

**EKSAMENSOPPGAVE I SØK3515**

**MIKRO- OG PANELDATAØKONOMETRI**

**MICRO AND PANEL DATA ECONOMETRICS**

Faglig kontakt under eksamen: Bjarne Strøm, Tlf.: 91933

Eksamensdato: Fredag 7. desember 2012

Eksamenssted: Dragvoll

Eksamenstid: 6 timer

Studiepoeng: 15

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 el. HP 30S.

Sensur: 7. januar 2013

Eksamensoppgaven består av 3 oppgaver med delspørsmål som alle skal besvares.

Antall sider bokmål: 2

Antall sider nynorsk: 2

Antall sider engelsk: 2

Vedlegg: Tabell

*Oppgave 1.*

Vi vil estimere sammenhengen mellom folks alkoholforbruk og inntektsnivå basert på tverrsnittsdata. Foreslå økonometrisk modell, forklar hvordan du vil gå fram for å estimere effekten av inntekt på drikking, og hvordan du vil tolke og bruke estimerte effekter i følgende tilfeller:

- a) Datamaterialet er individbasert og inneholder en variabel som tar verdien 0 hvis individet ikke drikker alkohol i det hele tatt, 1 hvis individet drikker alkohol av og til men mindre enn en gang i uka, 2 hvis individet drikker alkohol en gang hver uke eller oftere. I tillegg har vi informasjon om individenes årsinntekt.
- b) Datamaterialet inneholder informasjon om 1000 husholdningers utgifter til kjøp av alkohol i løpet av en uke. 30% av husholdningene rapporterer at de ikke har utgifter til kjøp av alkohol i løpet uka. I tillegg har vi informasjon om husholdningenes inntekt pr.uke.

*Oppgave 2.*

To forskere har studert sammenhengen mellom den regionale andelen elever som fullfører videregående skole og arbeidsledighetsraten i regionen. Avhengig variabel i alle estimeringene er prosentandelen av elevene som avsluttet grunnskolen i år t i regionen som har fullført videregående skole 5 år etterpå. Forklaringsvariabelen er arbeidsledighetsraten i regionen målt i prosent. Tabell 1 viser deskriptiv statistikk for fullføring og arbeidsledighet, mens tabell 2 viser estimeringsresultater ved to ulike metoder. Estimeringsperiode er 1981-2004. Antall regioner er 90.

Tabell 1. Deskriptiv statistikk

Variables	Antall observasjoner	Min	Max	Gjennomsnitt	Standardavvik
Regional fullføring (%)	2160	27,84	86,70	64,35	10,05
Regional arbeidsledighet (%)	2160	0,16	6,89	2,64	1,13

## SØK3515 – Mikro- og paneldataøkonometri

Tabell 2. Estimeringsresultater. Estimerte standardavvik i parentes.

	“Fixed regional effects”	“Random regional effects”
Regional arbeidsledighet (%)	0,964 (0,315)	-0,013 (0,010)
Årsdummier	Ja	Ja
Observasjoner	2160	2160

- Forklar hvordan «random effects»-modellen er estimert.
- Gjennomfør en Hausman-test og benytt eventuell annen informasjon til å avgjøre hvilken estimeringsmetode som gir det mest troverdige anslaget på effekten av ledighet på fullføringsraten.
- Beregn et 95 prosent konfidensintervall rundt effekten av ledigheten.
- Bruk opplysningene i tabellen til å beregne
  - Virkingen av et standardavviks økning i ledigheten
  - Elastisiteten av fullføringsandelen med hensyn på arbeidsledighetsraten.
- En kollega av forskerne argumenterer for at høyere lønn for ufaglært arbeidskraft kan føre til redusert fullføringsandel og at dette ikke er ivaretatt i den økonometriske modellen som dermed er feilspesifisert. Drøft dette argumentet og foreslå eventuelle alternative estimeringsopplegg for å imøtegå kritikken, gitt at lønnsdata ikke er tilgjengelig.

### Oppgave 3.

Gitt modellen  $y_{it} = a + x_{it}b + cy_{it-1} + d_i + u_{it}$  som skal estimeres på et panelmateriale med  $t=0,1,\dots,4$  og  $i=1,\dots,5000$ .  $x$  er en rekke-vektor av dimensjon  $1 \times k$  av forklaringsvariable med tilhørende koeffisientvektor  $b$ ,  $c$  er en koeffisient mellom 0 og 1 og  $y$  er den avhengige variabelen.  $d_i + u_{it}$  er stokastisk feilledd der  $d_i$  er den enhetsspesifikke komponenten.

