

First Assignment

Malgorzata, Kevin, Johannes

20th of December, 2018

Microeconometrics 2018/2019

Assesment 1

Group Members:

Johannes Wagner, ID: 598797, Msc Statistics, wagnejoh@hu-berlin.de Malgorzata Paulina Olesiewicz, ID:598939, Msc Statistics, malgorzata.paulina.olesiewicz@student.hu-berlin.de Kevin Hope, ID: 598247, Msc Statistics, Kevin.Hoppe1@web.de

Division of work:

Task	Johannes	Malgorzata	Kevin
1.a	100% programming + 80% theory	-	20% theory
1.b	100% programming + 70% theory	-	30% theory
1.b i)	100% programming + 40% theory	40% theory	20% theory
1.b ii)	50% theory	50% theory	-
1.b iii)	100% theory	-	10% theory
1.c	100% programming	-	-
1.c i)	40% theory	50% theory	10% theory
1.c ii)	90% theory	-	-
1.c iii)	50% theory	50% theory	-
2.a	-	100% theory + programming	-
2.b	-	100% theory	-
2.c	50% theory	50% theory + 100% programming	-
2.d i)	-	-	100% programming
2.d ii)	30% theory	40% theory	30% theory
2.d iii)	-	100% theory	-
2.d iv)	30% theory	40% theory + 100% programming	30% theory
2.d v)	30% theory	40% theory	30% theory

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

Consider the following JOINT distributions of random variables X_0 & ε_j^*

$$\begin{pmatrix} X_0 \\ \varepsilon_1^* \end{pmatrix} \sim N \left[\begin{pmatrix} 10 \\ 0 \end{pmatrix}, \begin{pmatrix} 2^2 & 0 \\ 0 & 1 \end{pmatrix} \right], \begin{pmatrix} X_0 \\ \varepsilon_2^* \end{pmatrix} \sim N \left[\begin{pmatrix} 10 \\ 0 \end{pmatrix}, \begin{pmatrix} 2^2 & 0 \\ 0 & 2^2 \end{pmatrix} \right],$$
$$\begin{pmatrix} X_0 \\ \varepsilon_3^* \end{pmatrix} \sim N \left[\begin{pmatrix} 10 \\ 0 \end{pmatrix}, \begin{pmatrix} 2^2 & 3 \\ 3 & 2^2 \end{pmatrix} \right]$$

with

$$\begin{aligned} \beta_0 &= -30 \\ \beta_1 &= 4 \end{aligned}$$

$$Y_j^* = \beta_0 + \beta_1 * X_1 + \varepsilon_j$$

$$Y_j = \begin{cases} 1 & \text{if } Y_j^* > 0 \\ 0, & \text{otherwise} \end{cases}$$

