

Institutt for samfunnsøkonomi

Eksamensoppgave i SØK3001 – Økonometri I / Econometrics I

Faglig kontakt under eksamen: Jørn Rattsø

Tlf.: 73 59 19 34

Eksamensdato: 29. mai 2013

Eksamenstid: 5 timer

Sensurdato: 19. juni 2013

Tillatte hjelpemidler: Flg 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.
Enkel kalkulator Citizen SR-270x, HP 30S eller SR-270X College

Annen informasjon: -

Målform/språk: Bokmål, nynorsk og engelsk

Antall sider: 10 (inkl. forside)

Antall sider vedlegg: 2

Bokmål

Alle spørsmål skal besvares.
To tabeller er vedlagt.

Oppgave 1

Den nedenstående regresjonsmodell er estimert med data for individer fra folketellingen i USA i år 2000 (805 observasjoner)

$$IWage_i = \beta_0 + \beta_1 Exp_i + \beta_2 Exp_i^2 + \beta_3 Edyrs_i + \beta_4 Part_i + \beta_5 Female_i + \beta_6 Black_i + u_i$$

hvo IWage er log timelønn, Exp er arbeidsmarkedserfaring i år, Exp^2 er erfaring opphøyd i andre, Edyrs er antall år utdanning, Part er en dummy variable lik 1 hvis individet arbeider deltid, Female er en dummy variabel lik 1 hvis individet er kvinne, Black er en dummy variabel som er lik 1 hvis individet er sort; alle variable er målt for individ j.

Regresjonsestimater og noe statistikk ('R i andre', Korrigert 'R i andre', og summen av kvadrerte avvik SSR) er vist i tabellen under:

IWage	Koeffisient	St. avvik
Exp	0.037	0.0063
Exp^2	-0.00049	0.00015
Edyrs	0.091	0.0095
Part	0.0030	0.0581
Female	-0.225	0.0427
Black	-0.170	0.0590
Const	0.923	
'R i andre'	0.2308	
Korr. 'R i andre'	0.2251	
SSR	282.584	

- Tolk koeffisientene, kommenter deres fortegn, størrelse og signifikans. Hva er de viktige faktorene som forklarer lønnsforskjeller gitt denne analysen?
- Etter hvor mange år i arbeidsmarkedet begynner individets lønn å falle?
- Vis hvordan hypotesen om at koeffisienten for Female er lik -0.1 kan testes og gjør testen.
- Tillegg av både kubisk og kvartisk ledd (tredje ordens og fjerde ordens) i Exp fører til en reduksjon i summen av kvadrerte avvik til 281.54. Vis hvordan hypotesen om at koeffisientene til disse tilleggsvariable samtidig er lik null kan testes og gjør testen.
- Forklar hva som måles med 'R i andre' og forskjellen mellom 'R i andre' og korrigert 'R i andre'.
- En regresjon med sum av kvadrerte avvik fra modellen over med de samme forklaringsvariablene gir en 'R i andre' på 0.0014. Diskuter hvordan du kan bruke denne informasjonen til å teste for heteroskedastisitet.

Oppgave 2

Britiske forskere prøver å estimere effekten av migrasjon på arbeidsmarkedet i Storbritannia. De har bare gode regionale data for 1999 og estimerer den følgende modellen med minste kvadraters metode:

$$\ln Wage_i = \beta_0 + \beta_1 MigPop_i + \beta_2 X_i + u_{it}$$

Her er $\ln Wage_i$ den gjennomsnittlige log av reallønnen til arbeidere født i Storbritannia som lever i region i , $MigPop_i$ er prosentandelen av innflyttere i befolkningen i region i , X_i representerer et sett av kontrollvariable i region i . De finner et estimat for koeffisienten for MigPop på 0.88 med standardavvik på 0.21.

