Exercise 06

Topic 1: The causal effect of smoking on depression

Exercise 6.1

a)

Several studies in the available literature have explored the relationship between smoking and depression. It is becoming increasingly clear that smoking is not innocuous to mental health, and may in fact aggravate mental illness or contribute to its onset (Pasco et al., 2008). Adjustment for confounding factors revealed persistent significant associations between nicotine-dependence symptoms and depressive symptoms (Boden et al., 2010). Additionally, Khaled (2012) also found that the heavy-smoking-to-depression pathway is causal in nature, rather than mainly due to confounding by shared-vulnerability factors. Based on these findings, a testable hypothesis emerges:

Hypothesis 1: Cigarette smoking is a risk factor for the development of depression.

This hypothesis can be tested through the longitudinal dataset easySHARE by studying the relationship between smoking behavior and the index of depression over time. Utilizing both the OLS regression and Fixed Effects model in longitudinal data analysis will allow for a comprehensive examination of the impact of smoking on the development of depression, leading to a more accurate understanding of the relationship between these variables.

b)

Gender differences in smoking behavior and mental health outcomes suggest the possibility of effect heterogeneities by gender. For instance, women have higher rates of major depression: the lifetime prevalence for men is 12%, whereas for women it is 21% (Paperwalla et al., 2004). The prevalence of current and lifetime dysthymia was approximately two times higher for women than men (Weinberger et al., 2013). Therefore, it's reasonable to expect effect heterogeneities by gender when studying the impact of smoking on the development of depression.

Exercise 6.2

a)

The summary statistics shown in **Table 1** (see below) provide an overview of the main independent variable (SMOKING) as well as socio-demographic variables.

Table 1. Summary Statistics of Main Independent Variable and Socio-Demographic Variables

	2-No (N=129037)	1-Yes (N=27375)	Overall (N=156412)
DEPRESSION			
Mean (SD)	2.17 (2.15)	2.32 (2.27)	2.20 (2.17)
Median [Min, Max]	2.00 [0, 12.0]	2.00 [0, 12.0]	2.00 [0, 12.0]
SMOKING			
2-No	129037 (100%)	0 (0%)	129037 (82.5%)
1-Yes	0 (0%)	27375 (100%)	27375 (17.5%)
Gender			
1-Female	74008 (57.4%)	13004 (47.5%)	87012 (55.6%)
2-Male	55029 (42.6%)	14371 (52.5%)	69400 (44.4%)
Age of respondent			
Mean (SD)	65.0 (7.75)	61.5 (7.29)	64.4 (7.78)
Median [Min, Max]	65.1 [50.0, 79.0]	60.6 [50.0, 79.0]	64.3 [50.0, 79.0]

Table 1 provides an overview of survey responses categorized by smoking status (smokers vs. non-smokers). Depression Index, as the dependent variable, is assessed using an index ranging from 0 to 12. The table indicates that smokers exhibit a higher mean depression index (2.32) compared to non-smokers (2.17), hinting at a potential correlation between smoking and elevated depression levels.

Here's a breakdown by smoking status:

