

Microeconometrics 2018/2019 - 1st Assignment

20th of December, 2018

Instructor: Tomas Repasky

Group Members:

- Johannes Wagner
Humboldt-Universität Berlin - ID 598797 - Msc Statistics
wagnejoh@hu-berlin.de
- Malgorzata Paulina Olesiewicz
Humboldt-Universität Berlin - ID 598939 - Msc Statistics
malgorzata.paulina.olesiewicz@student.hu-berlin.de
- Kevin Hoppe
Humboldt-Universität Berlin - ID 598247 - Msc Statistics
Kevin.Hoppe1@web.de

Approximate individual contributions:

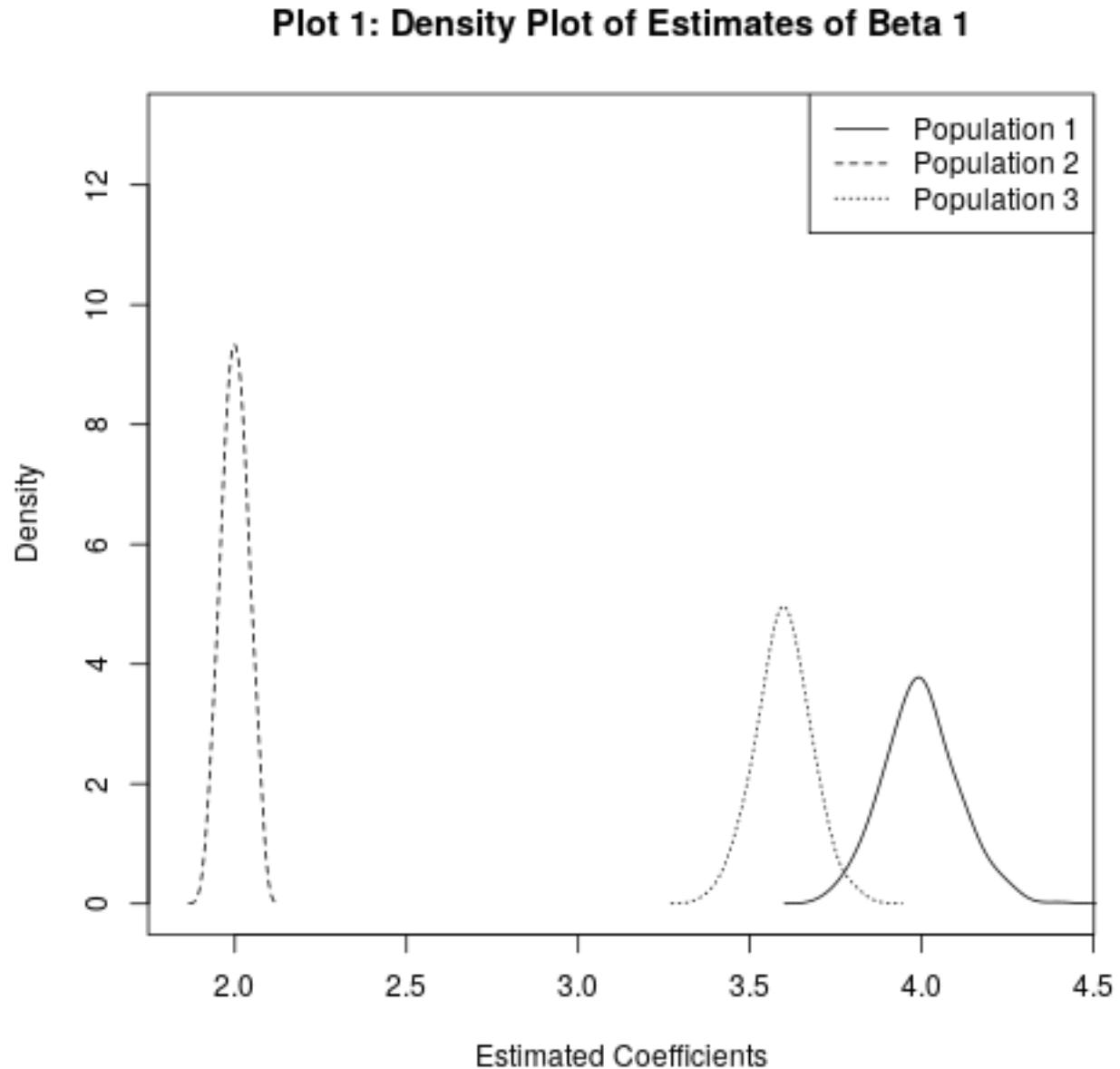
Task	Johannes	Malgorzata	Kevin
1.a	70% progr. + 50% theory	-	30% progr. + 50% theory
1.b	70% progr. + 70% theory	-	30% progr. + 30% theory
1.b i)	70% progr. + 30% theory	40% theory	30% progr. + 30% theory
1.b ii)	33% theory	33% theory	33% theory
1.b iii)	50% theory	-	50% theory
1.c	100% progr.	-	-
1.c i)	40% theory	50% theory	10% theory
1.c ii)	100% theory	-	-
1.c iii)	50% theory	50% theory	-
2.a	-	50% theory + 50% progr.	50% theory + 50% progr.
2.b	-	70% theory	30% theory
2.c	33% theory	33% theory + 50% progr.	33% theory + 50% progr.
2.d i)	-	40% progr.	60% progr.
2.d ii)	30% theory	40% theory	30% theory
2.d iii)	-	60% theory	40% theory
2.d iv)	30% theory	40% theory + 70% progr.	30% theory + 30% progr.
2.d v)	33% theory	33% theory	33% theory

