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

Exam in Panel and Evaluation Methods
Winter Term 2023/24

Remarks:

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: **[16.5 Points]**

You use OLS to analyze the effect of being hit by a natural disaster on support for environmental policies. You have survey data on individuals living in New Orleans and in Seattle for the years 2004 and 2006. In 2005, New Orleans was severely hit by Hurricane Katrina, while Seattle was not. Your dataset contains the following variables:

$support_{it}$	self-reported support for environmental policies of individual i in year t on a range from 1 (no support) to 5 (full support)
$NewO_{it}$	=1 if individual i in year t lives in New Orleans; =0 otherwise
$post_{it}$	=1 if year is 2006; =0 otherwise

1.1 Write down a regression model to estimate the effect of the hurricane on support for environmental policies with a Difference-in-Differences (DiD) estimation. The coefficient of which variable reflects the estimated treatment effect? What do the other explanatory variables control for? (5 points)

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

1.3 Calculate the estimated effect of the hurricane on support for environmental policies using a Difference-in-Differences (DiD) approach based on the following sample means of *support*: (1.5 points)

	2004	2006
if $NewO = 0$	2.72	3.02
if $NewO = 1$	2.20	2.88

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

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: **[13 Points]**

You use balanced panel data on 2,120 German cities to estimate the relationship between the local tax rate (x) and average local wages (y). In total, the data set contains 4,240 observations and no other variables are available.

2.1 Briefly explain how you could perform a *least-squares-dummy-variables* (LSDV) estimation with the available data. How many parameters would be estimated for this regression model? (2 Points)

2.2 In the present problem, two other estimation methods can provide the same parameter estimates as the LSDV estimator. Name the two estimators and explain the data transformation required for each. (4 Points)

2.3 State and explain the condition of strict exogeneity under which the within estimator yields consistent estimates. Is the condition fulfilled in this setting? (3 Points)

2.4 A Hausman test has been applied to determine whether a fixed-effects or a random-effects model should be used for estimation. State the null and alternative hypotheses and make a test decision based on the following test results. Do the results support the use of the random-effects estimator? (2 Points)

```
chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
        =        0.47
Prob>chi2 =      0.3288
```

2.5 You used uncorrected standard errors. Name one disadvantage of this approach compared to cluster-robust standard errors and a consequence for your inference. (2 Points)

Problem 3: **[19 Points]**

You want to estimate the effect having foreign-born coworkers on attitudes towards immigration. Your dataset contains the following variables:

$foreign_i$	Share of foreign-born workers (in % from 0 to 100) in the firm in which individual i is employed
$immigration_i$	Attitude towards immigrants of individual i on a range from 1 (very negative) to 5 (very positive)
$dist_i$	Distance from individual i's workplace to the nearest border, measured in kilometres

You estimate the following model using OLS and obtain the regression results below:
$immigration_i = \beta_0 + \beta_1 foreign_i + u_i$.

immigration	Coef.	[95% Conf.	Interval]
foreign	.2400	.1195	.3605

3.1 Interpret the coefficient for $foreign$ in terms of magnitude. Briefly explain whether the estimated coefficient is statistically significant at the 5%-level. (2 points)

3.2 Briefly explain the concept of omitted variable bias in general. Name one additional control variable that you would include, if available, in order to reduce the potential problem of omitted variable bias in this case and explain your choice. Do not use the variable $dist_i$. (3 Points)

3.3 You now use the variable $dist_i$ as an instrument for $foreign_i$.

3.3.1 List two conditions that valid instruments have to meet. (2 Points)

3.3.2 State the new moment condition for the IV estimation, using $dist_i$ as an instrument for $foreign_i$. (2 points)

3.3.3 Verbally explain how you estimate the effect of effect of foreign-born coworkers on attitudes towards immigration using a *two-stage-least-squares* (2SLS) estimation. Which coefficient gives you the 2SLS estimate? (4 Points)

3.3.4 Name two problems that arise when instruments are only weakly correlated with the endogenous variable. (2 points)

3.3.5 Verbally explain how you carry out an *indirect-least-squares* (ILS) estimation of the effect of foreign-born coworkers on attitudes towards immigration. Which coefficient gives you the ILS estimate? (4 Points)

Problem 4 (11.5 Points)

You want to analyze the effect of robot technologies on workers's wages. You have data on a large number of firms which applied for a technology subsidy program. The applications of the firms were rated on a scale from 0 (worst rating) to 100 (best rating) and all firms with a rating of at least 80 were granted the subsidy which covers a large part of the costs of implementing the robot technology. You construct a dataset containing the following variables:

$wage_i$	Average wage in firm i (in EUR)
$rating_i$	Rating of firm i's application, ranging from 0 (worst) to 100 (best)
$robot_i$	=1 if firm i has implemented the robot technology; =0 otherwise

4.1 You evaluate the effects of the robot technology on wages using a parametric regression discontinuity design (RD). Briefly describe your approach using the data and establish the relevant estimation equation. If necessary, define new variables. (3.5 points)

4.2 Give a brief definition of the *running variable*. Which of the variables in your data set is the running variable in this specific example? Briefly explain the data transformation needed to set the running variable to zero at the relevant cutoff. (2.5 points)

4.3 State the main identifying assumption of your RD approach to identify the causal effect of the robot technology on wages. (1.5 points)

4.4 A descriptive analysis yields the following output:

```
. sum wage robot if rating>=80

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |        51    69.84449   58.987440       14.9      156.7
       robot |        51      .71289     .216227          0          1

 . sum wage robot if  rating<80

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       229    90.54534   60.012834       22.1      161.4
       robot |       229      .02040     .221511          0          1
```

Given these results, explain briefly which RD design (sharp or fuzzy) would be appropriate in this case. What treatment effect would you identify with the chosen design? Calculate the effect using the output above. *Hint: Round the result to <u>two</u> decimal places.* (3 points)

4.5 Briefly explain the concept of *Statistical Power*. (1 Point)