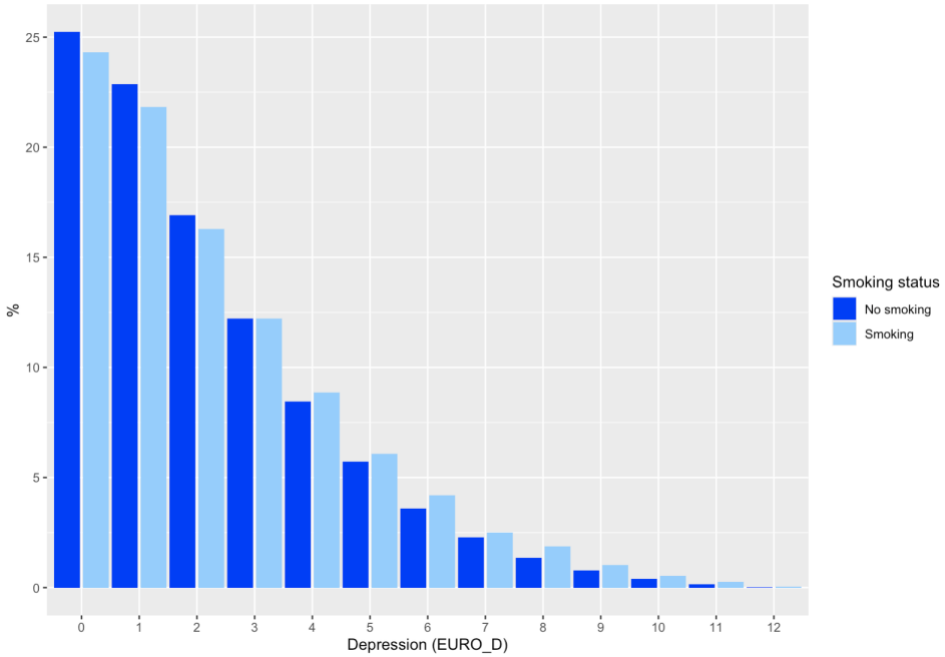
- No (82.5%): The mean depression index is 2.17 with a standard deviation of 2.15.
- Yes (17.5%): The mean depression index is 2.32 with a standard deviation of 2.27.

It is important to note that correlation does not imply causation. While this table demonstrates a disparity in depression index between smokers and non-smokers, it does not conclusively determine whether smoking causes depression or if individuals with depression are more inclined to smoke.

b)

Figure 1 displays the distribution of the outcome variable (Depression Index) among smokers and non-smokers.

Figure 1. Distribution of Depression Index by Smoking Status



The graph reveals a similar distribution of depression index for both groups, with the majority falling within the 0-2 depression index range. However, there are discernible differences. Smokers exhibit a slightly higher proportion of individuals in the higher depression index categories (8-12) compared to non-smokers, implying that smokers may be more susceptible to experiencing moderate to severe depression than non-smokers.

Exercise 6.3

a)

The results of the OLS model (see **Table 2** in Appendix), controlling for key variables such as smoking status, sports participation, job status, and other socio-demographic factors (gender, age, country, and wave), show that smoking is significantly associated with higher levels of depression in the easySHARE dataset. Individuals who smoke have a statistically significant depression index of 0.25 points higher compared to non-smokers. Gender also plays a role, with men having a significantly lower depression index (-0.66 points) compared to women. Age shows a weak positive association, with a depression index increase of 0.43 points per 100 years.

Sports participation has mixed effects. Engaging in sports "Hardly ever" is linked to a significantly higher depression index (0.65 points) compared to not participating at all. However, participating "One to Four times a

month" does not show a significant difference. Job status also matters. Being retired or unemployed is associated with a significantly higher depression index compared to being employed. The increase is 0.12 points for retired and 0.69 points for unemployed individuals.

The model accounts for baseline differences in the depression index between countries by including fixed effects for each country. Compared to a reference category (Austria), significant differences were found for several countries. Similarly, fixed effects for each wave of the survey control for potential changes in depression index over time. Here, significant differences were found for almost all waves.

b)

The second research (see **Table 2** in Appendix) utilizes an FE-model to examine the relationship between smoking status and depression index within individuals over time. This approach controls for unobserved characteristics of individuals that might influence depression but remain constant over time (e.g., genetics, personality).

The results reveal a negative and statistically significant coefficient for smoking (-0.09, p-value < 0.001). This suggests that within individuals, smokers tend to have a depression index 0.09 points lower on average compared to non-smokers. However, it's important to consider the possibility of reverse causality. If individuals who experience higher levels of depression are more likely to start smoking as a coping mechanism, then the negative association found in the FE-model may be due to reverse causality rather than a true causal effect.

The model also included fixed effects for each wave of the survey (WAVE1-WAVE8), excluding WAVE3 which has no information on smoking. These coefficients capture the average change in depression index for the entire population relative to the reference wave (WAVE1). Significant differences were found for all waves.

c)

The third research (see **Table 3**) focuses on estimating the FE-model separately for males and females to examine the relationship between smoking status and depression index within each gender over time.