First Problem: (Simulation: Latent Variable, Probit Model):

a) [1P]

The estimates for $\hat{\beta}_{1,1}$, $\hat{\beta}_{1,2}$ and $\hat{\beta}_{1,3}$ are 4.12, 2.03 and 3.69 respectively.

b) [1.5P]

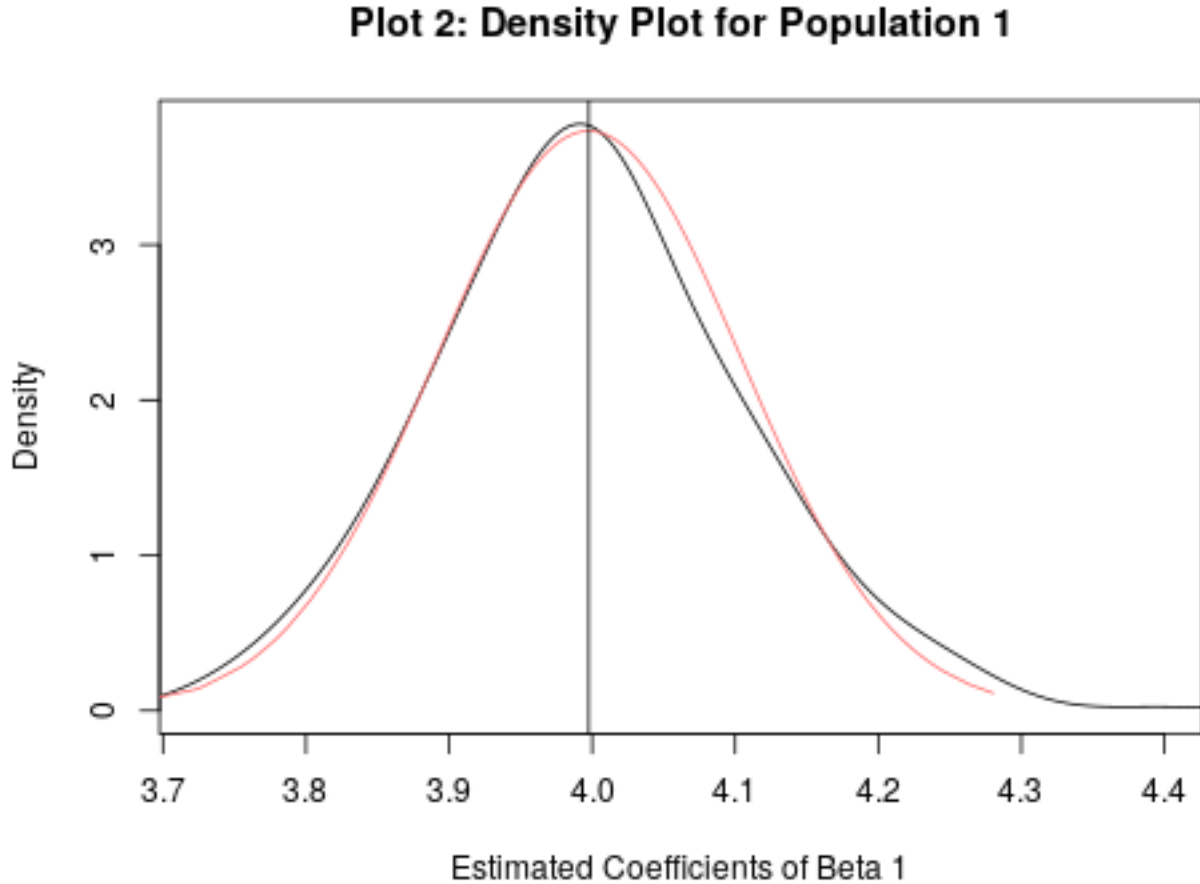


In order to interpret our distributions, we need to find a smooth representation for the probability density functions of our estimates. We can make use of a kernel estimator

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

where we choose K to be a *Gaussian Kernel*, which is the conventional choice and weighs observations x_i using a standard normal density function centered around x . That means that $K(\frac{x-x_i}{h}) = \phi(\frac{x-x_i}{h})$, where ϕ is the probability density function of the standard normal distribution. The smoothing parameter h is the bandwidth. It is chosen to be as small as possible to avoid loss of information and as big as necessary to avoid overfitting. We used “Silvermann’s *rule of thumb*” for the bandwidth (see ?density in R). It is the default setting and can be adjusted using a multiplicative parameter. We decided to increase the bandwidth by the factor 1.5 to get a more stable and informative distribution as with smaller values especially the distribution for population 1 appeared to be unstable.

i.) [1.5P]



Given the basic properties of an ML-estimator, we expect our estimates $\hat{\beta}_1$ to be consistent and asymptotically normally distributed.

Given our large sample, we expect the estimator to asymptotically converge in probability to the true parameter β_1 . In our probit model, we assume that the errors of the latent variable model are

a) independent of the explanatory variable x_1 and b) follow a standard normal distribution. We know from our parameters for the first population that those assumptions holds.

The plotted density confirms our expectation. We see that the plotted density of the estimates (black line in plot 2) approximately follows a normal distribution (indicated in red).

The estimates $\hat{\beta}_1$ for our first population are concentrated around their expected value of 4, which is equivalent to the true parameter β_1 in our latent model.

ii.) [1P]

As we can see in plot 1, all 3 distributions are approximately normal, but differ in their means and variances. For the means we find that $\hat{E}(\hat{\beta}_{1,1}) = 3.998$, $\hat{E}(\hat{\beta}_{1,2}) = 2.001$ and $\hat{E}(\hat{\beta}_{1,3}) = 3.598$.

Given the assumptions stated in i), we expect the estimate of $\hat{\beta}_{1,1}$ to be close to the true value of 4, as is the case. In the second population, the error variance is greater than one which results in a decreased estimate of $\hat{\beta}_{1,2}$ since the estimate is derived as being equal to β/σ with σ being the standard deviation of the error term (see iii)). The observed variance of the estimator is much smaller.

