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

Exam in Panel and Evaluation Methods
Winter Term 2018/19

Problem 1 (24 points)

You aim to estimate the effect of maternal education on children's mental health. Your data contains the following variables:

mh_i	mental health measure (standardized with mean 0 and standard deviation 1).
$meduc_i$	maternal education (in years of schooling).
yob_i	maternal year of birth.
$reform_i$	=1 if mother was affected by a compulsory schooling reform raising minimum school leaving age to 15; =0 otherwise.

Note: only parents born after a cut-off date were affected by the compulsory schooling reform.

1.1 You start out with a simple model using ordinary least squares:

$$mh_i = \beta_1 + \beta_2 meduc_i + \beta_3 yob_i + e_i$$

Name two different potential sources of endogeneity for maternal education. Discuss whether these sources are likely to apply in this case. (4 points)

1.2 You intend to use a regression discontinuity design (RDD) to estimate the causal effect of maternal education on children's mental health.

1.2.1 Write down the estimation equation using RDD. (2 points)

1.2.2 What kind of treatment effect does the RDD estimator identify? (1 point)

1.2.3 State two general disadvantages of the RDD approach. (2 points)

1.3 You intend to use an instrumental variables (IV) estimator to estimate the causal effect of maternal education on children's mental health.

1.3.1 Which variable available in the data could you use to instrument maternal education? Write down the two equations for the two-stage-least-squares (2SLS) estimator and define new variables if needed. (3 points)

1.3.2 What kind of treatment effect does IV identify? Which group of observations identifies the treatment effect? What characterizes this group? (3 points)

1.3.3 In your 2SLS estimation, you estimate a coefficient of 0.400 for your instrument in your first stage, with a standard error of 0.05. Define briefly what is meant by weak instruments. Does your analysis have a weak instruments problem? (2 points)

1.3.4 Can you use the Sargan test in this case to assess the instrument's exogeneity given the available data? Explain shortly. (1 point)

1.4 You also examine the effect of maternal education on children's mental health using a conditional quantile regression analysis.

The following table shows the results from a linear regression model in column (1). Columns (2) to (6) contain the coefficients for the quantile regressions.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS			quantiles		
		10%	25%	50%	75%	90%
$\hat{\beta}_{meduc}$	0.116	0.232	0.215	0.126	0.050	0.002
S.E.($\hat{\beta}_{meduc}$)	(0.034)	(0.070)	(0.048)	(0.050)	(0.090)	(0.047)

1.4.1 Interpret the estimated OLS coefficient and the coefficient for the 10th quantile economically and statistically at the 5% significance level. (5 points)

1.4.2 What can we learn from a comparison of the coefficients at the 10th and 90th quantile? (2 points)

Problem 2 (6 points)

2.1 Define the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT) formally. Use the notation of the potential outcomes framework, which indicates the treatment indicator by D_i, the potential outcome with treatment as $y_i(1)$ and the potential outcome without treatment as $y_i(0)$. (3 points)

2.2 Show how estimating the ATT using only observable means potentially biases the ATT. Briefly explain in which situation this ATT estimate would be unbiased. (3 points)

Problem 3 (12 points)

In 2010 the German federal state of Baden-Württemberg (BW) banned the sale of alcohol in gas stations and kiosks between 10 p.m. and 5 a.m. In Bavaria alcohol sales remained legal around the clock. The aim of the sales ban in BW was to reduce binge drinking and subsequent alcohol poisoning among young adults. You have data from all municipalities in BW and Bavaria from 2007 and 2011. Your data set contains the following variables:

$hosp_{it}$	yearly rate of hospitalization due to alcohol poisoning of inhabitants aged 14 to 21.
BW_{it}	=1 if municipality is in Baden-Württemberg; =0 otherwise.
$after_{it}$	=1 if the year is 2011; =0 otherwise.

3.1 Write down a regression model to estimate the causal effect of the alcohol sales ban on the hospitalization rate with a Difference-in-Difference (DiD) estimation. (2 points)

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

3.3 State and verbally explain the central assumption that has to hold for the DiD method to identify the causal effect of the alcohol sales ban. Give an example in which the assumption would be violated. (3 points)

3.4 Briefly state what kind of treatment effect is identified by a consistent DiD estimation. (1 point)

3.5 Define the stable unit treatment value assumption (SUTVA) and discuss briefly whether it is likely to hold in this scenario. (3 points)

Problem 4 (7 points)

4.1 State and briefly explain the implications for efficiency and bias of the effect estimate if propensity score matching is implemented with a caliper instead of without a caliper. (2 points)

4.2 Verbally define the strong Conditional Independence Assumption (CIA). (2 points)

4.3 If you are concerned that unobservables are still unbalanced after matching, how can you improve your matching procedure? Briefly describe three strategies. (3 points)

Problem 5 (11 points)

You analyze the relationship between private tutoring (x_{it} = hours of tutoring in one school year) and school grades (y_{it}) using a balanced panel of 1000 students who attend the same school. The data includes annual observations from 2013 to 2015.

5.1 Explain briefly how a least-squares-dummy-variables (LSDV) estimation would be implemented in this case. How many parameters would be estimated? (2 points)

5.2 You consider using either a Fixed-Effects or a Random-Effects model and, to this end, conduct a Hausman-Test at the 10% significance level with a test statistic of 2.84. Write down the hypotheses, degrees of freedom, critical value and test decision. (3 points)

5.3 What is an advantage of a Fixed-Effects against a Random-Effects estimation? (1 points)

5.4 Consider the following dynamic model for the regression of grades on tutoring:

$$y_{it} = x'_{it}\beta + \gamma_1 y_{i,t-1} + \alpha_i + \varepsilon_{it}$$

Set up the estimation model for the Anderson-Hsiao (AH) estimator in this case and describe the instrumentation the estimator uses. How many observations are used in the final estimation? (3 points)

5.5 Explain why a first-order autoregressive process (AR1-process) in ε_{it} would be a problem for the AH estimator. (2 points)

Percentiles of the *t*-distribution

Cell entries: x, such that $\text{Prob}[t_n \leq x] = P$, with n degrees of freedom

n \ P	0.75	0.9	0.95	0.975	0.99	0.995
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.817	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.500
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.813	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.696	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.320	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
35	0.682	1.306	1.690	2.030	2.438	2.724
40	0.681	1.303	1.684	2.021	2.423	2.705
45	0.680	1.301	1.679	2.014	2.412	2.690
50	0.679	1.299	1.676	2.009	2.403	2.678
60	0.679	1.296	1.671	2.000	2.390	2.660
70	0.678	1.294	1.667	1.994	2.381	2.648
80	0.678	1.292	1.664	1.990	2.374	2.639
90	0.677	1.291	1.662	1.987	2.368	2.632
100	0.677	1.290	1.660	1.984	2.364	2.626
∞	0.674	1.282	1.645	1.960	2.326	2.576

Source: Generated in R

Percentiles of the χ^2-distribution

Cell entries: c, such that Prob$\left[\chi_n^2 \leq c\right] = P$, with n degrees of freedom

n \ P	0.005	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	0.995
1	0.00004	0.0002	0.001	0.004	0.016	0.102	0.455	1.323	2.706	3.842	5.024	6.635	7.879
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
35	17.19	18.51	20.57	22.47	24.80	29.05	34.34	40.22	46.06	49.80	53.20	57.34	60.27
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77
45	24.31	25.90	28.37	30.61	33.35	38.29	44.34	50.98	57.51	61.66	65.41	69.96	73.17
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49

Source: Generated in R