1. Theory

Question 1

a) $\mathrm{E}(\epsilon_i) = \mu_\epsilon \neq 0$

The Formula for $\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$ according to slide 4 can be rewritten as $\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2}$, since $\sum x_i y_i$ is equal to $\sum x_i Y_i$. For Y_i, one can use the popular regression function and rewrite $\hat{\beta}_2$ as follows:

$$\hat{\beta}_2 = \frac{\sum x_i(\beta_1 + \beta_2 X_i + \epsilon_i)}{\sum x_i^2}$$

This fraction can be separated and the βs can be pulled out:

$$\hat{\beta}_2 = \beta_1 \frac{\sum x_i}{\sum x_i^2} + \beta_2 \frac{\sum x_i X_i}{\sum x_i^2} + \frac{\sum x_i \epsilon_i}{\sum x_i^2}$$

We know that $\sum x_i$ sums up to 0 and that $\sum x_i X_i$ is equal to $\sum x_i^2$, leaving the equation as follows:

$$\hat{\beta}_2 = \beta_2 + \frac{\sum x_i \epsilon_i}{\sum x_i^2}$$

Looking at the expected values, we modify this equation:

$$E[\hat{\beta}_2] = E[\beta_2 + \frac{\sum x_i \epsilon_i}{\sum x_i^2}]$$

$$E[\hat{\beta}_2] = E[\beta_2] + E[\frac{\sum x_i \epsilon_i}{\sum x_i^2}]$$

$$E[\hat{\beta}_2] = \beta_2 + \frac{\sum x_i E[\epsilon_i | X_i]}{\sum x_i^2}$$

$$E[\hat{\beta}_2] = \beta_2 + \frac{\sum x_i \mu_\epsilon}{\sum x_i^2}$$

With $\mu_\epsilon \neq 0$ we can see that $E[\hat{\beta}_2] \neq \beta_2$, meaning that $\hat{\beta}_2$ is biased.

The Formula for $\hat{\beta}_1$: $\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}$ includes this biased $\hat{\beta}_2$, meaning $\hat{\beta}_1$ is also biased.

Tarik Benli 15-719-818, Ramon Gmür 16-705-220

b) $\tilde{X} = 2X$

The formula for $\hat{\beta}$ is, according to slide 5 as follows: $\hat{\beta} = (X'X)^{-1}X'y$

For $\tilde{\beta}$, the formula looks like this:

$$\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y$$

Since we know that $\tilde{X} = 2X$, we can rewrite this formula:

$$\tilde{\beta} = (2X'2X)^{-1}2X'y$$

$$\tilde{\beta} = \frac{1}{4}(X'X)^{-1}2X'y$$

$$\tilde{\beta} = \frac{1}{2}(X'X)^{-1}X'y$$

$$\tilde{\beta} = \frac{1}{2}\hat{\beta}$$

This means that by doubling *X*, the coefficient $\hat{\beta}$ gets multiplied by $\frac{1}{2}$.

c) $y^* = 2y$

Using the same formula from slide 5 again ($\hat{\beta} = (X'X)^{-1}X'y$), the new equation for β^* looks as follows:

$$\beta^* = (X'X)^{-1}X'y^*$$

Knowing the new *y** being equal to 2*y*, we can rewrite this:

$$\beta^* = (X'X)^{-1}X'2y$$

$$\beta^* = 2(X'X)^{-1}X'y$$

$$\beta^* = 2\hat{\beta}$$

This means that by doubling *y*, the coefficient $\hat{\beta}$ gets multiplied by 2.

d) The results from the previous questions suggest that a linear change in the units in which *X* and *y* are measured does not affect the relationship between *X* and *y*.

e) From question b) we already know that $\tilde{\beta} = \frac{1}{2}\hat{\beta}$. Therefore, one can rewrite the variance of $\tilde{\beta}$:

$$V(\tilde{\beta}) = V(\frac{1}{2}\hat{\beta})$$

Since $\frac{1}{2}$ is a constant multiplied with $\hat{\beta}$, we can pull it out of the variance and square it:

$$V(\tilde{\beta}) = \frac{1}{4}V(\hat{\beta})$$

The formula on slide 5 ($V(\hat{\beta}) = \sigma^2(X'X)^{-1}$) gives the same results:

$$V(\tilde{\beta}) = \sigma^2(\tilde{X}'\tilde{X})^{-1} = \sigma^2(2X'2X)^{-1} = \frac{1}{4}\sigma^2(X'X)^{-1} = \frac{1}{4}V(\hat{\beta})$$

Tarik Benli 15-719-818, Ramon Gmür 16-705-220

This means that by doubling *X*, the variance of $\hat{\beta}$ gets multiplied by one quarter.

