

Oppgave 1.

a) og b) Må få fram at koeffisientene i kolonne (1) og (2) er koeffisienter i en underliggende ligning for latent variabel (se Verbeek kap 7.2.1), og denne bør formuleres eksplisitt som f.eks

$helse_i^* = x_i\beta + \varepsilon_i$ der $helse_i^*$ er en latent uobserverbar variabel for helsetilstand og ε_i er et stokastisk restledd som er identisk og uavhengig normalfordelt med forventning 0 og varians 1.

Observerbare variable 1,2 og 3 relateres til underliggende latent variabel ved hjelp av terskelverdiene. Likelihoodfunksjonen bør forklares og formuleres. F.eks, la $d_{ij} = 1$ hvis individ i velger $y=j$, $j=1,2, 3$, $i=1,2,\dots,N$ og 0 ellers. Gitt at terskelverdiene er representert ved α_1 og α_2 og x representerer forklaringsvariablene med tilhørende koeffisientvektor β , blir Likelihoodfunksjonen dermed:

$$L = \prod_{i=1}^N \{ \Phi(\alpha_1 - x_i\beta)^{d_{i1}} [\Phi(\alpha_2 - x_i\beta) - \Phi(\alpha_1 - x_i\beta)]^{d_{i2}} [1 - \Phi(\alpha_2 - x_i\beta)]^{d_{i3}} \}$$

Tar logaritmen til denne og får log-likelihoodfunksjonen som:

$$\ln L = \sum_{i=1}^N \{ d_{i1} \ln \Phi(\alpha_1 - x_i\beta) + d_{i2} \ln [\Phi(\alpha_2 - x_i\beta) - \Phi(\alpha_1 - x_i\beta)] + d_{i3} \ln [1 - \Phi(\alpha_2 - x_i\beta)] \}$$

Kandidatene bør forklare at estimerte koeffisienter for de ukjente terskelverdiene og koeffisientvektoren henholdsvis $\alpha_1, \alpha_2, \beta$ finnes som løsningen for det tilhørende settet av førsteordensbetingelser for maksimum av log-likelihoodfunksjonen og er rapportert i tabell 1. Kandidatene bør kommentere at identifikasjon av både konstantledd og to terskelverdier ikke er mulig og at en normaliseringsrestriksjon er nødvendig – restriksjon om konstantledd lik null er derfor nødvendig når begge terskelverdiene inngår, se Verbeek 7.2.1.

c) Rett fram formulering av sannsynligheter for de tre utfallene, innsetting av estimerte koeffisientverdier og evaluering i de angitte verdier av forklaringsvariablene ved hjelp av formelen og bruk av vedlagt tabell for normalfordelingen gir

	$\overline{P(helse = 1)}$	$\overline{P(helse = 2)}$	$\overline{P(helse = 3)}$
Utdanning=16	0.41	0.41	0.18
Utdanning=12	0.49	0.38	0.13

Resultatene tyder på at utdanning bidrar til bedre helse. Det eksemplifiseres ved at sannsynligheten for å ha svært god helse (helse=3) er 5% poeng høyere for en med 16 år utdanning enn en med 12 år utdanning, mens sannsynligheten for dårlig helse (helse=1) er 8% poeng høyere for en med 12 år utdanning enn en med 16 år utdanning.

d) Hypotesen som testes er nullrestriksjoner $\beta_{sivilstand} = \beta_{barn} = 0$ med Likelihood-ratio-test basert på log Likelihood-verdiene oppgitt i tabellen. Bør kort kommentere intuisjonen for testen og beregne observatoren $LR = 2(L_{(1)} - L_{(2)})$ der $L_{(1)}, L_{(2)}$ er Log-likelihoodverdien for modellen i henholdsvis kolonne (1) og (2) i tabell 1. $LR=3.53$ som er under kritisk verdi i Kvikvadratfordelinga med antall frihetsgrader (2) pålagt modellen i kolonne (2) under nullhypotesen. Hypotesen om at sivilstand og familiestørrelse ikke har betydning forkastes derfor ikke.

Oppgave 2.

a) Instruktivt å sette opp enkel modell

$lprod_{it} = \beta_0 + \beta_1 lkap_{it} + \beta_2 larb_{it} + \eta_i + \varepsilon_{it}$ der η_i og ε_{it} er hhv bedriftsspesifikk og idiosynkratisk komponent.

