

First and Last Name:

Student ID No.:

## Applied Microeconometrics using Stata

### Empirical Part

Prof. Dr. Uwe Jirjahn  
M. Sc. Mehrzad B. Baktash  
M. Sc. Thi Xuan Thu Le

Summer Semester 2023  
45 Points

**Task 1:** Please choose the correct answer(s) for the following questions, **without the help of Stata**. You may use Stata and move to Task 2 only after completing and handing over the Task 1.

*(Note: Multiple correct answers is possible. 2 Points/Question)*

**Q1.** Which of the following Stata commands can be used to calculate descriptive statistics for a variable?

- a. summarize      b. Tabulate      c. generate      d. codebook

**Q2.** When examining the relationship between two continuous variables in Stata, which command can be used to calculate the correlation coefficient?

- a. correlate      b. pwcorr      c. scatter      d. summarize

**Q3.** Which of the following commands can be used to assess multicollinearity in a regression model in Stata?

- a. collin      b. reg      c. vif      d. estat

**Q4.** When conducting panel data analysis in Stata, which command should be used to specify the panel structure?

- a. xtset      b. xtreg      c. paneldata      d. panelset

**Q5.** Which of the following commands is used in Stata to estimate a random effects panel regression model?

- a. xtreg, Re      b. mixed      c. xtreg, Fe      d. xtreg

**Q6.** Which of the following models do we use to estimate the determinants of the 'participating in the labor force' (PLF = 0,1)?

- a. Logit      b. Probit      c. Poisson      d. OLS

**Q7.** Which of the following codes are wrong?

- a. gen interaction=X1\*x2  
b. gen age2==age\*age  
c. drop if gender!=2  
d. keep if gender==.

First and Last Name:

Student ID No.:

**Task 2:** Please solve the following question **using Stata**. Create a **do-file** and start a **log-file** with your full name and matriculation number in the file names. Do not use the menu or the command line to create your solutions, as the results from these options are not reproducible. Instead, always use and store the appropriate commands in the do-file. Do not forget to close your log-file at the end of your do-file. After the exam, send me your do-file and log-file to **baktash@uni-trier.de**, I will let you know immediately whether I received your files.

Load the **exam\_task\_2.dta** dataset. It consists of a longitudinal data from workers residing in Germany. The variable names and definitions are as follows:

<i>Variable name</i>	<i>Definition</i>
id	Person identifier
gender	Male=1, Female=2, Other=3, Missing= -1
t	Survey year
income	Income in Euros
stress	Score of stress (1= Lowest, 10= Highest).
birth	The worker's year of birth.
education	Years of attained education.
experience	Work experience in years.
distance	The distance between permanent home and nearest educational institution.
parental	Parental education (0=Low educated, 1=Medium educated, 2=Highly educated).

Please solve the following:

- a. Generate a variable for the age of the workers, and label it as 'Worker Age.' (2p)
- b. Exclude cases with missing values from the analysis. (2p)
- c. What is the share of female workers who have highly educated parents and at least 10 years of education and work experience? (3p)

Now please make sure, if your dataset is ready for the following analysis.

- d. Estimate the returns to schooling:
  - a. Use an appropriate estimation strategy to estimate the returns to schooling, except instrumental variable approach. Explain shortly why this method is appropriate and any shortcomings associated with it. (9p)
  - b. Perform any necessary tests for (a). (2p)
  - c. Now, use an instrumental variable approach. Explain what advantage/disadvantage it has to your approach used in (a). (9p)
  - d. Perform any necessary tests for (c). (4p)