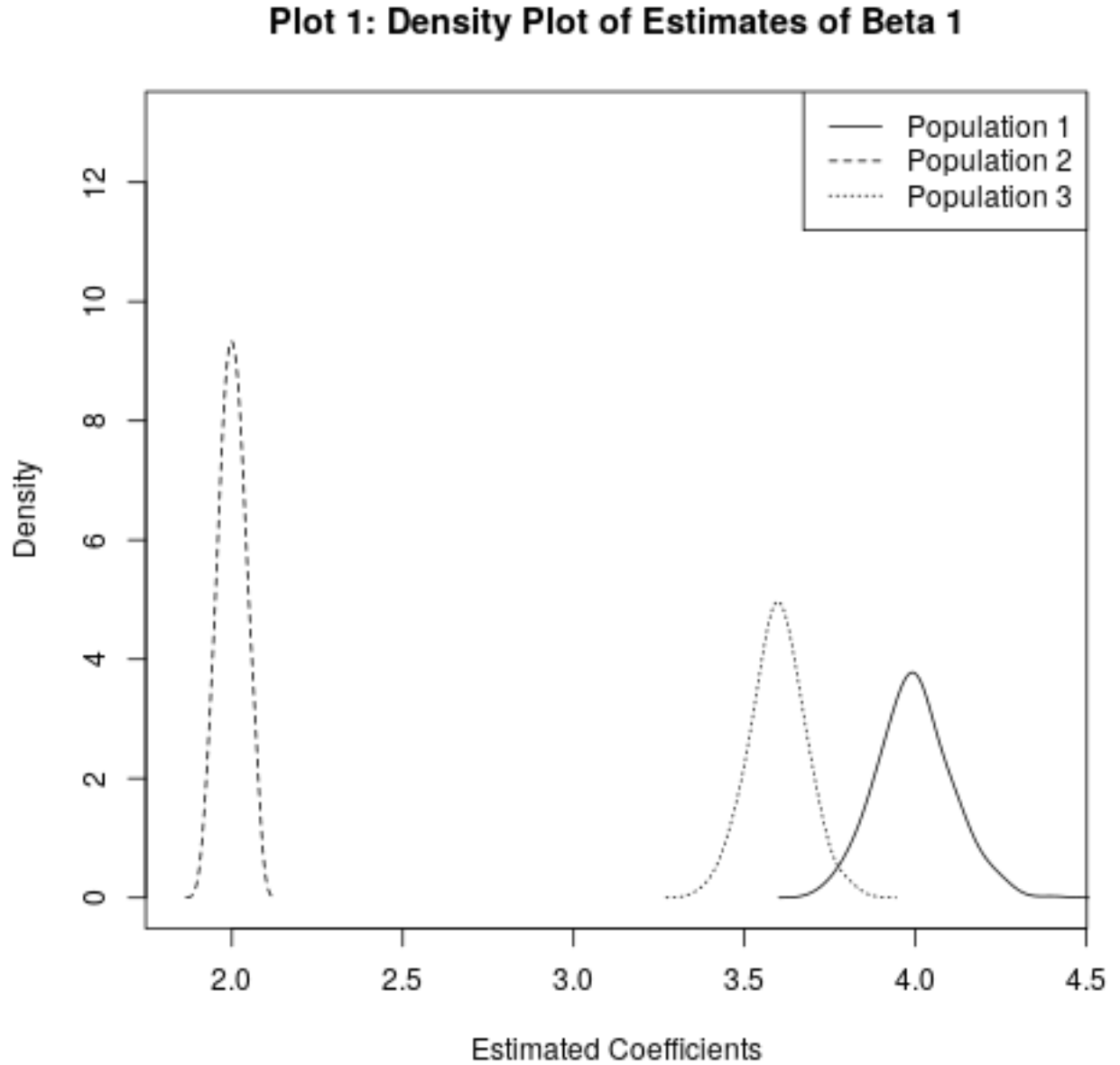
$$j \in \{1, 2, 3\}$$

a) [1P]

Given our simulated Data of independent samples, the parameter estimate in the respective Probit model $\hat{\beta}_{1,j}$ for $j = 1, 2, 3$ can be obtained through the maximization of the log likelihood of the probability function for our binary dependend variable y_i given x_1 .

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]



