NTNU
Kunnskap for en bedre verden

Institutt for samfunnsøkonomi

Eksamensoppgave i SØK3514/SØK8614
Anvendt Økonometri/Applied Econometrics

Faglig kontakt under eksamen: Kåre Johansen
Tlf.: 73 59 19 36

Eksamensdato:	30. november 2019
Eksamenstid (fra-til):	6 timer (09.00-15.00)
Hjelpemiddelkode	C

Formelsamling:

Knut Sydsæter, Arne Strøm og Peter Berck (2006): Matematisk formelsamling for økonomer, 4utg. Gyldendal akademiske.

Knut Sydsæter, Arne Strøm, og Peter Berck (2005): Economists' mathematical manual, Berlin.

Kalkulator:

Casio fx-82ES PLUS, Casio fx-82EX Citizen SR-270x, SR-270X College eller HP 30S.

Målform/språk:	Norsk og engelsk
Antall sider norsk (uten forside):	3
Antall sider engelsk (uten forside):	3

Informasjon om trykking av eksamensoppgave

Originalen er:

1-sidig ☐ **2-sidig** ☐

sort/hvit ☐ **farger** ☐

skal ha flervalgskjema □

Kontrollert av:

Dato Sign

Oppgave 1

Artikkelen "Education and Catch-up in the Industrial Revolution" av Becker, Hornung og Woessman undersøker empirisk sammenhengen mellom utdanningsnivå og industrialisering i forskjellige regioner i Preussen.

a) Forklar hypotesene som forfatterne ønsker å teste, hva er de viktigste økonometriske utfordringene i den forbindelse og hvilken strategi forfatterne bruker for å håndtere disse økonometriske utfordringene.

b) Tabell 1 rapporterer utvalgte estimeringsresultater for forholdet mellom utdanningsnivå og et mål for industrialisering (andelen fabrikkarbeidere) i 1849. Forklar hvilke spesifikasjoner som er estimert, hvilke metoder som brukes og diskuter resultatene gitt i tabellen. Hva kan vi si om styrken til instrumentvariabelen som brukes ved estimering av ligning (3) og (5)? Forklar også hvorfor forfatterne bruker kontrollvariabler datert (rundt) 1816 i (4) og (5).

Table 1: Education and industrialization 1849

Dependent variable	(1) OLS	(2) First stage	(3) IV	(4) First stage	(5) IV
Years of schooling 1849	0,177 (0,077)		0,132 (0,077)		0,187 (0,080)
School enrollment rate 1816		0,061 (0,001)		0,060 (0,001)	
Control variables 1849	Yes	Yes	Yes	Yes	Yes
Control variables 1816	No	No	No	Yes	Yes
Observations	334	334	334	334	334
First stage F-statistic		6206		5507	

Notes: Results for all factories. Dependent variable in (1), (3) and (5) is Share of factory workers in total population 1849. Dependent variable in (2) and (4) is Years of schooling 1849. The Table reports estimated effects with standard errors in parentheses.

c) Tabell 2 rapporterer resultater for utviklingen av industrialiseringen mellom 1849 og 1882. Utfallsvariabelen er nå et mål på industrialisering i 1882. Diskuter og tolk resultatene ved bruk av informasjonen gitt i tabellen.

Table 2: Education and industrialization 1882

Dependent variable	(6) First stage	(7) IV
Literacy rate1871		0,101 (0,036)
School enrollment rate 1816	0,301 (0,037)	
Share of factory workers in total population 1849	1,183 (0,312)	0,923 (0,168)
Control variables 1849	Yes	Yes
Control variables 1816	Yes	Yes
Observations	334	334
First stage F-statistic	65	

Notes: Results for all manufacturing. Dependent variable in (6) is Literacy rate 1871. Dependent variable in (7) is the Share of manufacturing workers in total population 1882. The Table reports estimated effects with standard errors in parentheses.