Table 3. Fixed Effects Model for Depression Index by Gender

	Female			Male		
Characteristic	**Beta**	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**
SMOKING						
2-No	—	—		—	—	
1-Yes	-0.04	-0.12, 0.05	0.4	-0.13	-0.20, -0.07	<0.001
WAVE						
1	—	—		—	—	
2	-0.11	-0.16, -0.06	<0.001	-0.05	-0.09, 0.00	0.034
4	0.07	0.02, 0.12	0.008	0.19	0.14, 0.24	<0.001
5	0.03	-0.02, 0.08	0.3	0.14	0.10, 0.19	<0.001
6	0.09	0.04, 0.15	0.001	0.26	0.20, 0.32	<0.001
7	0.10	0.03, 0.17	0.004	0.21	0.14, 0.28	<0.001
8	0.02	-0.04, 0.08	0.5	0.19	0.14, 0.25	<0.001

[1] CI = Confidence Interval

The results reveal a potential gender difference. For females, the effect of smoking on the depression index is negligible (the coefficient for smoking is not statistically significant). In contrast, for males, smokers tend to have a depression index 0.13 points lower on average compared to non-smokers within the model (statistically

significant negative coefficient). However, it's important to consider the possibility of reverse causality. Specifically, it suggests that males experiencing higher levels of depression may be more inclined to initiate smoking as a coping mechanism. Overall, the evidence suggests a bidirectional association between smoking and depression.

Both models included fixed effects for each wave of the survey to capture average changes in depression index within each gender group over time. There are significant differences in these wave effects. For females, the average depression index fluctuated slightly over time. For males, the average depression index shows a more pronounced increase across later waves compared to the reference wave.

In summary, the findings suggest a potential difference in how smoking is associated with depression for males and females. Further research is needed to explore the reasons behind this gender difference and the observed changes in the average depression index within each gender group over time.

Exercise 6.4

In Exercise 6.3, we employed both OLS regression and Fixed Effects models to examine the relationship between smoking status and depression scores within the easySHARE dataset. While these models provide valuable insights, it's essential to consider whether they generated causal effects.

The Fixed Effects model effectively controls for unobserved time-invariant individual-specific factors, such as genetic predispositions, personality traits, or early life experiences, which may confound the relationship between smoking and depression. For instance, the rewarding effects of smoking and the beneficial effects of nicotine replacement therapy for depressed smokers may depend, in part, on genetic factors involved in dopamine transmission (Lerman et al., 1998).

While numerous studies suggest that smoking increases the risk of developing depression, it's also plausible that individuals experiencing depressive symptoms may turn to smoking as a coping mechanism. Such conditions are amenable to management by nicotine because of its ability to produce small but reliable adjustments in relevant cognitive and behavioral functions (Pomerleau, 1997).

The relationship between smoking and depression may vary across different subgroups of the population, such as gender. The FE model, by accounting for individual-specific heterogeneity, may unveil nuanced patterns or interactions that are not apparent in aggregated analyses. This capability allows for a more comprehensive understanding of how smoking influences depression across diverse demographic and clinical profiles.

In conclusion, while these analyses provide valuable insights into the association between smoking and depression, they should be interpreted with caution.