Rett fram om estimering av OLS FE og RE. Koeffisientene foran lkap og larb i kolonne (1), (3) og (5) er grense-elasticitetene av hhv realkapital (β_1) og arbeidskraft (β_2) i en underliggende Cobb Douglas produktfunksjon. Må innse at modellene estimert i kolonne (2), (4) og (6) er rene transformasjoner av utgangsmodellen. Koeffisienten foran larb i disse blir $(\beta_1 + \beta_2 - 1)$ mens koeffisienten foran lkap-larb er β_1 .

b) Det følger rett fram av svaret i a) at test av konstant skala-avkastning teste hypotesen om koeffisienten foran larb i (2), (4) og (6) er lik null. Forkaster nullhypotesen i alle disse ved enkel t-test på koeffisienten foran larb i (2), (4) og (6).

c) Bør få fram at relevansen av innvendingen er avhengig av hvilke forutsetninger vi er villig til å gjøre om restleddskomponentene som generelt vil inneholde det uobserverte arbeidsmiljøet. Må få fram at OLS og RE er konsistent bare dersom arbeidsmiljøet som generelt kan inngå i begge komponentene er ukorrelert med arbeidskraft og kapitalmengde, i dette tilfellet er RE effisient estimator.

Kan så betrakte to tilfeller:

i) Arbeidsmiljøet er konstant over tid innen hver bedrift,

Da er vil FE være konsistent estimator for grense-elasticitetene, mens både OLS og RE er inkonsistent hvis kollegaens innvending stemmer.

ii) Arbeidsmiljøet varierer både mellom bedriftene og også over tid innen hver bedrift

Da er heller ikke FE konsistent estimator. Relevant å kommentere at vi da i prinsippet kunne brukt en instrumentvariabelstrategi med instrumentvariable for lkap og larb for å oppnå konsistente estimatorene. Dvs, vi måtte ha variable som både har bedrifts og tidsvariasjon og som påvirker kapitalmengde og arbeidskraft uten å påvirke produksjonen direkte. Men slike variable er selvsagt vanskelig å forestille seg i praksis.

d) Rett fram beskrivelse av Hausman-testen, og hva den tester, se Verbeek kap 10.2.4.

Testresultatet gir klar indikasjon på at FE framstår som mer troverdig enn RE.

e) Kandidatene bør innse at alle modellene inneholder årsummier som ivaretar effekten av nasjonale forurensingsforskrifter som påvirker alle bedriftene i bransjen.

f) Kandidatene bør innse at temperaturvariabelen slik den er formulert bare vil ha én verdi for hver bedrift og således vil ivaretas av faste bedriftseffekter (FE-variantene). Det betyr at **nivå-effekten** av temperaturvariabelen ikke vil la seg identifisere i FE-variantene. Noen vil kanskje også foreslå interaksjon mellom temperatur og øvrige variable (kapital, arbeidskraft eller tidsdummier) i modellen. Relevant å påpeke at effekten av slike interaksjonsvariable vil kunne identifiseres i FE-variant.

g) Bør formulere ulike dynamiske spesifikasjoner og diskutere metodeutfordringer:

i) inkludere lag i forklaringsvariablene, f.eks formulere modellen som

$$lprod_{it} = \beta_0 + \beta_{10}lkap_{it} + \beta_{11}lkap_{it-1} + \beta_{20}larb_{it} + \beta_{21}larb_{it-1} + \eta_i + \varepsilon_{it}$$

Dette gir ingen spesielle estimeringsproblemer, men kan påpeke at $\beta_{10} + \beta_{11}$ og $\beta_{20} + \beta_{21}$ da kan tolkes som langtidselastisiteter.

ii) Mer generell dynamisk formulering: Inkludere lag i lprod som forklaringsvariabel.

$$lprod_{it} = \beta_0 + \alpha_1 lprod_{it-1} + \beta_1 lkap_{it} + \beta_2 larb_{it} + \eta_i + \varepsilon_{it}$$

Vise at FE (og OLS) da vil være skjev siden lagga endogen variabel vil være korrelert med det transformerte restleddet, se Verbeek kap. 10.4. Bør diskutere alternative estimatorer som Anderson&Hsiao's IV metode og Arellano Bond sin GMM-versjon.

Oppgave 3.

a) Sentralt problem her er at sysselsatte i offentlig sektor er en selektert gruppe. Håndtering av seleksjon i treatment-modeller er behandlet i Verbeek kap. 7.7. Bør vise hvordan seleksjonsproblemet kan håndteres med en to-stepsprosedyre ala det som er beskrevet i kap.7.7 i Verbeek og presentert på forelesning.