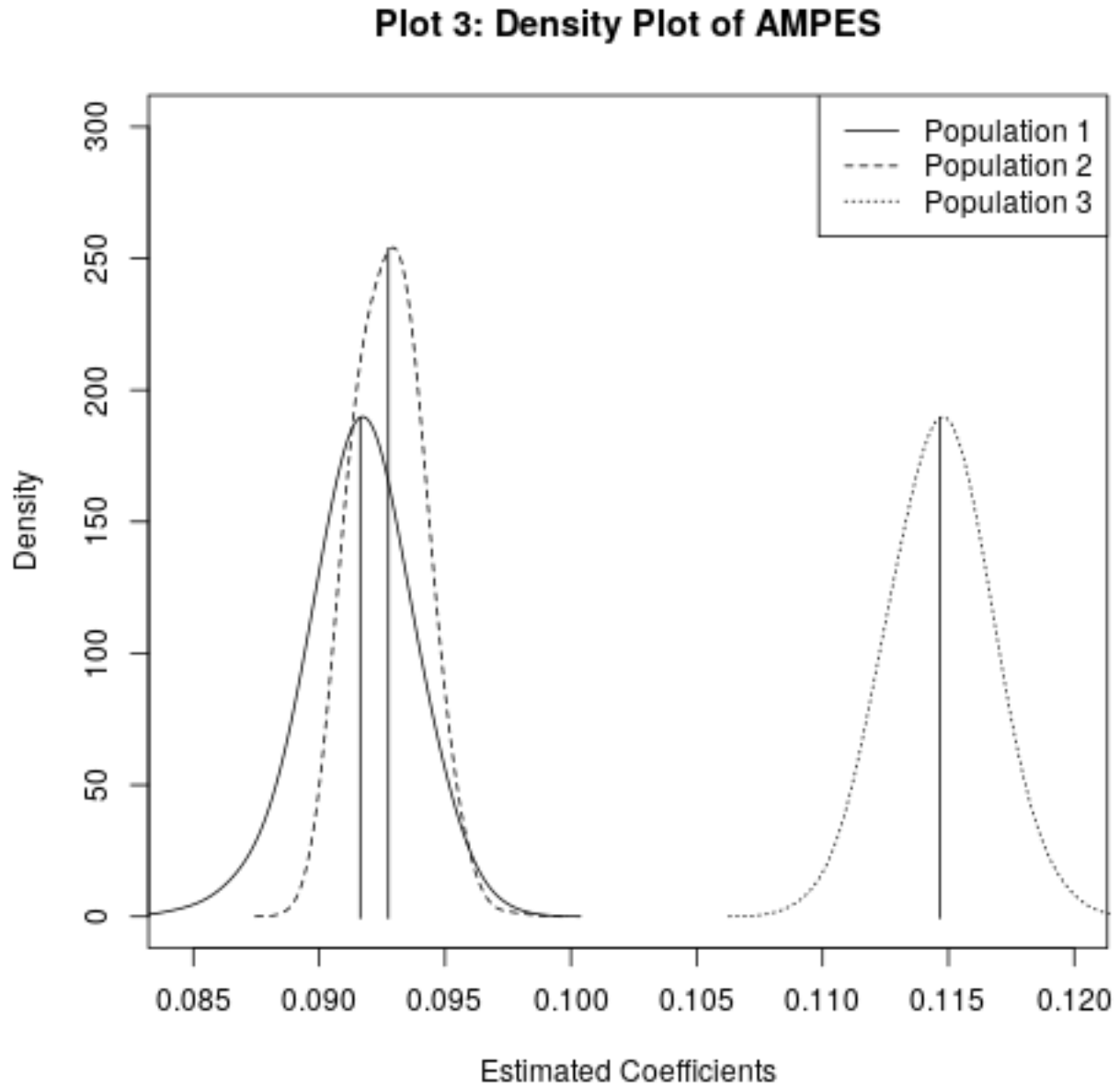
In the third population, there is covariance between the explanatory variable and the error term in the latent model. Therefore, the error term is not independent of the explanatory variable in our latent model as required in a.). In this case, it does not make much sense to compare the estimates from population three to the other ones, since the probit estimator can not be consistently estimated as being equal to β/σ anymore.

iii.) [1P]

The mean estimates for $\hat{E}(\hat{\beta}_{1,2}) = 2.001$.

We know that in the Probit model the coefficients for the regressors are consistently estimated at the true value β standardized by the standard deviation of the error term - we consistently estimate β/σ rather than just β . As long as the error term in the latent model is normalized to be one, this does not change the estimate. But in population 2, we specifically set the standard deviation of the error term to two. We therefore expect the estimates for this population to be consistently estimated at $\hat{\beta}_{1,2} = \beta_j/\sigma = 4/2 = 2$.

c.) [1P]



Again we can see that the sample distributions are approximately normal and centered around their means. We get 0.092, 0.093 and 0.115 for $\hat{E}[AMPE|j]$ given the populations $j = 1, 2$ and 3.

i.) [1.5P]

We calculate the relative difference:

$$\text{relative difference} = \left| \frac{x - y}{\max(x, y)} \right|$$

The relative difference between $\hat{E}[AMPE|j = 1]$ and $\hat{E}[AMPE|j = 3]$ is 20.08 percent.

In this case, the estimate of $\hat{E}[AMPE|j = 1]$ is more reliable, since in population 1 a) the error term is independent of the regressor X_1 and b) the error term is standard normally distributed. In population 3 on the other hand, these conditions are not met and the assumptions made in the estimation of our Probit model are therefore violated.

ii.) [0.5P]

The relative difference between $\hat{E}[AMPE|j = 1]$ and $\hat{E}[AMPE|j = 2]$ is 1.21 percent.

iii.) [3P]

Taking into account that the AMPE depends on the value of $\hat{\beta}_1$, the results in c-i) and c-ii) are not intuitive, since in plot 1 $\hat{\beta}_1$ for population 1 and 3 are much closer to each other than to distribution of the $\hat{\beta}_{1,2}$.