Oppgave 2

a) Forklar hvordan strategien "Regression Discontinuity Design" (RDD) kan brukes til å estimere kausale effekter i økonometriske studier og diskuter forskjellen mellom "sharp design" og "fuzzy design".

b) Artikkelen "Does the size of the legislature affect the size of government? Evidence from two countries" av Petterson-Lidbom inneholder en empirisk studie av sammenhengen mellom kommunale utgifter og størrelsen på kommunestyret målt med antall medlemmer i kommunestyrene i Finland og Sverige. Forklar hvordan Petterson-Lidbom i sin empiriske strategi benytter seg av institusjonelle ordninger om hvordan størrelsen på kommunestyrene blir bestemt i Finland og Sverige. Diskuter modellspesifikasjoner, estimeringsmetoder og datautvalg brukt i studien.

c) Pensumlisten inneholder andre anvendelser av RDD-strategien. Ta utgangspunkt i noen av disse anvendelsene, forklar hvorfor RDD er et fornuftig valg av empirisk strategi, forklar hvilke av disse som er «sharp» design og hvilke som er «fuzzy» design og pek kort på de viktigste forskjellene og likhetene mellom de forskjellige anvendelsene.

Oppgave 3

a) Artikkelen av Bratsberg og Raaum "Immigration and wages: Evidence from construction" undersøker effekten av arbeidsinnvandring på lønnsnivået for norske arbeidere i bygg og anlegg. Diskuter økonometriske utfordringer i en empirisk analyse av effekten av arbeidsinnvandring på lønn for innfødte arbeidere og forklar den empiriske strategien som brukes av Bratsberg og Raaum for å håndtere disse økonometriske utfordringene.

b) Tabell 3 rapporterer resultater fra estimering av forskjellige varianter av en lønnsligning basert på individuelle data 1995-2005 for norske arbeidere. Bruk informasjonen gitt i tabellen og diskuter årsaker til at den estimerte effekten av arbeidsinnvandring er så forskjellig mellom de forskjellige regresjonene, avhengig av modellspesifikasjoner og utvalg som benyttes. Hvilken variant gir det mest troverdige estimatet på kausaleffekten av arbeidsinnvandring på lønn for innfødte (norske) arbeidere?

Table 3: OLS-estimates of the effect of immigrant labor on native log wage

	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient of ln(1+M/N)	-0.103 (0.16)	-0.724 (0.20)	-0.554 (0.18)	-0.570 (0.18)	-0.032 (0.18)	-0.569 (0.18)
Number of observations	918 082	918 082	296 152	296 152	843 567	840 747
Individual fixed effects	No	Yes	No	Yes	No	No
Sample used	Full sample	Full sample	Balanced panel	Balanced panel	Drop entrants	Drop leavers

Notes: The table reports estimated coefficient with standard errors in parentheses. The outcome variable is the log of individual wages for Norwegian workers. M is the number of immigrant workers and N is the number of Norwegian workers in 16 different activities in the construction sector. The sample in column (3) and (4) is restricted to individuals who are included in the sample all eight years and who remain employed in the same activity all eight years. The sample in column (5) drops entrants – workers who did not hold a construction job during the first half of the sample period but entered the sector after 2001. The sample in column (6) drops leavers – worker who were present in the data during the first half of the sample period and who ended their construction job before 2002. All regressions control for age (third order polynomial), years of schooling, gender, activity, and year of observation.

c) En kommentator foreslår at forfatterne kan definere en dummyvariabel for «entrants» og en dummyvariabel for «leavers». Definer deretter interaksjonsledd mellom disse dummyvariablene og ln(1 + M/N) og estimer lønnsligningene utvidet med disse nye variablene ved bruk av hele utvalget i stedet for å estimere lønnsligningen for forskjellige delutvalg. Hvordan ville estimeringsresultatene fra en slik modell mest sannsynlig blitt sammenlignet med de som er rapportert i tabell 3? Har det betydning om den utvidede modellen inkluderer individspesifikke faste effekter?

English

Question 1

The article "Education and Catch-up in the Industrial Revolution" by Becker, Hornung and Woessman investigates empirically the relationship between educational level and industrialization in different regions within Prussia.

a) Explain the hypotheses that the authors want to test, what are the most important econometric challenges in this regard and what strategy the authors use to handle these econometric challenges.

b) Table 1 reports selected estimation results for the relationship between educational level and a measure of industrialization (the share of factory workers) in 1849. Explain which specifications are estimated, which methods are used and discuss the results given in the table. What can we say about the strength of the instrumental variable used in estimating equations (3) and (5)? Also explain why the authors use control variables dated (around) 1816 in (4) and (5).