For å få fram problemet med seleksjon på uobserverte variable spesifiseres en seleksjonsligning for sektor tillegg til lønnsligninga. P^* er en latent variabel og seleksjonsligninga er

$$P_i^* = z_i b + \varepsilon_i$$

z_i inneholder x pluss eventuelle ekstra variable som kan bidra til identifikasjon. b er korresponderende koeffisientvektor. Dersom ε_i og u_i følger en bivariat normalfordeling med forventning 0 og kovarians $\sigma_{u\varepsilon}$ mellom dem blir

$$E(w_i | x_i, P_i) = x_i \beta + \delta P_i + \sigma_{\varepsilon u} \underbrace{\left[P_i \lambda_{1i} + (1 - P_i) \lambda_{0i} \right]}_{\text{"LAMBDA"}} \quad \text{der } \lambda_{1i} = \frac{\phi(z_i b)}{\Phi(z_i b)} \quad \text{og } \lambda_{0i} = \frac{-\phi(z_i b)}{1 - \Phi(z_i b)}$$

ϕ og Φ er henholdsvis tetthet og kumulativ tetthetsfunksjon (fordelingsfunksjonen) i standard normalfordeling. Dette innebærer at en naiv OLS regresjon på lønnsligninga uten å korrigere for seleksjon på uobserverte variable vil feilestimere lønnsforskjellen når $\sigma_{\varepsilon u} \neq 0$. Et eksempel kan være at uobservert produktivitet som påvirker lønna uavhengig av sektor også påvirker sektorvalg.

I dette tilfellet kan δ estimeres konsistent ved en tottrinnsprosedyre der lønnsligninga estimeres i et andre trinn med OLS der det inkluderes et ledd som kontrollerer for seleksjonen

$$w_i = x_i \beta + \delta r_i + \sigma_{\varepsilon \eta} \underbrace{\left[r_i \hat{\lambda}_{1i} + (1 - r_i) \hat{\lambda}_{0i} \right]}_{\text{"LAMBDA"}} + \text{restledd}$$

$$\hat{\lambda}_{1i} = \frac{\phi(z_i \hat{b})}{\Phi(z_i \hat{b})} \quad \text{og} \quad \hat{\lambda}_{0i} = \frac{-\phi(z_i \hat{b})}{1 - \Phi(z_i \hat{b})} \quad \text{som inngår i 2.trinns-ligninga estimeres i et første trinn}$$

ved en probit med observert sektorvalg representert ved dummy-variabelen P som utfallsvariabel. Gode kandidater vil påpeke at metoden i prinsippet kan gjennomføres også dersom seleksjonsligninga og lønnsligninga inneholder det samme settet av variable.

Metoden blir imidlertid langt mer troverdig dersom det kan pålegges en eksklusjonsrestriksjon om at noen variable bare inngår i seleksjonsligninga.

b) Her er det mulig å demonstrere betydelig innsikt i identifikasjonsproblemet med utgangspunkt i en enkel panelmodell med observasjoner i tidsdimensjonen for hvert individ i tillegg til tverrsnittsdimensjonen. Restleddet i panelmodellen dekomponeres i en individspesifikk og en idiosynkratisk komponent og kandidatene bør få fram hvordan estimering av modell med individfaste effekter kan gjennomføres (transformasjon fra individsnitt eller første-differensiering) og hvordan utelatte variable kan ivaretas med slik framgangsmåte. Kandidatene bør påpeke at identifikasjon av lønnsforskjell i et slikt opplegg krever

1) variasjon i sektortilhørighet for enkeltindivider mellom 2006 og 2010. M.a.o. påpeke at det er de som *skifter sektor i perioden* som bidrar til identifikasjon av lønnsforskjellen.

2) kausal tolking (troverdig anslag på lønnspremien) i modell med individfaste effekter krever at *endring* i sektortilhørighet (offentlig vs privat) for individer er ukorrelert med den idiosynkratiske restleddskomponenten. Dvs. at valget mellom offentlig/privat sektor bare er en funksjon av individspesifikke variable som er konstante over tid.

Ekstra pluss til kandidater som kan resonnerer på at dersom 2) ikke er oppfylt, kan IV/2SLS-metode med instrumentvariabel som bare påvirker endring i sektortilhørighet over tid innen individ, uten å påvirke lønna direkte være relevant. Men slik variabel er selvsagt vanskelig å finne i praksis.