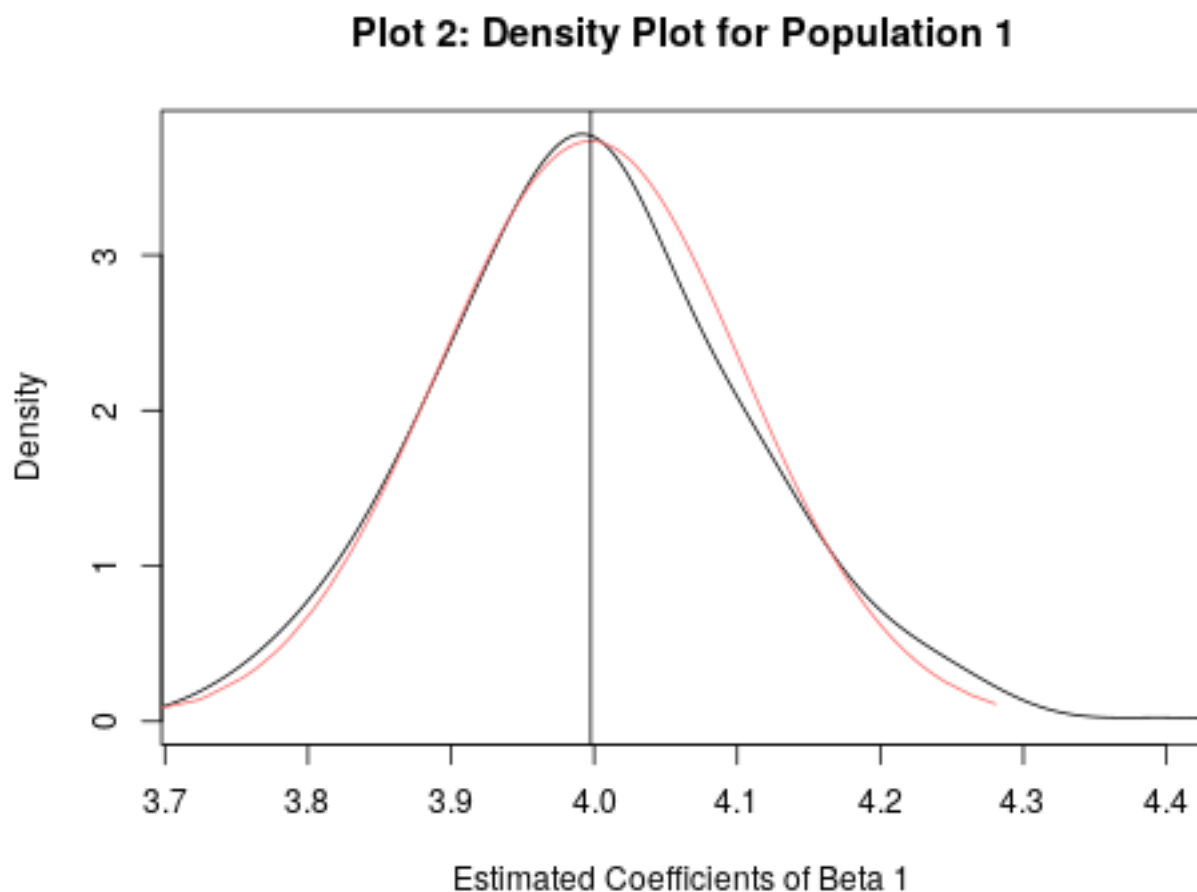
In order to be able to interpret our distributions, we need to find a smooth representation for the probability density function of our estimates. We can make use of a kernel estimator

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

where we choose K to be a *Gaussian Kernel*, which is the conventional choice and produces a standard normal density function. That means that $K(x) = \phi(x)$, where ϕ is the probability density function of the standard normal. Given that we expect our estimator β_1 to have approximately normal distribution (see answer b.ii. below), this is the best choice for our kernel density estimation. The

smoothing parameter h is called the bandwidth and had to be chosen accordingly to the data. The goal is to find a bandwidth that is as small as possible to avoid loss of information and as big as necessary to represent the true density of our estimates. A rule-of-thumb for the bandwidth in R is “Silvermann’s *rule of thumb*” (See man page of function `density()` in R). It is used as the default for calculating the density in R and can be adjusted through the multiplication with an adjustment parameter. We can reduce the bandwidth with an adjustment parameter smaller than zero. This resulted in an unstable distribution picture, especially for population 1. Therefore we decided to increase the bandwidth by the factor 1.5 to get a stable and informative distribution.

i.) [1.5P]



Given the basic properties of a ML-estimator, we expect our estimates of $\hat{\beta}_{1,j}$ in general to be consistent and asymptotically normally distributed.

Given our relatively large number of samples, we can expect our estimator to asymptotically converge in probability to the true parameter β_1 , given that the assumptions of our probit model are not violated. 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 our estimates (black line in Plot 2) approximately follows a normal distribution, which is illustrated with the red line. The estimates of $\hat{\beta}_1$ 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 distributed, but differ in their means and variances. For the means of our estimates we find that $\hat{E}(\hat{\beta}_{1,1}) = 3.998$, $\hat{E}(\hat{\beta}_{1,2}) = 2.001$ and $\hat{E}(\hat{\beta}_{1,3}) = 3.598$.