- Gi en tolkning av koeffisienten. Tror du dette estimat med bruk av minste kvadraters metode gir et forventningsrett estimat for virkningen av migrasjon på lønninger?
- Forskerne argumenterer at nye innflyttere delvis baserer beslutningen om hvor de skal bo i Storbritannia på eksisterende beholdning av tilsvarende innflyttere. De bruker beholdningen av innflyttere for 10 år siden som instrument for MigPop. Hvilke forutsetninger må dette instrumentet tilfredsstillere? Kan du tenke deg noen grunner for at det foreslåtte instrumentet ikke tilfredsstiller disse forutsetningene.
- Forklar hvordan forskerne kan gå fram for å finne et instrumentvariabel (IV) estimat for effekten av MigPop.
- Med bruk av standard IV estimeringsmetode finner de en koeffisient for MigPop på 0.94 med standardavvik på 0.58. Kommenter deres funn.
- Diskuter hvordan forskerne kunne gå fram for å teste bruken av instrumentet.

Oppgave 3

Finansdepartementet I Storbritannia har vært bekymret for at økningen i merverdiavgift (MVA) fra 17,5% til 20% den 1. Januar 2011 vil redusere samlede salg. Økonomer klarte å overbevise departementet om at økningen burde innføres i noen regioner og ikke andre. Ti regioner er unntatt for økningen i en periode på 3 måneder. Du samler inn data for de 10 regionene som er unntatt og ytterligere 10 tilsvarende regioner som har hatt økning i merverdiavgift. Den følgende tabellen viser gjennomsnittlig detaljsalg per butikk (i millioner pund) for 3 måneder før og etter økningen i MVA:

	Økning før MVA-økning	Økning etter MVA-økning
Regioner med økning	4.77	4.15
Regioner med unntak	4.34	4.21

- Kalkuler 'difference in difference' estimat av effekten av MVA økning på detaljsalget.
- Diskuter nødvendige forutsetninger for å tolke 'difference in difference' estimatet som en årsakseffekt av MVA på økningen i salg.

Du samler nå inn data fra alle detaljbutikker i de 10 behandlingsregionene (med MVA økning) og de 10 kontrollregionene (med unntak) om salget de tre siste månedene i 2009 og de 3 første månedene i 2010 (alle før økningen i MVA).

Du estimerer følgende regresjon basert på individ-data for de to tidsperiodene:

$$Sales_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Time_t + \alpha_{DiD}(Treat_i \times Time_t) + u_{it}$$

Her er Sales detaljsalg fra butikk i tidsperiode t, Treat er en dummy variabel lik 1 hvis butikken er i en region med MVA økning og null hvis den er i en region med unntak, Time er en dummy variabel som er lik null i de siste 3 månedene i 2009 og 1 for de første 3 månedene i 2010.

- c) Du estimerer en α_{DiD} lik -0.1 med standardavvik 0.02. Hva betyr dette for din 'difference in difference' estimering. Tror du estimatene i a) over- eller under-estimerer effekten av MVA på salg?

Ved bruk av data i dine behandlingsregioner (med MVA økning), estimerer du den følgende regresjon for endring i salget i butikk i over perioden med MVA økning:

$$\Delta Sales_i = \beta_0 + \beta_1 Distance_i + u_i$$

Her måler Distance avstanden fra butikk i til den nærmeste kontrollregion (region som har unntak for MVA økning). Du finner en estimert koeffisient for Distance på 0.011 med standardavvik på 0.0012.

- d) Hva betyr dette for din 'difference in difference' estimering. Tror du estimatene i a) over- eller under-estimerer effekten av MVA på salg?

Nynorsk

Alle spørsmål skal besvarast.
To tabellar er vedlagt.

