Question 4. (25 marks) The researcher investigates the effect of having vocational training available in high school on the probability of currently living in poverty for the population of men who grew up with disadvantaged background. Let *pov* be a dummy variable equal to one if a man is currently living below the poverty line, and zero otherwise. The variable *age* is age and *edu* is total years of schooling. Let *voc* be an indicator equal to unity if a man's high school offered vocational training. Using a random sample of 850 men, the researcher obtains

$$\hat{\Pr}(pov = 1 \mid age, edu, voc) = F(0.453 - 0.016age - 0.087edu - 0.149voc) \qquad \textbf{(1)}$$

where $F(z) = \dfrac{\exp(z)}{(1+\exp(z))}$ is the logit function.

(a) (10 marks) □ Why is model (1) estimated by maximum likelihood and not OLS? Explain the meaning of the maximum likelihood method. What properties do estimates obtained by the maximum likelihood method have?

The logit model includes a non-linear transformation $F(z) = \dfrac{\exp(z)}{(1+\exp(z))}$ and is therefore non-linear in parameters, so using OLS is not possible. An alternative to the logit model is the linear probability model LPM, which can be estimated by OLS, but this introduces a number of problems, which we will discuss next.
Let us denote dependent variable by $X = (age, edu, voc) = F(0.453 - 0.016age - 0.087edu - 0.149voc)$ and denote by $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ the vector of model parameters $z = \beta_0 + \beta_1 age + \beta_2 edu + \beta_3 voc$.

The idea of Maximum Likelihood Method is to find the unknown values of parameters $\beta_0, \beta_1, \beta_2, \beta_3$ that maximize the probability of obtaining a sample that we actually have $P(X_1, ..., X_{850}) = L(\beta \mid, X_1, ..., X_{850}))$.
To do this one need to take the likelihood function as a product of the probability density function for all sample values

$$L(\beta \mid, X_1, ..., X_{850}) = f(\beta \mid, X_1) \cdot f(\beta \mid, X_2) \cdot ... \cdot f(\beta \mid, X_{850})$$

where $f(\beta \mid, x) = f(x)$ is the probability density function. For reasons of simplification of taking the derivatives in order to maximize the likelihood function we usually instead take the log-likelihood function $\log L(\beta \mid, X_1, ..., X_{850}) = \log f(\beta \mid, X_1) + \log f(\beta \mid, X_2) + ... + \log f(\beta \mid, X_{850})$.
The maximum likelihood function usually gives more **efficient estimators** then OLS when GM conditions are violated. In particular, MLE is **not very sensitive to the heteroscedasticity** inherent in binary choice models.
The estimators are generally **asymptotically consistent,** so it is large sample method, for finite samples it could give biased estimate.
In the case of normal distribution ML method gives asymptotically the same estimates of parameters as LS method. In estimation of simple linear regression coefficients under assumption of normal distribution of the disturbance term the ML estimators of regression coefficients are equal to the conventional OLS estimators, **but the estimations of the standard deviation of the disturbance term are different in two methods**.

□ Discuss the benefits/drawback of using the logit regression model when trying to explain a binary variable *pov*.

The logit model has two main advantages over the linear probability model (LPM): predicted probabilities are restricted to **lie in** $[0, 1]$ and MLE is (asymptotically) efficient whereas OLS (LPM) will be **inefficient** given the inherent presence of heteroskedasticity. Additionally the marginal effects of factors are not constant in logit model unlike LPM where they do not depend on the point chosen.
The main drawbacks of the logit model relative to the linear probability model are that the **coefficients cannot be directly interpreted as the marginal effects of the regressor(s)** of interest and it is also computationally more complicated. The estimations of marginal effects obtained by the direct calculation and by the derivative method can be slightly different.

□ Equation (1) contains information only about the estimated coefficients of the model. What additional information is needed to be able to judge the statistical quality of the econometric model (1)? What tests can be carried out for this purpose?

To assess the significance of factors, it is necessary to know the **standard errors** of the logit regression coefficients, while using normal distribution tables. Instead of conventional R-square, maximum likelihood models use **Mc-Fadden pseudo R-square**, which does not have the meaning of the proportion of the explained variance of the dependent variable. To assess the significance of the equation as a whole, information about the **LR statistics** is needed, its value is compared with **the critical chi-square value for three degrees of freedom**.

(b) (7 marks) □ Use the direct comparison of two probabilities of living in poverty calculated by logit function to evaluate effect of having vocational training available in high school for a 40-year old man, with 12 years of education. Give details and interpret the results.

The logistic function has the following expression

$$F(z) = \frac{1}{1+\exp(-z)} = \frac{\exp(z)}{1+\exp(z)}$$

where $z = 0.453 - 0.016age - 0.087edu - 0.149voc$.

For a 40-year old man with 12 years of education with vocational training the estimated probability of living in poverty is given by:

$$\Pr\overbrace{(y_i = 1 \mid age_i = 40, edu_i = 12, voc_i = 1)}^{\wedge} = \frac{1}{1+\exp(-z_1)} \approx 0.2010,$$

where $z_1 = 0.453 - 0.016 \times 40 - 0.087 \times 12 - 0.149 \times 1 \approx -1.38$.

The estimated probability of living in poverty for the same man without the vocational training is given by:

$$\Pr\overbrace{(y_i = 1 \mid age_i = 40, edu_i = 12, voc_i = 0)}^{\wedge} = \frac{\exp(z_2)}{1+\exp(z_2)} \approx 0.226,$$

where $z_2 = 0.453 - 0.016 \times 40 - 0.087 \times 12 - 0.149 \times 0 \approx -1.23$.

The difference between the two probabilities is $0.201 - 0.226 = -0.025$

Therefore, for a 40-year old man with 12 years of education having vocational training in high school decreases the probability of living in poverty by 2.5 percentage points.

(c) (8 marks) □ Now do the same estimation of marginal effect of vocational education as in (b) using derivatives.

The derivative for the logistic function is $\dfrac{\partial F(z)}{\partial z} = \dfrac{d}{dz}\left(\dfrac{1}{1+\exp(-z)}\right) = \dfrac{\exp(-z)}{(1+\exp(-z))^2} = \dfrac{\exp(z)}{(1+\exp(z))^2}$.

For a 40-year old man with 12 years of education without vocational training the estimated marginal effect of training is

$$\frac{\partial \Lambda(z)}{\partial E} = \frac{\partial \Lambda(z)}{\partial z} \cdot \frac{dz}{dE} = \frac{\exp(-z)}{(1+\exp(-z))^2} \cdot \beta_E = \frac{\exp(-z_2)}{(1+\exp(-z_2))^2} \cdot (-0.149) = 0.175 \cdot (-0.149) = -0.0261,$$

where $z_2 = 0.453 - 0.016 \times 40 - 0.087 \times 12 - 0.049 \times 0 \approx -1.23$.

Therefore, for a 40-year old man with 12 years of education having vocational training in high school decreases the probability of living in poverty by 2.6 percentage points, the result is close to that obtained in (b) 2.5 percentage points.

□ What percentage is the calculated marginal effect of the maximum possible?

The maximum possible effect is achieved at the point $z = 0$:

$$\left.\frac{\partial \Lambda(z)}{\partial E}\right|_{\max} = \frac{\exp(0)}{(1+\exp(0))^2} \cdot (-0.049) = 0.25 \cdot (-0.149) = -0.0373$$

The difference is considerable: the current marginal effect is 69,97% (about 70%) of maximum