2. Empirical Application

Question 2

a) Observations

This dataset has 807 observations with 10 variables each.

b) Summary statistics

Descriptive Statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
cigs	807	8.686	13.722	0	80
educ	807	12.471	3.057	6	18
age	807	41.238	17.027	17	88
income	807	19304.83	9142.958	500	30000
white	807	.879	.327	0	1
restaurn	807	.247	.431	0	1

The summary statistics show that all individuals have answered all the questions. The individuals have at least 6 years of schooling. However, the minimum age is 17, meaning that there are probably individuals still being in school, falsifying possible results of the effect of education on cigarettes. All the observed incomes are between 500 and 30'000 with a mean income of 19'304.83. White and restaurn are dummy variables, but the means show that most individuals are white and that only a quarter of the states, in which the individuals live, restrict smoking in restaurants.

c) i) The estimators are computed with the following formulas:

$$\hat{\beta}_2 = \frac{Cov(x,y)}{Va\ (x)} \text{ and } \hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{X}$$

The results are: $\hat{\beta}_2 = -.21855212$ and $\hat{\beta}_1 = 11.41203$.

Tarik Benli 15-719-818, Ramon Gmür 16-705-220

ii) The table below shows the regression results:

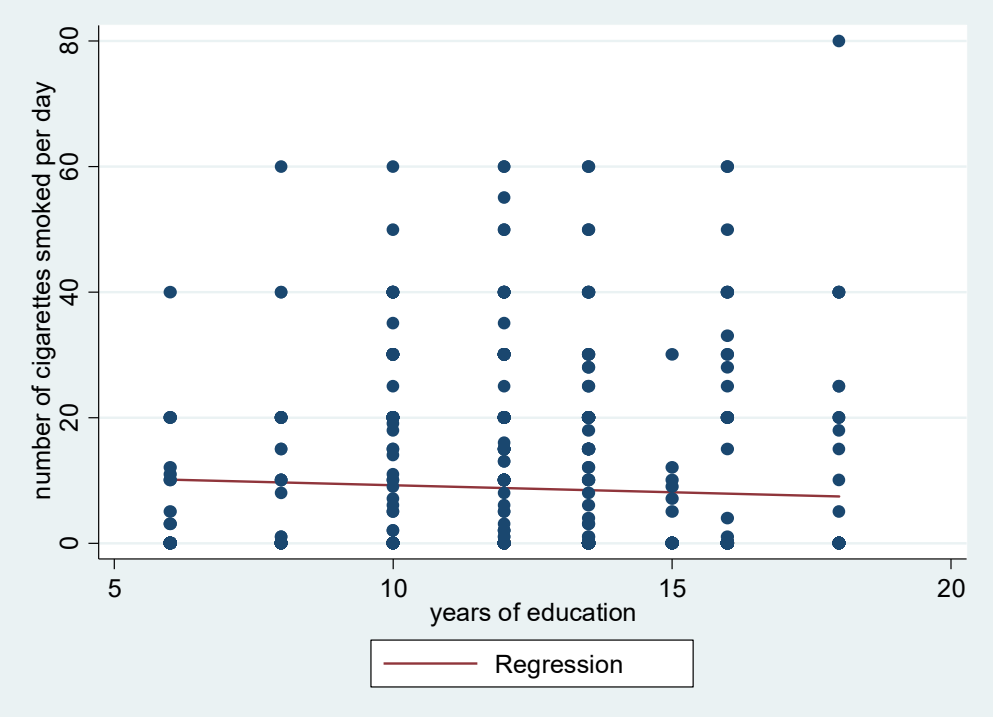
VARIABLES	(1) Reg
educ	-0.219
	(0.158)
Constant	11.41***
	(2.029)
Observations	807
R-squared	0.002

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The estimators in the regression show the exact same results as in 2.c)i).

iii) If the mean-zero error (assumption 2) is satisfied, we know that the estimator is unbiased. The effect of -0.219 means that with one year of additional education, an average individual smokes 0.219 cigarettes fewer per day. On one hand, this effect makes sense, since more education possibly increases awareness of the health risks of smoking. On the other hand, 0.2 cigarettes fewer per day seems like a rather small effect per additional year of education. The effect's standard error is rather large compared to the effect itself, making it insignificant.

