsymbol{x\beta})}{1+\exp(\boldsymbol{x\beta})} =: \Lambda(\boldsymbol{x\beta}) \tag{28}$$

which is between zero and one as this is the cumulative distribution function (CDF) for a logistic variable. In the probit model, G is the standard normal CDF, expressed as an integral:

$$G(\boldsymbol{x\beta}) = \Phi(\boldsymbol{x\beta}) = \int_{-\inf}^{\boldsymbol{x\beta}} \phi(v)dv \tag{29}$$

$$\phi(v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) \tag{30}$$

where $\phi(.)$is the standard normal density. This choice of G also ensures that the probability of success is strictly between zero and one for all values of the parameters and the explanatory variables. Probits and logits are traditionally viewed as models suitable for estimating parameters of interest when the dependent variable is not fully observed:

$$y^* = \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e \tag{31}$$

where e is a residual, assumed uncorrelated with x (i.e.x is not endogenous). While we do not observe y, we do observe the discrete choice made by the individual, according to the following choice rule:

$$y = \begin{cases} 1 & if \;\; y^* > 0 \\ 0 & if \;\; y^* \leq 0 \end{cases} \tag{32}$$

Think about y as representing net utility of, say, buying a car. The individual undertakes a cost-benefit analysis and decides to purchase the car if the net utility is positive. We do not observe (because we cannot measure) the amount of net utility; all we observe is the actual outcome of whether or not the individual does buy a car. We want to model the probability that a positive choice is made (e.g. buying, as distinct from not buying, a car):

$$\begin{aligned} \mathbb{P}[y=1|\boldsymbol{x}] &= \mathbb{P}[y^* > 0|\boldsymbol{x}] \\ &= \mathbb{P}[\boldsymbol{x\beta} + e > 0|\boldsymbol{x}] \\ &= \mathbb{P}[e > -\boldsymbol{x\beta}|\boldsymbol{x}] \\ &= 1 - \Lambda(-\boldsymbol{x\beta}) \\ &= \Lambda(\boldsymbol{x\beta}) \end{aligned}$$

Notice that the last step here exploits the fact that the logistic distribution is symmetric, so that $G(z) = 1 - G(z)$ for all z. This equation is the binary response model for the logit model showing how the response can be derived from an underlying latent variable. The associated likelihood function is:

$$\begin{aligned} L(y|\boldsymbol{x};\boldsymbol{\beta}) &= \prod_{i\in I_1} G(\boldsymbol{x\beta}) \prod_{i\in I_0} [1 - G(\boldsymbol{x\beta})] \\ &= \prod_{i}^{N} [G(\boldsymbol{x\beta})]^{y_i} [1 - G(\boldsymbol{x\beta})]^{1-y_i} \end{aligned} \tag{33}$$

where I_0 refers to observations for which $y = 0$ and I_1 to those with $y = 1$. We can take log from both sides of this equation to write (when logit $G(.) = \Lambda(.)$):

$$\begin{aligned} \ell \equiv \log L(y|\boldsymbol{x};\boldsymbol{\beta}) &= \sum_{i=1}^{N} \{y_i \log [G(\boldsymbol{x\beta})] + (1-y_i)\log[1-G(\boldsymbol{x\beta})]\} \\ &= \sum_{i=1}^{N} \left\{ y_i \log \left[\frac{\exp(\boldsymbol{x\beta})}{1+\exp(\boldsymbol{x\beta})} \right] + (1-y_i)\log \left[\frac{1}{1+\exp(\boldsymbol{x\beta})} \right] \right\} \quad (34) \\ &= \sum_{i=1}^{N} \{y_i.\boldsymbol{x\beta} - y_i \log [1+\exp(\boldsymbol{x\beta})] - (1-y_i)\log[1+\exp(\boldsymbol{x\beta})]\} \quad (35) \end{aligned}$$

and if G is the standard normal CDF we get the probit estimator (when logit $G(.) = \Phi(.)$):

$$= \sum_{i=1}^{N} \{y_i \log \Phi(\boldsymbol{x\beta}) + (1-y_i)\log[1-\Phi(\boldsymbol{x\beta})]\} \quad (36)$$

differentiating w.r.t. $\boldsymbol{\beta}$:

$$\mathbf{0}_{k\times 1} = \sum_{i=1}^{N} \left\{ y_i \frac{g(\boldsymbol{x\beta})}{G(\boldsymbol{x\beta})} + (1-y_i)\frac{g(\boldsymbol{x\beta})}{1-G(\boldsymbol{x\beta})} \right\} \quad (37)$$

It is typically not possible to solve analytically for $\boldsymbol{\beta}$, instead, to obtain parameter estimates, we rely on some numerical techniques (most common ones are based on first and sometimes second derivatives of the log likelihood function).

(4.2) In linear models the marginal effect of a unit change in some explanatory variable on the dependent variable is simply the associated coefficient on the relevant explanatory variable. However, for logit and probit models obtaining measures of the marginal effect is more complicated. When x is a *continuous* variable, its partial effect on $\mathbb{P}[y = 1|\boldsymbol{x}]$ is obtained from the partial derivative:

$$\frac{\partial \mathbb{P}[y=1|\boldsymbol{x}]}{\partial x_j} = \frac{\partial G(\boldsymbol{x\beta})}{\partial x_j} = g(\boldsymbol{x\beta}).\beta_j \quad (38)$$

where $g(.) \equiv \partial G(z)/\partial z$ is the associated probability density function ($G(.)$ is the probability distribution function). Because the density function is non-negative, the partial effect of x_j will always have the same sign as β_j. Notice that the partial effect depends on $g(.)$, i.e. for different values of x's the partial effect is different.

If x_j is a *discrete* variable then we should not rely on calculus in evaluating the effect on the response probability. To keep things simple, suppose x_2 is binary. In this case the partial effect from changing x_2 from zero to one, holding all other variables fixed is:

$$G(\beta_1 + \beta_2 \times 1 + \ldots + \beta_k x_k) - G(\beta_1 + \beta_2 \times 0 + \ldots + \beta_k x_k) \quad (39)$$

Similarly, this depends on all the values of the other explanatory variables and the values of all the other coefficients. Also, knowing the sign of β_2 is sufficient for determining whether the effect is positive or not, but to find the magnitude (and examining significance) of the effect we have to use the formula above.