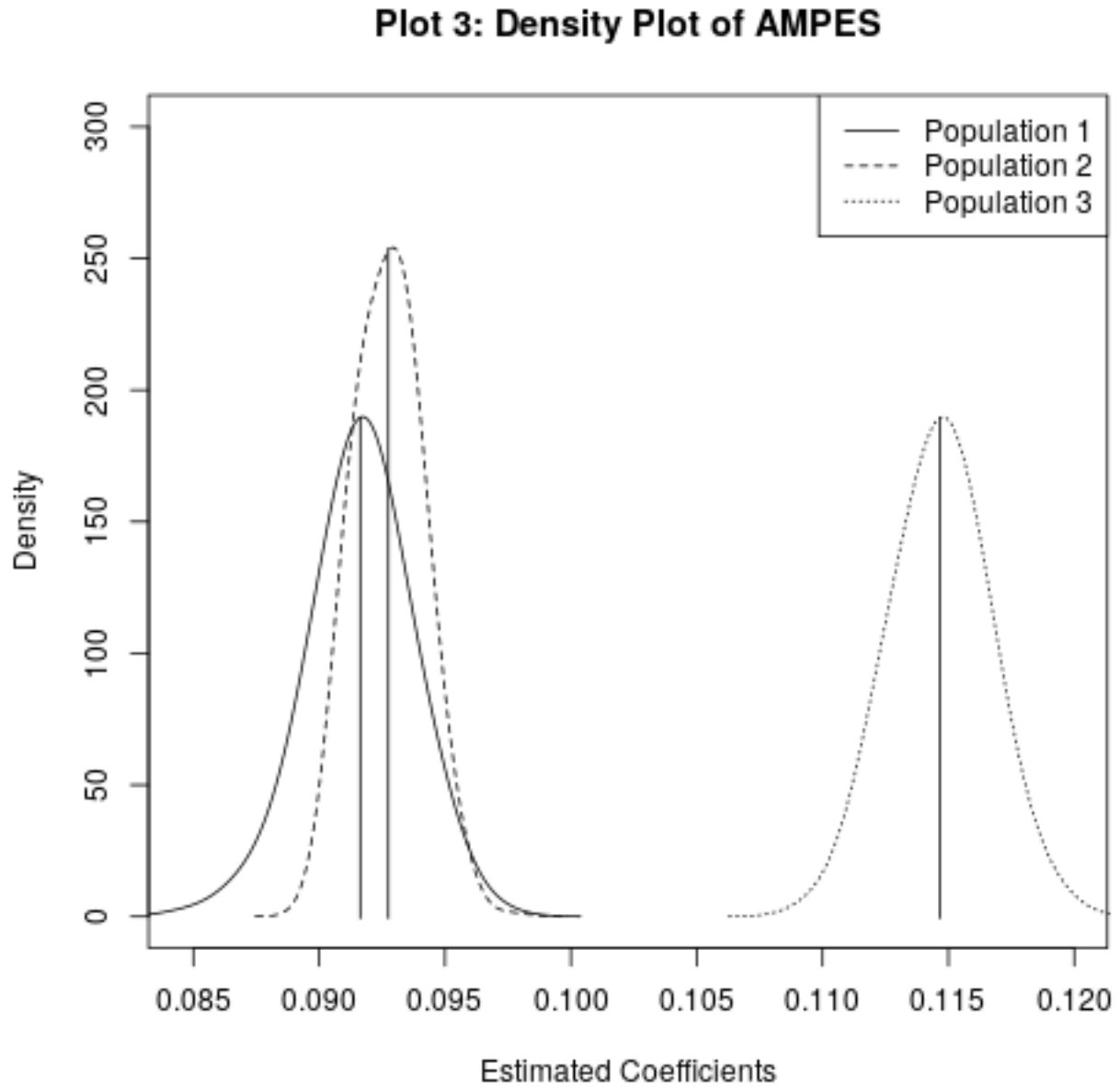
Given the assumptions stated in i), we expect the estimate of $\hat{\beta}_1$ in this population to be close to the true value of 4, as it is the case. In the second population our errors have a variance greater than 1. This results in a decreased estimate of $\hat{\beta}_1$, since the probit estimator is derived as being equal to β/σ with σ being the standard deviation of our error term. In the third population we can additionally observe a covariance greater than zero between the explanatory variable and the error term in our latent model. Therefore our error term is not independent of the explanatory variable in our latent model as requested by a.) and our estimate of $\hat{\beta}_1$ from population $j = 3$ is inconsistent. 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 derived as being equal to β/σ anymore.