Oppgåve 1

Regresjonsmodellen nedanfor er estimert med data for individ fra folketellinga i USA i år 2000 (805 observasjonar)

$$lWage_i = \beta_0 + \beta_1 Exp_i + \beta_2 Exp_i^2 + \beta_3 Edyrs_i + \beta_4 Part_i + \beta_5 Female_i + \beta_6 Black_i + u_i$$

kor lWage er log timelønn, Exp er arbeidsmarknadserfaring i år, Exp^2 er erfaring opphøgd i andre, Edyrs er antall år utdanning, Part er en dummy variable lik 1 hvis individet arbeider deltid, Female er en dummy variabel lik 1 hvis individet er kvinne, Black er en dummy variabel som er lik 1 hvis individet er sort; alle variable er målt for individ j.

Regresjonsestimater og noe statistikk ('R i andre', Korrigert 'R i andre', og summen av kvadrerte avvik SSR) er vist i tabellen under:

lWage	Koeffisient	St. avvik
Exp	0.037	0.0063
Exp^2	-0.00049	0.00015
Edyrs	0.091	0.0095
Part	0.0030	0.0581
Female	-0.225	0.0427
Black	-0.170	0.0590
Const	0.923	
'R i andre'	0.2308	
Korr. 'R i andre'	0.2251	
SSR	282.584	

- Tolk koeffisientane, kommenter deira fortegn, størrelse og signifikans. Kva er dei viktige faktorane som forklarar lønnsforskjellar gitt denne analysen?
- Etter kor mange år i arbeidsmarknaden begynner individets lønn å falle?
- Vis korleis hypotesa om at koeffisienten for Female er lik -0.1 kan testes og gjør testen.
- Tillegg av både kubisk og kvartisk ledd (tredje ordens og fjerde ordens) i Exp fører til en reduksjon i summen av kvadrerte avvik til 281.54. Vis korleis hypotesa om at koeffisientane til disse tilleggsvariable samtidig er lik null kan testes og gjør testen.
- Forklar kva som måles med 'R i andre' og forskjellen mellom 'R i andre' og korrigert 'R i andre'.
- Ein regresjon med sum av kvadrerte avvik fra modellen over med dei same forklaringsvariablane gir en 'R i andre' på 0.0014. Diskuter korleis du kan bruke denne informasjonen til å teste for heteroskedastisitet.

Oppgave 2

Britiske forskarar prøver å estimere effekten av migrasjon på arbeidsmarknaden i Storbritannia. Dei har bare gode regionale data for 1999 og estimerer den fylgjande modellen med minste kvadraters metode:

$$\ln Wage_i = \beta_0 + \beta_1 MigPop_i + \beta_2 X_i + u_{it}$$

Her er $\ln Wage_i$ den gjennomsnittlige log av reallønnen til arbeidere født i Storbritannia som lever i region i, $MigPop_i$ er prosentandelen av innflyttere i befolkningen i region i, X_i representerer et sett av kontrollvariable i region i. De finner et estimat for koeffisienten for MigPop på 0.88 med standardavvik på 0.21.

- Gje ei tolking av koeffisienten. Tror du dette estimat med bruk av minste kvadraters metode gir et forventningsrett estimat for verknaden av migrasjon på lønninger?
- Forskerane argumenterer at nye innflyttarar delvis baserer beslutningen om kor dei skal bo i Storbritannia på eksisterande beholdning av tilsvarende innflyttarar. Dei bruker beholdningen av innflyttarar for 10 år sidan som instrument for MigPop. Hvilke forutsetningar må dette instrumentet tilfredstille? Kan du tenke deg noen grunner for at det foreslåtte instrumentet ikke tilfredstillar desse forutsetningane.
- Forklar korleis forskerane kan gå fram for å finne eit instrumentvariabel (IV) estimat for effekten av MigPop.
- Med bruk av standard IV estimeringsmetode finner dei ein koeffisient for MigPop på 0.94 med standardavvik på 0.58. Kommenter funnet deira.
- Diskuter korleis forskerane kunne gå fram for å teste bruken av instrumentet.