Appendix

Table 2. OLS Model and Fixed Effects Model for Depression Index

	OLS			Fixed Effects		
Characteristic	**Beta**	**95% CI**[1]	**p-value**	**Beta**	**95% CI**[1]	**p-value**
SMOKING						
2-No	—	—		—	—	
1-Yes	0.25	0.22, 0.28	<0.001	-0.09	-0.14, -0.04	<0.001
Gender						
1-Female	—	—				
2-Male	-0.66	-0.68, -0.64	<0.001			
Age of respondent	0.43	0.24, 0.61	<0.001			
Sports Frequency						
1-More than once per week	—	—				
2-One to Four times a month	0.01	-0.01, 0.04	0.3			
3-Hardly ever	0.65	0.62, 0.67	<0.001			
Job Situation						
2-Employed	—	—				
1-Retired	0.12	0.09, 0.15	<0.001			
3-Unemployed	0.69	0.66, 0.72	<0.001			
COUNTRY						
Austria	—	—				
Belgium	0.46	0.42, 0.51	<0.001			
Denmark	-0.04	-0.09, 0.02	0.2			
France	0.79	0.74, 0.84	<0.001			
Germany	0.30	0.25, 0.35	<0.001			
Greece	0.11	0.05, 0.16	<0.001			
Italy	0.66	0.61, 0.70	<0.001			
Netherlands	0.01	-0.05, 0.06	0.8			
Spain	0.38	0.33, 0.43	<0.001			
Sweden	0.11	0.06, 0.16	<0.001			
Switzerland	0.04	-0.01, 0.09	0.14			
WAVE						
1	—	—		—	—	
2	-0.10	-0.14, -0.06	<0.001	-0.08	-0.11, -0.05	<0.001
4	0.07	0.04, 0.11	<0.001	0.12	0.09, 0.16	<0.001
5	0.03	0.00, 0.06	0.083	0.08	0.04, 0.11	<0.001
6	0.07	0.03, 0.11	<0.001	0.16	0.12, 0.20	<0.001
7	-0.07	-0.12, -0.01	0.031	0.15	0.10, 0.20	<0.001
8	-0.03	-0.07, 0.01	0.2	0.09	0.05, 0.13	<0.001

[1] CI = Confidence Interval

References:

Boden, J. M., Fergusson, D. M., & Horwood, L. J. (2010). Cigarette smoking and depression: Tests of causal linkages using a longitudinal birth cohort. *British Journal of Psychiatry*, *196*(6), 440–446. https://doi.org/10.1192/bjp.bp.109.065912

Börsch-Supan, A. & S. Gruber (2020): easySHARE. Release version: 8.0.0. SHARE-ERIC. Dataset. doi: 10.6103/SHARE.easy.800

Khaled, S. M., Bulloch, A. G., Williams, J. V. A., Hill, J. C., Lavorato, D. H., & Patten, S. B. (2012). Persistent heavy smoking as risk factor for major depression (MD) incidence – Evidence from a longitudinal Canadian cohort of the National Population Health Survey. *Journal of Psychiatric Research*, *46*(4), 436–443. https://doi.org/10.1016/j.jpsychires.2011.11.011

Lerman, C., Caporaso, N., Main, D., Audrain, J., Boyd, N. R., Bowman, E. D., & Shields, P. G. (1998). Depression and self-medication with nicotine: The modifying influence of the dopamine D4 receptor gene. *Health Psychology*, *17*(1), 56–62. https://doi.org/10.1037/0278-6133.17.1.56

Paperwalla, K. N., Levin, T. T., Weiner, J., & Saravay, S. M. (2004). Smoking and depression. *Medical Clinics of North America*, *88*(6), 1483–1494. https://doi.org/10.1016/j.mcna.2004.06.007

Pasco, J. A., Williams, L. J., Jacka, F. N., Ng, F., Henry, M. J., Nicholson, G. C., Kotowicz, M. A., & Berk, M. (2008). Tobacco smoking as a risk factor for major depressive disorder: Population-based study. *British Journal of Psychiatry*, *193*(4), 322–326. https://doi.org/10.1192/bjp.bp.107.046706

Pomerleau, C. S. (1997). Co-factors for smoking and evolutionary psychobiology. *Addiction*, *92*(4), 397–408. https://doi.org/10.1111/j.1360-0443.1997.tb03371.x

Weinberger, A. H., Pilver, C. E., Desai, R. A., Mazure, C. M., & McKee, S. A. (2013). The relationship of dysthymia, minor depression, and gender to changes in smoking for current and former smokers: Longitudinal evaluation in the U.S. population. *Drug and Alcohol Dependence*, *127*(1–3), 170–176. https://doi.org/10.1016/j.drugalcdep.2012.06.028

R packages:

Croissant Y, Millo G (2008). “Panel Data Econometrics in R: The plm Package.” _Journal of Statistical Software_, *27*(2), 1-43. doi:10.18637/jss.v027.i02 <https://doi.org/10.18637/jss.v027.i02>.