iii.) [1P]

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

We know that probit of y on x consistently estimates β/σ rather than just β . When the error term in our latent model is not standard normal distributed, as this is the case in population 2, with a variance of the error term of 4 (or standard deviation of 2), we expect our estimator to be consistently estimated at $\hat{\beta}_1 = \beta_j/\sigma = 4/2 = 2$.

c.) [1P]



Again we can see that the sample distributions are approximately normal and concentrated 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 the model for the population 1 we observe a) the error term to be uncorrelated with the variable x_1 and b) the standard normality of the error term. Therefore, $\hat{E}[AMPE|j = 1]$ can be estimated consistently as oppose to $\hat{E}[AMPE|j = 3]$ which is estimated from a model of population 3, which violates both assumptions necessary for asymptotic consistency of our estimator.

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 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 $\hat{\beta}_1$ for population 2.

However, since we know that in case of population 2 - where the error term does not follow a standard normal, but is still uncorrelated with x_1 - the attenuation bias in β/σ will average out across its distribution during the AMPE estimation. Since both our AMPE estimators of population $j = 1$ and $j = 2$ are consistent, we expect the AMPE estimates to concentrate approximately around the same value and their relative difference to be very small (c-ii). If we look at Plot 3, we see that both distributions are in fact close, with $\hat{E}[AMPE|j = 2]$ being slightly bigger than $\hat{E}[AMPE|j = 1]$. Since the variance of the error term in population $j = 2$ is greater than 1, we assume to be dealing with neglected heterogeneity here, which adds an unknown constant to our partial effects in the probit.

In population 3 we also deal with suspected unobserved heterogeneity with the variance of the error term being greater than 1. However, here the error term is correlated with x_1 and the probit is clearly heteroscedastic with regard to the explanatory variable. In this case it is not possible to estimate $\hat{\beta}_1$ and $\hat{E}[AMPE|j = 3]$ consistently. The violation of our model assumptions in the probit and the resulting inconsistency of the estimator explains 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 interpreted as marginal effects. In probit model, the coefficients could be interpreted with respect to the latent variable but since it has no specific unit, we cannot use any meaningful scale. Therefore, the coefficients give only the direction (signs) of the partial effect of each x on the response probability, and its statistical significance by giving us the information whether we can reject null hypothesis at sufficiently small significance level. Therefore, *ldl*-cholesterol increases the probability of suffering from the disease and is significantly different from zero.

b.) [0.5P]

The MLE β_1 is derived from the log-likelihood function, which is a logarithmic transformation of a joint probability density function of the given sample, where the distribution of y depends on x . Because y is a binary (Bernoulli) random variable, the variance depends on $P(y_i = 1|\mathbf{x}_i)$, which in general is a function of x_i , i.e. 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 have asymptotic properties and can be used to conduct the asymptotic t-test and confidence intervals.