To økonometrikere foretar hver sin transformasjon av modellen for å fjerne den enhetsspesifikke effekten  $d_i$ . A transformerer modellen ved å subtrahere enhetsspesifikke gjennomsnitt av hver variabel, mens B transformerer modellen ved å subtrahere lagget verdi av hver variabel. Hvilke egenskaper vil OLS-estimatorene for modellene til henholdsvis A og B ha? Hvilken av transformasjonsmetodene foretrekker du, og hvilken estimeringsmetode ville du bruke for å oppnå troverdige estimater på de ukjente parametrene i modellen? Begrunn svaret.

### Oppgave 1.

Vi vil estimera samanhengen mellom folk sitt forbruk av alkohol og inntektsnivå basert på tverrsnittsdata. Foreslå økonometrisk modell, forklår korleis du vil gå fram for å estimere effekten av inntekt på drikking, og korleis du vil tolke og bruka estimerte effektar i fylgjande tilfelle:

a) Datamaterialet er individbasert og inneheld ein variabel som tar verdien 0 dersom individet ikkje drikk alkohol i det heile, 1 dersom individet drikk alkohol av og til men mindre enn ein gong i veka, 2 dersom individet drikk alkohol ein gong i veka eller oftare. I tillegg har vi informasjon om individa si årlege inntekt.

b) Datamaterialet inneheld informasjon om 1000 hushald sine utgifter til kjøp av alkohol i løpet av ei veke. 30% av hushalda rapporterer at dei ikkje har utgifter til kjøp av alkohol i løpet av veka. I tillegg har vi informasjon om hushalda si inntekt pr. veke.

### Oppgave 2.

To forskarar har studert samanhengen mellom den regionale andelen elevar som fullfører vidaregåande skule og arbeidsledighetsraten i regionen. Avhengig variabel i alle estimeringane er prosentandelen av elevane som avslutta grunnskulen i år  $t$  i regionen som fullførte vidaregåande skule 5 år etter. Forklåringsvariabelen er arbeidsledighetsraten i regionen målt i prosent. Tabell 1 syner deskriptiv statistikk for fullføring og arbeidsledighet, medan tabell 2 syner estimeringsresultat ved to ulike metodar. Estimeringsperiode er 1981-2004. Talet på regionar er 90.

Table 1. Deskriptiv statistikk

Variable	Antall observasjonar	Min	Max	Gjennomsnitt	Standardavvik
Regional fullføring (%)	2160	27,84	86,70	64,35	10,05
Regional arbeidsledighet (%)	2160	0,16	6,89	2,64	1,13

## SØK 3515 Mikro-og paneldata økonometri

Table 2. Estimeringsresultat. Estimerte standardavvik i parentes.

	“Fixed regional effects”	“Random regional effects”
Regional arbeidsledighet (%)	0,964 (0,315)	-0,013 (0,010)
Årsdummiar	Ja	Ja
Observasjonar	2160	2160

- Forklår korleis «random effects»-modellen er estimert.
- Gjennomfør ein Hausman-test og nytt eventuell annan informasjon til å avgjere kva for estimeringsmetode som gjev det mest truverdige anslaget på effekten av ledighet på fullføringsraten.
- Berekn eit 95 prosent konfidensintervall rundt effekten av ledigheten.
- Bruk opplysningane i tabellen til å rekne ut
  - Verknaden av eit standardavviks auke i ledigheten
  - Elastisiteten av fullføringsandelen med omsyn til arbeidsledighetsrata.
- Ein kollega av forskarane argumenterer for at høgare løn for ufaglært arbeidskraft kan føra til redusert fullføringsandel og at dette ikkje er teke vare på i den økonometriske modellen som difor er feilspesifisert. Drøft dette argumentet og eventuelt gjer framlegg om alternative estimeringsopplegg for å imøtegå kritikken, gjeve at lønnsdata ikkje er tilgjengeleg.

