Exam in Panel and Evaluation Methods
Summer Term 2023

<u>**Remarks:**</u>

Grading:

- The exam consists of four problems.
- The total number of points is 60. The number of points for each problem is given in parentheses. It corresponds approximately to the recommended time spent on solving the problem (in minutes).

Important:

- Answers in German will be graded as well.
- If relevant information (necessary to solve a problem) is missing, make a plausible assumption for the missing item and briefly explain it in your answer.
- Whole sentences in your answers are not necessary, but your line of arguments should be clear and precise!

Problem 1: **[17.5 Points]**

The effect of smoking during pregnancy (S, smoking) on children's birth weight ($bweight$, logged and in kg) is represented by the following model $bweight_i = \beta_0 + \beta_1 S_i + \beta_2 mage + u_i$, where $mage$ denotes the age of the mother. The model is estimated with an ordinary least squares regression.

1.1 You estimate: $\hat{\beta}_1 = -0.15$, with an estimated standard error $se(\hat{\beta}_1) = 0.05$. Interpret the coefficient in terms of economic and statistical significance at the 5% level. (2.5 points)

1.2 Using an example, explain why the variable *smoking* during pregnancy might be endogenous. (2 points)

1.3 Assume that the variable *cigarette taxes* (T) is a valid instrument for S.

  1.3.1 Explain two characteristics of valid instruments. (2 points)

  1.3.2 State the moment condition(s) required for IV estimation. In which situation can you use the Sargan test to check the exogeneity of the instrumental variable? (2 points)

  1.3.3 Explain how to estimate the effect of smoking using a *two-stage-least-squares* estimator. (4 points)

  1.3.4 Assume that the variables T and $mage$ are exogenous, name a test that can be used to check whether the variable S is endogenous. Sketch the test procedure including the necessary estimation equations. (5 points)

Problem 2: **[14.5 Points]**

In 2008, the Swiss canton Zurich introduced a mandatory kindergarten. In the neighboring canton Zug, no mandatory kindergarten was implemented. You use a data set that includes a sample of individuals living in these cantons in the years 2007 and 2009 and the following variables:

$mearnings_{it}$	self-reported annual earnings of mother i in year t (in Swiss Franc)
$Zurich_{it}$	=1 if individual i in year t lives in Zurich; =0 otherwise
$post_{it}$	=1 if year is 2009; =0 otherwise

2.1 Which treatment effect do you identify with the Difference-in-Differences (DiD) estimation procedure? (1 point)

2.2 Write down a regression model to estimate the causal effect of the mandatory kindergarten on maternal earnings with a DiD estimation. (2 points)

2.3 Define the causal effect using only conditional expectations. (3 points)

2.4 Calculate the effect of mandatory kindergarten on maternal earnings using a DiD approach based on the following sample means of *mearnings*: (1.5 points)

	2007	2009
if $Zurich = 0$	30,000	31,500
if $Zurich = 1$	32,800	34,300

2.5 State and verbally explain the central assumption that has to hold to identify the causal effect of the mandatory kindergarten using the DiD method. Give an example in which the assumption would be violated. (4 points)

2.6 Verbally define the stable unit treatment value assumption (SUTVA). Briefly explain one reason why this assumption might not hold in this specific case. (3 points)

Problem 3: **[15.5 Points]**

You want to estimate the effect of health expenditures on hospital admissions based on a monthly panel data set for 45 countries from 2002-2022. The following variables are available:

$$\begin{aligned} admission_{it} &= \text{number of hospital admissions per capita in country } i \text{ in month } t \\ expenditure_{it} &= \text{health expenditures (in Euro) in country } i \text{ in month } t \\ age_{it} &= \text{average age of the population in country } i \text{ in month } t \end{aligned}$$

You estimate the following equation:

$$admission_{it} = \beta_0 + \beta_1 ln(expenditure_{it}) + \beta_2 age_{it} + \epsilon_{it}$$

OLS-, within- and random-effects estimations yield the following results (standard errors in parentheses):

	(1) OLS	(2) Within	(3) Random-Effects
Log Expenditures	-38.028***	-30.016***	-31.023***
	(5.083)	(7.003)	(6.952)
Age	0.802	0.754	0.765
	(0.675)	(0.712)	(0.695)
Constant	0.099	-1.720***	-0.760*
	(0.534)	(0.438)	(0.308)

3.1 Interpret the within-estimator for β_1 economically and statistically at the 5%-level. (2 points)

3.2 You assume that the number of general practitioners (*doctors*) is an important explanatory variable and add it to the initial estimation model. However, when performing a within-transformation, you find that the coefficient for *doctors* can not be estimated. Briefly explain why. (2 points)

3.3 Assume consistency of the within-estimator in the given example. Which additional assumption is required for a consistent random effects-estimator? Name and briefly explain the intuition behind the test of the validity of this assumption. (4.5 points)

3.4 The Hausman-Taylor-estimator can be considered as an alternative to the within-estimator. Explain the general approach of the Hausman-Taylor-estimator. (5 points)

3.5 Name the conditions under which the within-estimator is more efficient than the first difference estimator? (2 points)

Problem 4: **[12.5 Points]**

You want to estimate the causal effect of the school entry age on individuals' educational outcomes. Children being born on or before June 30 have to enter school in the current year (with 6 years), whereas children born after this cut-off have to attend school in the next year (with (almost) 7 years). Assume that parents can postpone the enrolment of their children in exceptional cases and that some make use of this. Your dataset includes the following information:

$days_i$	Number of days child i was born before the school entry date (= negative value) or after the school entry date (= positive value) respectively
$grade_i$	Average grade of child i after the third class on a scale from 1 (very good) to 5 (very bad)
$post_i$	=0 if child i was born before the cut-off; =1 if child i was born after the cut-off
age_i	School-entry age of child i in days

4.1 Which treatment effect do you identify with an Regression Discontinuity Design (RDD) in this setting? (1 point)

4.2 What is a sharp design in the context of regression discontinuity design? What is meant by a fuzzy design? Which design would you prefer in the given setting? Please provide a brief explanation for your choice. (3 points)

4.3 What is the central assumption of the RDD approach to identify the causal effect of the school entry age? Give an example where the assumption is violated. (3 points)

4.4 Give a brief definition of running variable. Which of the variables in your data set is the running variable in this specific example? (1.5 points)

4.5 Mention two characteristics (advantages or disadvantages) of the RDD approach. (2 points)

4.6 Explain how a regression kink design differs from a regression discontinuity design. (2 points)