Oppgave 3

Finansdepartementet i Storbritannia har vært bekymra for at auken i merverdiavgift (MVA) fra 17,5% til 20% den 1. Januar 2011 vil redusere samlede salg. Økonomer klarte å overbevise departementet om at auken burde innføres i noen regioner og ikke andre. Ti regioner er unntatt for auken i en periode på 3 måneder. Du samler inn data for de 10 regionene som er unntatt og ytterligere 10 tilsvarende regioner som har hatt auke i merverdiavgift. Den fylgjande tabellen viser gjennomsnittlig detaljsalg per butikk (i millioner pund) for 3 måneder før og etter auken i MVA:

	Auking før MVA-auke	Auking etter MVA-auke
Regioner med auking	4.77	4.15
Regioner med unntak	4.34	4.21

- Kalkuler 'difference in difference' estimat av effekten av MVA auking på detaljsalget.
- Diskuter nødvendige forutsetningar for å tolke 'difference in difference' estimatet som en årsakseffekt av MVA på auken i salg.

Du samler nå inn data fra alle detaljbutikker i de 10 behandlingsregionane (med MVA auke) og de 10 kontrollregionene (med unntak) om salget de tre siste månedene i 2009 og de 3 første månedene i 2010 (alle før auken i MVA).

Du estimerer følgende regresjon basert på individ-data for de to tidsperiodene:

$$Sales_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Time_t + \alpha_{DID}(Treat_i \times Time_t) + u_{it}$$

Her er Sales detaljsalg fra butikk i tidsperiode t, Treat er en dummy variabel lik 1 hvis butikken er i en region med MVA auke og null hvis den er i en region med unntak, Time er en dummy variabel som er lik null i dei siste 3 månedene i 2009 og 1 for dei første 3 månedene i 2010.

- c) Du estimerer en α_{DID} lik -0.1 med standardavvik 0.02. Kva betyr dette for din 'difference in difference' estimering. Tror du estimatane i a) over- eller under-estimerer effekten av MVA på salg?

Ved bruk av data i dine behandlingsregioner (med MVA auke), estimerer du den følgende regresjon for endring i salget i butikk i over perioden med MVA auke:

$$\Delta Sales_i = \beta_0 + \beta_1 Distance_i + u_i$$

Her måler Distance avstanden fra butikk i til den nærmeste kontrollregion (region som har unntak for MVA auke). Du finner en estimert koeffisient for Distance på 0.011 med standardavvik på 0.0012.

- d) Kva betyr dette for din 'difference in difference' estimering. Tror du estimatane i a) over- eller under-estimerer effekten av MVA på salg?

English

All questions to be answered.
Two tables enclosed.

Question 1

The following regression is estimated on individuals from the US Census in the year 2000 (805 observations)

$$lWage_i = \beta_0 + \beta_1 Exp_i + \beta_2 Exp_i^2 + \beta_3 Edyrs_i + \beta_4 Part_i + \beta_5 Female_i + \beta_6 Black_i + u_i$$

where $lWage$ is the log of hourly wage, Exp is the labour market experience in years, Exp^2 is experience squared, $Edyrs$ is the number of years in education, $Part$ is a dummy variable equal to 1 if the individual works part time, $Female$ is a dummy variable equal to 1 if the individual is female, $Black$ is a dummy variable which is equal to 1 if individual is black; all variables measured for individual i .

The regression estimates and some statistics (R-squared, Adjusted R-squared and sum of squared residuals SSR) are shown in the table below:

$lWage$	Coefficient	St.error
Exp	0.037	0.0063
Exp^2	-0.00049	0.00015
$Edyrs$	0.091	0.0095
$Part$	0.0030	0.0581
$Female$	-0.225	0.0427
$Black$	-0.170	0.0590
$Const$	0.923	
R-squared	0.2308	
Adj. R-squared	0.2251	
SSR	282.584	

- Interpret the coefficients, comment their sign, magnitude and significance. What are the major factors explaining wage differences based on this evidence?
- After how many years in the labour market does an individual's wage begin to decline?
- Show how to test the hypothesis that the coefficient on $Female$ is equal to -0.1 and do the test.
- Adding both a cubic and a quartic term (third order and fourth order) in Exp leads to a reduction in the sum of squared residuals to 281.54. Show how to test the hypothesis that the coefficients of these additional variables are jointly equal to zero and do the test.
- Explain what is measured by R-squared and the difference between R-squared and Adjusted R-squared.
- A regression of the squared errors from the regression above on the same variables yields an R-squared value of 0.0014. Discuss how you can use this information to test for heteroskedasticity.