### Oppgåve 3.

Gjeve modellen  $y_{it} = a + x_{it}b + cy_{it-1} + d_i + u_{it}$  som skal estimerast på eit panelmateriale med  $t=0,1,\dots,4$  og  $i=1,\dots,5000$ .  $x$  er ein rekke-vektor av dimensjon  $1 \times k$  av forklaringsvariablar med tilhøyrande koeffisientvektor  $b$ ,  $c$  er ein koeffisient mellom 0 og 1 og  $y$  er den avhengige variabelen.  $d_i + u_{it}$  er stokastisk feilledd der  $d_i$  er den einskapsspesifikke komponenten.

To økonometrikarar foretar kvar sin transformasjon av modellen for å fjerna den einskapsspesifikke effekten  $d_i$ . A transformerer modellen ved å subtrahera einskapsspesifikke gjennomsnitt av kvar variabel, mens B transformerer modellen ved å subtrahera lagga verdi av kvar variabel. Kva for eigenskaper vil OLS-estimatorane for modellane til henholdsvis A og B ha? Kva for ein av transformasjonsmetodane ville du foretrekke, og kva for estimeringsmetode ville du valgt for å oppnå truverdige estimat på dei ukjente parametrane i modellen?

Grunngje svaret.

*Question 1.*

We will estimate the relationship between individual's alcohol consumption and income level based on cross section data. Suggest econometric model, explain how you will estimate the impact of income on drinking and explain how you will interpret and use estimated coefficients in the following cases:

a) Data is based on individuals and contains a variable taking the value 0 if the individual do not drink alcohol at all, 1 if the individual drinks alcohol now and then, but not every week and 2 if the individual drinks alcohol once a week or more. In addition, we have data on yearly income for all individuals

b) Data contains information about alcohol expenditure during one week for 1000 households. 30% of the households report zero expenditure on alcohol during the week. In addition we have data on weekly income for all households.

*Question 2.*

Two researchers have studied the relationship between the regional share of students completing upper secondary school and the unemployment rate in the region. The dependent variable is the percentage share of students finishing lower secondary school in year  $t$  in the region that have completed upper secondary school 5 years afterwards. The explanatory variable is the unemployment rate in the region measured in percent. Estimation period is 1981-2004. The number of regions is 90. Table 1 presents descriptive statistics for completion rate and unemployment rate, while Table 2 shows the estimation results using two different methods. Estimation period is 1981-2004. The number of regions is 90.

Table 1. Descriptive statistics

Variables	Obs	Min	Max	Mean	Standard deviation
Regional completion (%)	2160	27,84	86,70	64,35	10,05
Regional unemployment (%)	2160	0,16	6,89	2,64	1,13

## SØK3515 – Micro and panel data econometrics

Table 2. Estimation results. Estimated standard errors in parenthesis.

	Fixed regional effects	Random regional effects
Regional unemployment (%)	0,964 (0,315)	-0,013 (0,010)
Year dummies	Yes	Yes
Observations	2160	2160

- a) Explain how the “random effects” model is estimated.
- b) Conduct a Hausman-test and eventually use other information to decide which of the two estimation methods is most likely to give credible estimate on the effect of unemployment on completion rate.
- c) Compute a 95 percent confidence interval for the unemployment effect.
- d) Use information in the tables to compute:
  - i) The effect of one standard deviation increase in the unemployment rate
  - ii) The elasticity of completion rate with respect to unemployment.
- e) A college of the researchers argues that higher wages for unskilled workers can reduce the completion rate and since this is not captured in the econometric model, the model is miss-specified. Discuss this argument, and eventually suggest alternative estimation methods to account for this critique, given that wage data is not available.

### Question 3.

Given the model:  $y_{it} = a + x_{it}b + cy_{it-1} + d_i + u_{it}$  which is to be estimated on a panel data set with  $t=0, \dots, 4$  and  $i=1, \dots, 5000$ .  $x$  is a row vector of dimension  $1 \times k$  of explanatory variables with corresponding coefficient vector  $b$ ,  $c$  is a coefficient between 0 and 1 and  $y$  is the dependent variable.  $d_i + u_{it}$  is the error term where  $d_i$  is the unit specific component. Two econometricians conduct a transformation of the model to remove the unit specific effect  $d_i$ . A transforms the model by subtracting unit specific means from each variable, while B transforms the model by subtracting lagged values of each variable. What are the properties of the OLS estimators for the models of A and B, respectively? What transformation method would you prefer, and what estimation method would you use to obtain credible estimates of the unknown parameters in the model. Explain your answer.