Enzmann JAIRscadwbD, Schwartz M, Jain N, Kraft S (2023). _descr: Descriptive Statistics_. R package version 1.1.8, <https://CRAN.R-project.org/package=descr>.

Göran Broström (2022). Event History Analysis with R, Second Edition. Chapman and Hall/CRC, Boca Raton.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Rich B (2023). _table1: Tables of Descriptive Statistics in HTML_. R package version 1.4.3, <https://CRAN.R-project.org/package=table1>.

Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. Reproducible summary tables with the gtsummary package. The R Journal 2021;13:570–80. https://doi.org/10.32614/RJ-2021-053.

Wickham H, Miller E, Smith D (2023). _haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files_. R package version 2.5.4, <https://CRAN.R-project.org/package=haven>.

R code:

```
library(haven)
library(descr)
library(table1)
library(gtsummary)
library(ggplot2)
library(plm)


#### Step 1: Load data ####
rm(list=ls())

# easySHARE data
DATA <- read_dta("assets/easySHARE_rel8-0-0.dta")

# DATA01: Reduce the sample to Austria, Germany, Sweden, Netherlands,
# Spain, Italy, France, Denmark, Greece, Switzerland, Belgium
DATA01 <- subset(DATA,DATA$country>10 & DATA$country<25)
# Reduce the sample to respondents aged 50-79
DATA02 <- subset(DATA01,DATA01$age>=50 & DATA01$age<=79)
# Reduce the sample to cases with valid information on smoking status
DATA03 <- subset(DATA02,DATA02$smoking==1 | DATA02$smoking==5)  # 1=Yes, 5=No, wave 3
has no information on smoking
# Reduce the sample to cases with valid information on the dependent variable: Depression Index
DATA04 <- subset(DATA03, DATA03$eurod>=0)
# Reduce the sample to cases with valid information on the Sports Paticipation
DATA05 <- subset(DATA04, DATA04$br015_>0)
# Reduce the sample to cases with valid information on the Job Situation
DATA06 <- subset(DATA05, DATA05$ep005_>0 & DATA05$ep005_<10)

#### Step 2: Generate New Variables ####

#Var: Smoking
DATA06$SMOKING <- NA
DATA06$SMOKING[DATA06$smoking==1] <- "1-Yes"
DATA06$SMOKING[DATA06$smoking==5] <- "2-No"
DATA06$SMOKING <- as.factor(DATA06$SMOKING)
# choose yes as reference
DATA06$SMOKING <- relevel(DATA06$SMOKING, ref = "2-No")

#Var: Id & WAVE
DATA06$ID   <- as.factor(DATA06$mergeid)
DATA06$WAVE <- as.factor(DATA06$wave)

#Var: Continuous variables
DATA06$EUROD <- as.numeric(DATA06$eurod)
DATA06$AGE   <- as.numeric(DATA06$age)

#Var: Gender
DATA06$GENDER<- "NA"
DATA06$GENDER[DATA06$female==1]<-"1-Female"
DATA06$GENDER[DATA06$female==0]<-"2-Male"
DATA06$GENDER<-as.factor(DATA06$GENDER)

#Var: Country
DATA06$COUNTRY<-"NA"
DATA06$COUNTRY[DATA06$country==11]<-"Austria"
DATA06$COUNTRY[DATA06$country==12]<-"Germany"
DATA06$COUNTRY[DATA06$country==13]<-"Sweden"
DATA06$COUNTRY[DATA06$country==14]<-"Netherlands"
```

```
DATA06$COUNTRY[DATA06$country==15]<-"Spain"