Question 2

British researchers attempt at estimating the impact of migration on the UK labour market. They only have good regional data for 1999 and estimate the following model using OLS:

$$lWage_i = \beta_0 + \beta_1 MigPop_i + \beta_2 X_i + u_{it}$$

Here $lWage_i$ is the mean log real wage of british born workers who live in region i , $MigPop_i$ is the percent of migrants in the population in region i , X_i represents a set of control variable in region i . They find an estimate of the coefficient on MigPop of 0.88 with a standard error of 0.21.

- Provide an interpretation of this coefficient. Do you think that this OLS estimate gives an unbiased estimate of the impact of migration on wages?
- The researchers argue that new migrants partly base their decision on where to reside in the UK on the current stock of similar migrants. They use the stock of migrants ten years ago in the region as an instrument for MigPop. What conditions do this instrument need to satisfy? Can you think of any reasons why the proposed instrument does not satisfy these conditions?
- Explain how the researchers may proceed to obtain an instrument variable (IV) estimate of the effect of MigPop.
- Using standard IV estimation methods they estimate the coefficient of MigPop to 0.94 with a standard error of 0.58. Comment on their findings.
- Discuss how the researchers could proceed to test the use of the instrument.

Question 3

The Ministry of Finance in the UK has been worried that the increase in the value added tax (VAT) from 17,5% to 20% January 1 2011 will reduce aggregate sales. Economists managed to convince the Ministry to introduce the rise in some regions and not others. Ten counties of the UK are exempt from the increase for a 3 month period. You collect data on sales for the exempt 10 counties and for another 10 similar counties who are subject to the VAT rise. The following table shows the average retail sales per outlet (in million pounds) for the 3 months before and after the VAT increase:

	Pre-VAT increase	Post-VAT increase
VAT rise counties	4.77	4.15
Exempt counties	4.34	4.21

- Calculate the difference in difference estimate of the impact of the VAT increase on retail sales.
- Discuss the assumptions needed to interpret this difference in difference estimate as the causal effect of the VAT rise on sales.

You now collect data from all individual retail outlets in the 10 treatment (VAT rise) and 10 control (exempt) counties on sales for the last 3 months of 2009 and the first 3 months of 2010 (all before the VAT increase).

You estimate the following regression based on the individual data for the two time periods:

$$Sales_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Time_t + \alpha_{DiD}(Treat_i \times Time_t) + u_{it}$$

Here Sales are retail sales from outlet i in time period t , Treat is a dummy variable equal to 1 if the outlet is in a VAT rise county and zero if it is in a VAT exempt country, Time is a dummy variable that is equal to zero for the last 3 months of 2009 and 1 for the first 3 months of 2010.

- c) You estimate an α_{DiD} equal to -0.1 with a standard error of 0.02. What does this mean for your difference in difference estimation. Do you think the estimates in a) over- or under-estimates the effect of VAT on sales?

Using data in your treated (VAT rise) counties, you estimate the following regression on the change of retail sales in outlet i over the period of the VAT rise:

$$\Delta Sales_i = \beta_0 + \beta_1 Distance_i + u_i$$

Here Distance measures the distance from the outlet i from the nearest control county (county that is exempt from the VAT increase). You find an estimated coefficient of Distance of 0.011 with a standard error of 0.0012.

- d) What does this mean for your difference in difference estimation? Do you think that the estimates in part a) over- or under-estimate the impact of VAT on sales?