However, since we know that in case of population 2 - where the error term is not standard normally distributed but still independent of X_1 - the attenuation bias in β/σ will average out across its distribution in the AMPE estimation. Since both our AMPE estimators of populations 1 and 2 are consistent, we expect the AMPE estimates to approximately center around the same value and their relative difference to be very small (c-ii). In plot 3 we can see that both distributions are in fact close with $\hat{E}[AMPE|j = 2]$ being slightly larger than $\hat{E}[AMPE|j = 1]$. Since the variance of the error term in population $j = 2$ is greater than one, we could be dealing with neglected heterogeneity which is an independent part of the error term.

In population 3 we also deal the variance of the error term being greater than one. However, here the error term is correlated with X_1 and the latent variable is clearly heteroscedastic over the range of the explanatory variable. In this case, it is not possible to consistently estimate $\hat{\beta}_{1,3}$ and $\hat{E}[AMPE|j = 3]$. The violation of our model assumptions explain the relatively big discrepancy between $\hat{E}[AMPE|j = 1]$ and $\hat{E}[AMPE|j = 3]$. This can be observed in Plot 3 and results in the relative big difference calculated in c)i.

Problem 2: (Marginal effects estimation & Interpretation):

a.) [0.5P]

In non-linear regression models such as the probit model, coefficients cannot be directly interpreted as marginal effects. The coefficients give only the direction of the marginal effect of each regressor X_i on the response probability (as indicated by their sign), and we can test for the significance level of the coefficients. Therefore, we can tell from the obtained coefficients that ldl-cholesterol increases the probability of suffering from heart disease and is significantly different from zero in this sample.

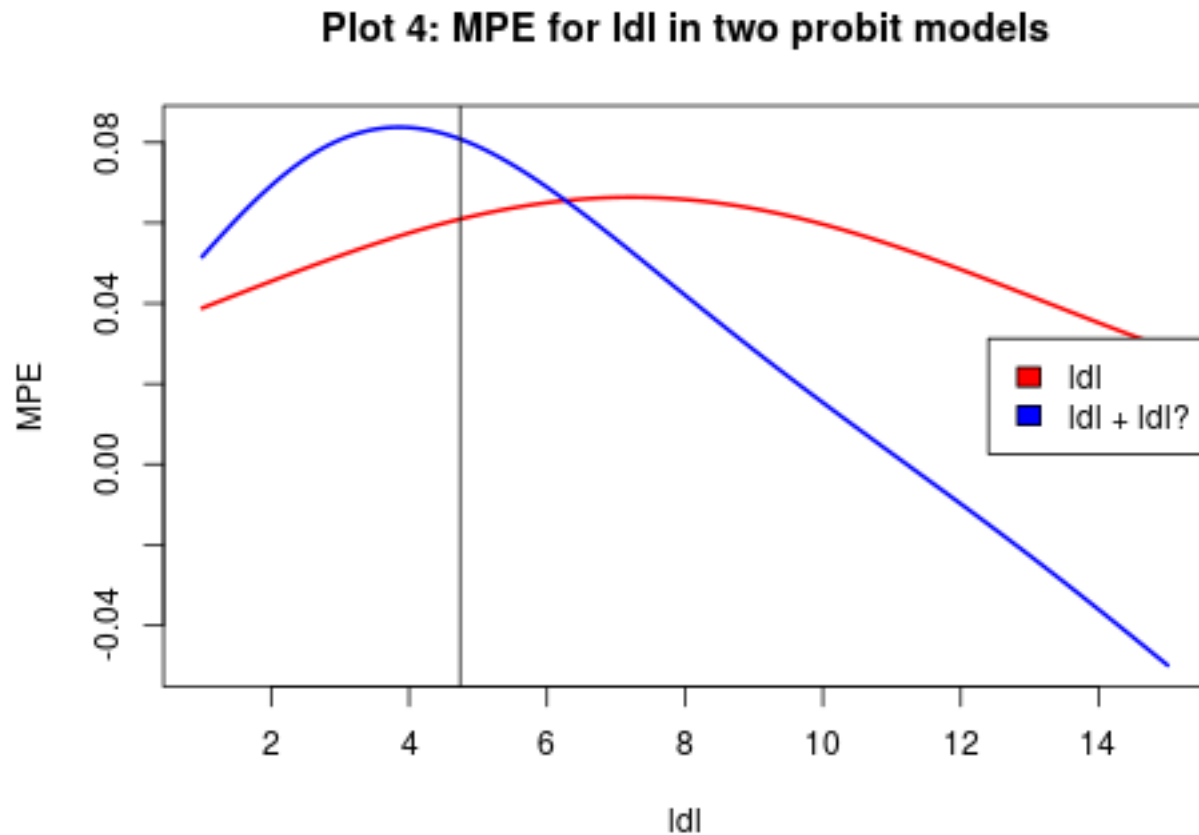
b.) [0.5P]