DATA06$COUNTRY[DATA06$country==16]<-"Italy"
DATA06$COUNTRY[DATA06$country==17]<-"France"
DATA06$COUNTRY[DATA06$country==18]<-"Denmark"
DATA06$COUNTRY[DATA06$country==19]<-"Greece"
DATA06$COUNTRY[DATA06$country==20]<-"Switzerland"
DATA06$COUNTRY[DATA06$country==23]<-"Belgium"
DATA06$COUNTRY<-as.factor(DATA06$COUNTRY)

#Var: Sports Participation
DATA06$SPORTS <- 'NA'
DATA06$SPORTS[DATA06$br015_==1] <- "1-More than once per week"
DATA06$SPORTS[DATA06$br015_==2] <- "2-One to Four times a month"
DATA06$SPORTS[DATA06$br015_==3] <- "2-One to Four times a month"
DATA06$SPORTS[DATA06$br015_==4] <- "3-Hardly ever"
DATA06$SPORTS <- as.factor(DATA06$SPORTS)

#Var: Job Situation
DATA06$JOB <- 'NA'
DATA06$JOB[DATA06$ep005_==1] <- "2-Retired"
DATA06$JOB[DATA06$ep005_==2] <- "1-Employed"
DATA06$JOB[DATA06$ep005_==3] <- "3-Unemployed"
DATA06$JOB[DATA06$ep005_==4] <- "3-Unemployed"
DATA06$JOB[DATA06$ep005_==5] <- "3-Unemployed"
DATA06$JOB <- as.factor(DATA06$JOB)
# choose yes as reference
DATA06$JOB <- relevel(DATA06$JOB, ref = "1-Employed")

# Sample Statistics
table1::label(DATA06$EUROD) <- "DEPRESSION"
table1::label(DATA06$SMOKING) <- "SMOKING"
table1::label(DATA06$GENDER) <- "Gender"
table1::label(DATA06$AGE) <- "Age of respondent"
table1::label(DATA06$SPORTS) <- "Sports Frequency"
table1::label(DATA06$JOB) <- "Job Situation"

table1::table1(~ EUROD + SMOKING + GENDER + AGE | SMOKING, data = DATA06)

#### Step 3: analysis of the effect of smoking on depression ####

TABLE1 <- table(DATA06$SMOKING,DATA06$EUROD)
TABLE2 <- prop.table(TABLE1,1)
TABLE3 <- as.data.frame(TABLE2)
TABLE3$Percent <- TABLE3$Freq*100

ggplot(TABLE3, aes(fill=Var1, x=Var2, y=Percent)) +
  geom_bar(stat="identity", position=position_dodge(width=1)) +
  ylab("%") +
  xlab("Depression (EURO_D)") +
  scale_fill_manual(name = "Smoking status", labels = c("No smoking", "Smoking"), values = c("blue1",
"skyblue1"))

# Regression analysis
# AGE * 100 for better interpretation
DATA06$AGE100 <- DATA06$AGE/100
OLS1 <- lm(EUROD ~ SMOKING+GENDER+AGE100+SPORTS+JOB+COUNTRY+WAVE,
data=DATA06)

# Fixed Effects on
FE1 <- plm(EUROD ~ SMOKING+WAVE, data=DATA06, index=c("ID","WAVE"), model = "within")

OUTPUT_OLS <- tbl_regression(OLS1)
```

```
OUTPUT_FE <- tbl_regression(FE1)
tbl_merge( tbls = list(OUTPUT_OLS, OUTPUT_FE),
        tab_spanner = c("OLS", "Fixed Effects"))

# split the sample by GENDER
DATA_F <- subset(DATA06,DATA06$GENDER=="1-Female")
DATA_M <- subset(DATA06,DATA06$GENDER=="2-Male")

FE_F <- plm(EUROD ~ SMOKING+WAVE, data=DATA_F, index=c("ID","WAVE"), model = "within")
FE_M <- plm(EUROD ~ SMOKING+WAVE, data=DATA_M, index=c("ID","WAVE"), model = "within")

OUTPUT1 <- tbl_regression(FE_F)
OUTPUT2 <- tbl_regression(FE_M)
tbl_merge( tbls = list(OUTPUT1, OUTPUT2),
        tab_spanner = c("Female", "Male"))
```