Chair for Statistics and Empirical Economics
Prof. Regina T. Riphahn, Ph.D.

Exam in Panel and Evaluation Methods
Winter Term 2020/2021

Remarks:

Number of tasks:

- The exam consists of 4 problems.

Grading:

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

Important:

- Answers in German will be graded as well.
- If relevant information (necessary to solve a task) is missing, make a plausible assumption for the missing item and briefly explain it in your answer.
- Most sub-questions can be solved independent of each other. Even if you are not able to answer a certain subquestion, you nevertheless might be able to answer the following parts of the question.
- Whole sentences in your answers are not necessary, but your line of arguments should be accessible, clear and precise! Please stick to the point and avoid general discussions that are only loosely related to the question.

Problem 1 (14 Points)

Between 1994 and 1999, the state of Bavaria implemented the *Future of Bavaria Offensive*. You want to estimate the effect of the policy on R&D (research and development) expenditure of firms located in Bavaria. You have data for the years 1993 and 2000 on firms located in Bavaria and firms located in Hesse. Your data set contains the following variables:

$rdexp_{it}$	R&D expenditure of firm i in year t
$Bavaria_{it}$	=1 if firm i in year t is located in Bavaria; =0 otherwise
$after_{it}$	=1 if year is 2000; =0 otherwise

1.1 Write down a regression model to estimate the causal effect of the policy on R&D expenditure with a Difference-in-Differences (DiD) estimation. (2 points)

1.2 Define the effect using conditional expectations. (3 points)

1.3 State and verbally explain the central assumption that has to hold to identify the causal effect of the policy using the DiD method. Give two examples in which the assumption would be violated. (5 points)

1.4 Briefly state (without abbreviation) what kind of treatment effect is identified by a consistent DiD estimation. (1 point)

1.5 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 2 (16 Points)

You estimate the effect of GDP per capita on gun-related deaths using a simple linear regression model and using country-level panel data.

2.1 State the assumption which has to be fulfilled for the consistency of the OLS estimator. Give a specific example in which this assumption is violated in the problem under study. (2 points)

2.2 Verbally explain the within-estimator for this specific case. (3 points)

2.3 Briefly explain an alternative estimator which yields the same point estimates as the within-estimator. What is the name of this estimator? (2 points)

2.4 Explain what kind of problem (of OLS) is solved by the within-estimator in this specific example. (2 points)

2.5 Assume that the within-estimator is consistent in this example. Under which additional assumption is the random effects-estimator consistent as well? (2 points)

2.6 You want to perform a Hausman test of the consistency of the random effects estimator. What is the the number of degrees of freedom of the test statistic in this case? (2 points)

2.7 You regress gun-related deaths per 100,000 inhabitants (*gundeaths*) on GDP per capita, measured in 1,000 Dollars (*gdpcapita*). Applying the within-estimator to a panel of countries, you obtain the following result:

gundeaths	Coef.	Std. Err.	t	[95% Conf.	Interval]
gdpcapita	-.1.2732	.3685	-3.46	-1.9955	-.5509

Interpret the coefficient for *gdpcapita* in terms of magnitude. Briefly explain if the estimated coefficient is statistically significant at the 5%-level. (2 points)

2.8 Suppose that in the estimation of 2.7 you additionally include a dummy variable for the USA, which is one of the countries included in your panel data. How does this affect the estimated coefficient of *gdpcapita*? Briefly explain. (1 point)

Problem 3 (16 Points)

Using a sample of primary school students (grade 1-4), you estimate the effect of class size (w) on students' math performance (y) measured in points (0-100) using a simple linear regression model.

3.1 Using a specific example, explain why class size could be an endogenous variable. What would be the nature of the potential bias in an OLS estimation? Briefly explain your answer (you might use a graphical illustration). (3 points)

3.2 Consider the grade-specific average class size (z_1) and the grade-specific number of teachers (z_2) as potential instruments for class size (w).

3.2.1 Name and briefly explain two conditions that z_1 and z_2 have to fulfill to be valid instruments. How can you test each of them? (4 points)

3.2.2 You reject z_2 as a valid instrument and proceed with z_1. Describe verbally the steps of a *two-stage-least-squares* estimation using z_1 as an instrument for class size. (4 points)

3.2.3 Next, you perform an *indirect-least-squares* estimation using z_1 as an instrument for w. The Stata output yields a first-stage effect of 0.8 and a reduced-form effect of -0.1. Provide a formal notation for the two estimated equations. Define formally the ISL estimator, calculate its value, and interpret its magnitude. (5 points)

Problem 4 (14 Points)

Consider the following model to estimate the gender pay gap

$$log(wage)_i = \beta_0 + \beta_1 female_i + u_i \tag{1}$$

where $log(wage)_i$ is the natural logarithm of hourly wage of a person i and $female_i$ is an indicator taking a value of 1 for women and 0 for men. The following table shows the estimates for β_1 obtained using quantile regressions (where q indicates the relevant quantile).

	(1)	(2)	(3)	(4)	(5)
q	0.1	0.25	0.5	0.75	0.9
female	-0.055	-0.071	-0.101	-0.138	-0.168

4.1 Interpret the estimated coefficient at the 25th percentile. (1 point)

4.2 Name two advantages of quantile regressions compared to OLS. (2 points)

4.3 Explain the relationship between the LAD estimator and quantile regressions. Show formally the loss function of the LAD estimator. (2 points)

4.4 Are the wages (conditional on x_i) for women more unequally distributed than those for men? Explain briefly. (2 points)

4.5 You are worried about a potential bias due to unobserved heterogeneity and decide to apply a radius-caliper matching to make the subsamples of men and women more comparable. For this purpose, you use a vector of characteristics x_i that includes age, education, and occupation. Define the concept of "propensity score" and provide the estimation equation necessary to estimate propensity score in this particular example. (2 points)

4.6 Define the term "caliper". What trade-off exists when deciding about the size of the caliper? (5 points)