iv) $\widehat{cigs}_i = 11.41 - 0.219 \times \text{educ}_i$

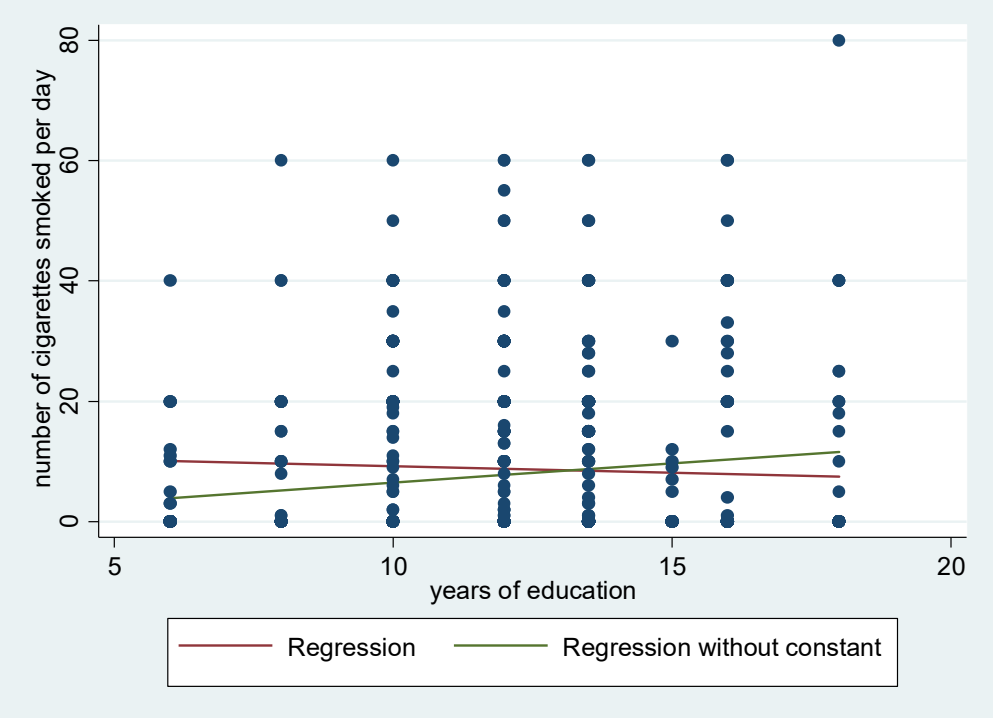


v) The table below shows the regression results:

VARIABLES	(1) Reg
educ	0.645*** (0.0383)
Observations	807
R-squared	0.260

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The effect of education on cigarettes is now positive. Meaning with one additional year of education, an average individual smokes 0.645 cigarettes more. This regression without a constant makes little sense, because it also suggests that an individual with 0 years of education does not smoke at all.



d) i) The table below shows the regression results:

VARIABLES	(1) Reg
educ	-0.452***
	(0.162)
age	0.826***
	(0.154)
age2	-0.00963***
	(0.00168)
white	-0.624
	(1.456)
restaurn	-2.796**
	(1.104)
Constant	0.669
	(3.707)
Observations	807
R-squared	0.051

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The coefficient of white is -0.624. This could lead to the interpretation that on average, individuals of white ethnicity smoke 0.624 cigarettes fewer per day. However, compared to the mean this effect seems rather small and it is also not significantly different from 0. The coefficient of restaurn is -2.796. This leads to the conclusion that individuals who live in a state where smoking in restaurants is restricted smoke on average 2.796 cigarettes fewer per day. This effect is not quite small and the coefficient is significantly different from 0 on a 10%-level.

ii) Marginal effects

In order to find the marginal effect of age, one has to derive the regression with respect to age:

$$\widehat{cigs}' = 0.826 + 2 \times (-0.0096) \times age$$

The calculated marginal effect for 20 years is:

$$\Delta\widehat{cigs}' = 0.826 + 2 \times (-0.0096) \times 20 = 0.442$$

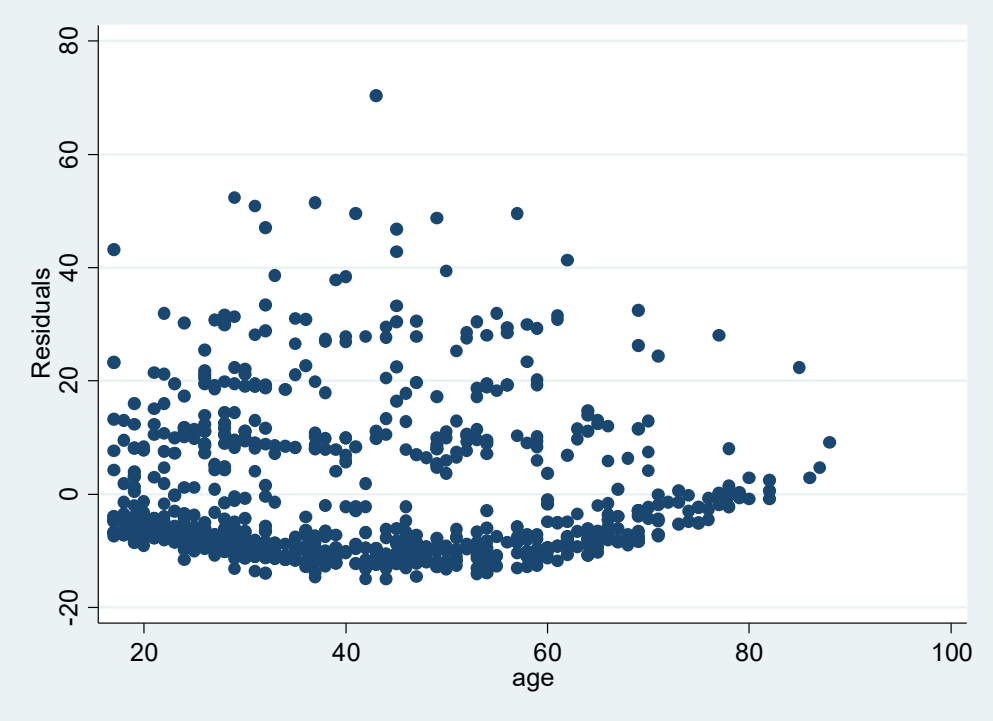
The calculated marginal effect for 40 years is:

$$\Delta\widehat{cigs}' = 0.826 + 2 \times (-0.0096) \times 40 = 0.058$$

The calculated marginal effect for 60 years is:

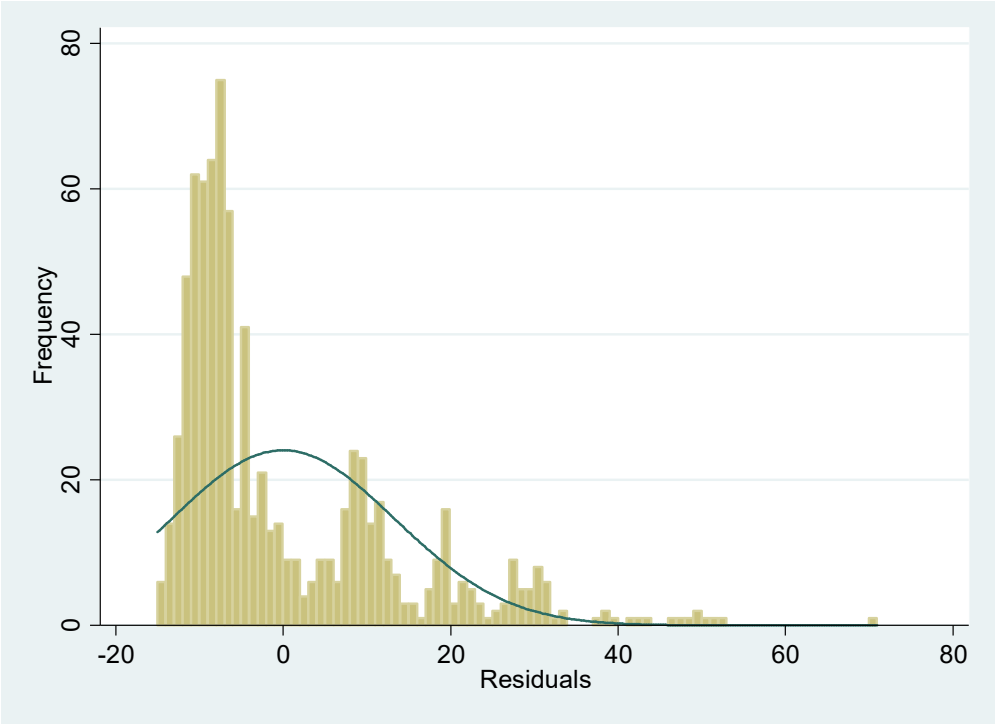
$$\Delta\widehat{cigs}' = 0.826 + 2 \times (-0.0096) \times 60 = -0.326$$

iii) A) In general, it seems that the variance is constant throughout all ages. However, for older individuals the variance seems to be smaller. This might be explained with relatively fewer observations starting the age of 60. Furthermore, the reason for this model to overestimate the cigarette consumption of individuals older than 60 might be that individuals who consume cigarettes already passed away due to health issues caused by smoking. Therefore, we think that assumption 3 holds.



B) The correlation between the residuals and the lagged residuals is -.0042332 and not significantly different from 0. Therefore, we suggest that assumption 4 holds.

C) The distribution of the residuals, compared to a normal distribution, shows a positive skewness, a negative mean and a negative median. This suggests that assumption 5 does not hold.



3. Log-file

See attachment