Since Y is a Bernoulli random variable, its variance depends on $P(y_i = 1|\mathbf{x}_i)$, which in general is a function of regressor X_i so that Y is heteroskedastic. Therefore, in the estimation of our coefficient $\hat{\beta}_1$ through the MLE method, the heteroskedasticity in $\text{Var}(y|x)$ is automatically accounted for and there is no need to further test for it. The estimated standard errors are valid and can be used to conduct the asymptotic t-test and to construct confidence intervals.

c.) [2P]

In the extended model, the coefficients are estimated under the ceteris paribus assumption (all else being equal), meaning that while estimating each of them we control for the effect of all other predictors in the model. If both variables were independent, the coefficient of *ldl* should not change in the second model when controlling for *age*. Therefore, seeing that the coefficient for *ldl* has changed compared to the estimated model without *age*, we may suspect that *ldl* and *age* are not independent. This is plausible: older people in the sample may tend to have higher levels of cholesterol. In the data there is a linear correlation between *ldl* and *age* equal to 0.31. As expected, *ldl* and *age* are not independent. In the first model, the *ldl* coefficient captured some of the effect of *age* on heart disease. When explicitly entering *age* as a regressor, this part of the variance in *ldl* is removed when estimating its regression coefficient.

d. i.) [1P]



ii.) [0.5P]

In the probit model, the MPE is given using the non-linear standard normal pdf to transform a linear combination of data weighted by the regression coefficients. Therefore, none of the MPE is linear. An increase in ldl by one unit leads to an incremental change in the probability by $\phi(\mathbf{x}'\boldsymbol{\beta})(\beta_1)$. The value of the MPE with respect to regressor X_1 also depends on the value of all other regressors and is different for every observation.

iii.) [1P]

Adding ldl^2 variable allows us to estimate whether the marginal effect of ldl on the probability increases or decreases according to the value of ldl , as opposed to model a) where we assume strict monotonicity of the effect of ldl on the marginal probability. In model d), the marginal effect of ldl^2 has a negative coefficient, which tells us that the marginal probability function of $P(Y = 1|ldl)$ is concave and will reach its maximum when $\frac{dP}{dldl} = 0$.

Interpretation: Until a certain level (in our case ~ 11 ; see plot 4) an increase of ldl is associated with an increased chance of heart disease. Above that level, the marginal probability effect of ldl

in model d.) is negative. Once the individual reaches certain level of ldl it is very likely they have a heart disease; further increase of ldl has not a positive impact on probability of getting heart disease. This interpretation would not be possible in model a), where the marginal effects are always positive.

iv.) [1.5P]

In model a), $\frac{dP}{dldl}$ of ldl at its mean equals 0.0609.

Interpretation: This marginal probability effect suggest that for an individual with the exact mean characteristics of the sample an increase of their ldl -cholesterol by one unit increases their chances of heart disease by 6.09 percentage points.

In model d), $\frac{dP}{dldl}$ of ldl at its mean equals 0.0808.

Interpretation: For an individual with the exact mean characteristics, an increase of their ldl -cholesterol by one unit increases their probability of getting the heart disease by 8.08 percentage points.

v.) [1P]

No, none of the computed effects in iv) are ceteris paribus since we have only one regressor variable ldl . In the model from d), once we fix the value of ldl at its mean, the value of ldl^2 is also automatically fixed at $(mean(ldl))^2$. This means that once we fix the value of ldl at its mean, there are no other variables that we need to control for - there can be no in the outcome caused by another variable.