Table 1: Education and industrialization 1849

Dependent variable	(1) OLS	(2) First stage	(3) IV	(4) First stage	(5) IV
Years of schooling 1849	0,177 (0,077)		0,132 (0,077)		0,187 (0,080)
School enrollment rate 1816		0,061 (0,001)		0,060 (0,001)	
Control variables 1849	Yes	Yes	Yes	Yes	Yes
Control variables 1816	No	No	No	Yes	Yes
Observations	334	334	334	334	334
First stage F-statistic		6206		5507	

Notes: Results for all factories. Dependent variable in (1), (3) and (5) is Share of factory workers in total population 1849. Dependent variable in (2) and (4) is Years of schooling 1849. The Table reports estimated effects with standard errors in parentheses.

c) Table 2 reports results for the progress of industrialization between 1849 and 1882. The outcome variable is now a measure of industrialization in 1882. Discuss and interpret the results using the information given in the table.

Table 2: Education and industrialization 1882

Dependent variable	(6) First stage	(7) IV
Literacy rate1871		0,101 (0,036)
School enrollment rate 1816	0,301 (0,037)	
Share of factory workers in total population 1849	1,183 (0,312)	0,923 (0,168)
Control variables 1849	Yes	Yes
Control variables 1816	Yes	Yes
Observations	334	334
First stage F-statistic	65	

Notes: Results for all manufacturing. Dependent variable in (6) is Literacy rate 1871. Dependent variable in (7) is the Share of manufacturing workers in total population 1882. The Table reports estimated effects with standard errors in parentheses.

Question 2

a) Explain how the strategy "Regression Discontinuity Design" (RDD) can be used to estimate causal effects in econometric studies and discuss the difference between "sharp design" and "fuzzy design".

b) The article "Does the size of the legislature affect the size of government? Evidence from two countries" by Petterson-Lidbom contains an empirical study of the relationship between local government expenditures and the size of the local government council measured by the number of council members in Finland and Sweden. Explain how Petterson-Lidbom in his empirical strategy utilizes institutional arrangements about how the size of local government councils is determined in Finland and Sweden. Discuss model specifications, estimation methods and data samples used in the study.

c) The reading list contains other applications of the RDD strategy. Take as a starting point some of these applications, explain why RDD is a sensible choice of empirical strategy, explain which of these are sharp designs and which are fuzzy designs and briefly point to the most important differences and similarities between the different applications.

Question 3

a) The article by Bratsberg and Raaum "Immigration and wages: Evidence from construction" investigates the effects of immigrant labor on the wage level for Norwegian workers in the construction sector. Discuss econometric challenges in an empirical analysis of the effect of immigrant labor on wages for native workers and explain the empirical strategy used by Bratsberg and Raaum to handle these econometric challenges.

b) Table 3 reports results from estimating different variants of a wage equation based on individual data 1995-2005 for Norwegian workers. Use the information given in the table and discuss reasons why the estimated effect of labor immigration is so different between the different regressions depending on model specifications and samples used. Which variant provides the most credible estimate of the causal effect of immigrant labor on native wages?

Table 3: OLS-estimates of the effect of immigrant labor on native log wage

	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient of ln(1+M/N)	-0.103 (0.16)	-0.724 (0.20)	-0.554 (0.18)	-0.570 (0.18)	-0.032 (0.18)	-0.569 (0.18)
Number of observations	918 082	918 082	296 152	296 152	843 567	840 747
Individual fixed effects	No	Yes	No	Yes	No	No
Sample used	Full sample	Full sample	Balanced panel	Balanced panel	Drop entrants	Drop leavers

Notes: The table reports estimated coefficient with standard errors in parentheses. The outcome variable is the log of individual wages for Norwegian workers. M is the number of immigrant workers and N is the number of Norwegian workers in 16 different activities in the construction sector. The sample in column (3) and (4) is restricted to individuals who are included in the sample all eight years and who remain employed in the same activity all eight years. The sample in column (5) drops entrants – workers who did not hold a construction job during the first half of the sample period but entered the sector after 2001. The sample in column (6) drops leavers – worker who were present in the data during the first half of the sample period and who ended their construction job before 2002. All regressions control for age (third order polynomial), years of schooling, gender, activity, and year of observation.

c) A commentator suggests that the authors could define a dummy variable for entrants and a dummy variable for leavers. Then define interaction terms between these dummy variables and ln(1+M/N) and estimate wage equations expanded with these new variables using the full sample instead of estimating the wage equation for different sub-samples. How would the estimation results from such a model most likely have been compared to those reported in Table 3? Does it matter whether the expanded model includes individual fixed effects?