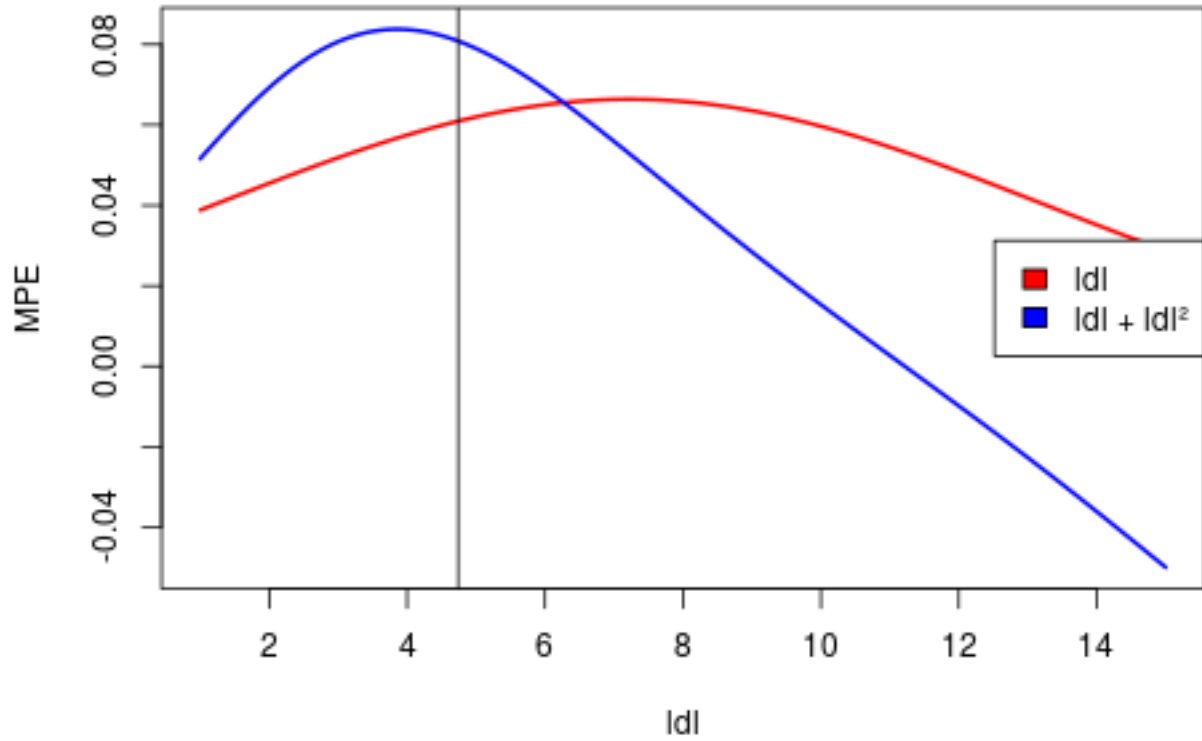
c.) [2P]

In the second 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. Plausibility check: older people in the sample may tend to have higher levels of cholesterol. Now, let's check for that in the data: Correlation between variables *ldl* and *age* = 0.31. As expected, *ldl* and *age* are not independent. In the first model, the *ldl* coefficient captured some of the *ldl* variations on *age*. When explicitly entering *age* as a regressor, that part of the variance in *ldl* is removed from the coefficient.

d. i.) [1P]

test

Plot 4: MPE for ldl in two probit models



ii.) [0.5P]

In Probit model, to calculate the MPE we are transforming the linear model of the beta estimators through the pdf (non-linear function) multiplication, which indicates that non of the the MPE is therefore linear. An increase in ldl by one unit leads to an incremental change in the probability by $\phi(\mathbf{x}\beta)(\beta_1)$. The value of the MPE with respect to that one regressor therefore depends on the value of all other regressors and is estimated for every observation sperepatly.

iii.) [1P]

Adding ldl^2 variable allows us to estimate whether the marginal effect of ldl on probability increases or decreases according to the value of ldl , as oppose to model a) where we assume strictly monotonic effect of ldl on marginal probability. In estimation d) the marginal effect of ldl^2 has a negative coefficient, which tells us that the marginal popability function of $P(Y = 1|ldl)$ is concaved and will reach its maximum when $\frac{dP}{dldl}$ is equal to 0. Interpretation: Until certain level (in our case ~ 11 see plot 4) the increase of ldl will increase the chances of the getting the heart disease, however above this level the marginal probability effect of ldl in model d.) will become negative. Once the individual reaches certain level of ldl it is very likely they have a heart disease, therefore even further increase of ldl has not a positive impact on probability of getting heart disease anymore. This interpretation

is not possible based on the estimation a), where we can only observed that increase in the *ldl* level increases the probability of getting the heart disease.

iv.) [1.5P]

In the model a), $\frac{dP}{dldl}$ of *ldl* at mean is equal to 0.0609395.

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

In the model d), $\frac{dP}{dldl}$ of *ldl* at mean is equal to 0.080844.

Interpretation: For an individual with the exact mean characteristics, 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 control variable ldl. In the model from d), once we fix the value of ldl at the 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 mean, there is no other variables that we need to control for - there is no chance for change in the outcome caused by another variable.