

Oppgave 1.

a) Ordered probit-koeffisienter. Rett fram tolking av koeffisientene foran forklaringsvariablene i tabell 1 som koeffisienter i underliggende ligning for latent variabel (se Verbeek s. 221), og denne bør formuleres eksplisitt. Terskelkoeffisientene forklares og relateres til underliggende latent variabel.

b) Sannsynligheter: Tilsvarende Verbeek s. 222, med eksplisitt angivelse av koeffisienter, forklaringsvariable og terskelverdier. La  $\alpha_1, \alpha_2$  representere terskelverdiene svarende til de estimerte cut/1 og cut/2 i tabellen,  $y_i^*$  den latente variabelen,  $x$  forklaringsvariablene angitt i tabellen og  $\Phi$  fordelingsfunksjonen for standard normalfordelt variabel. Dette leder til følgende sannsynligheter for henholdsvis utfall 1 (Ikke mer utdanning), 2 (to-årig college) og 3 (fireårig college):

$$P(\text{psechoice}_i = 1|x) = P(y_i^* \leq \alpha_1 | x) = P(x_i\beta + e_i \leq \alpha_1 | x) = P(e_i \leq \alpha_1 - x_i\beta | x) = \Phi(\alpha_1 - x_i\beta)$$

$$P(\text{psechoice}_i = 2|x) = P(\alpha_1 < y_i^* \leq \alpha_2 | x) = P(\alpha_1 < x_i\beta + e_i \leq \alpha_2 | x) = P(\alpha_1 - x_i\beta < e_i \leq \alpha_2 - x_i\beta | x) = \Phi(\alpha_2 - x_i\beta) - \Phi(\alpha_1 - x_i\beta)$$

$$P(\text{psechoice}_i = 3|x) = P(y_i^* > \alpha_2 | x) = 1 - P(y_i^* \leq \alpha_2 | x) = 1 - \Phi(\alpha_2 - x_i\beta)$$

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.

Likelihoodfunksjonen blir 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  $\alpha_1, \alpha_2, \beta$  finnes som løsningen for det tilhørende settet av førsteordensbetingelser for maksimum av log-likelihoodfunksjonen.

c) I besvarelsen av dette spørsmålet bør kandidatene ta utgangspunkt i hvordan marginaleffekten av faminc beregnes, f.eks forklare at

marginaleffekten for utfallet psechoice=1 (ikke mer utdanning) blir

$$\frac{\partial P(\text{psechoice} = 1|x)}{\partial \text{faminc}} = -\beta_{\text{faminc}} \phi(\alpha_1 - x_i \beta)$$

Marginaleffekten for utfallet psechoice=2 (2-årig college) blir

$$\frac{\partial P(\text{psechoice} = 2|x)}{\partial \text{faminc}} = [\phi(\alpha_1 - x_i \beta) - \phi(\alpha_2 - x_i \beta)] \beta_{\text{faminc}}$$

Marginaleffekten for utfallet psechoice=3 (4-årig college) blir

$$\frac{\partial P(\text{psechoice} = 3|x)}{\partial \text{faminc}} = \beta_{\text{faminc}} \phi(\alpha_2 - x_i \beta)$$

der  $\phi(\cdot)$  er tetthetsfunksjonen til standard normalfordelt variabel.

Kandidatene bør bruke dette til å få fram at kommentatorens tolking i beste fall er uklar.

Kandidatene bør forklare hvordan estimert marginaleffekt av familieinntekt på sannsynligheten for å ta 2-årig college (f.eks evaluert i utvalgsgjennomsnittene av forklaringsvariablene og estimerte koeffisienter) avhenger av differansen mellom de to positive leddene i klammeparentesen og således kan ha ukjent fortegn. Bør få fram at kommentatorens påstand uten videre er riktig bare når det gjelder sannsynligheten for å gå videre til 4-årig college.

d) 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 **eller**  $\alpha_1 = 0$  duger, se Verbeek s. 222. Bør innse at konstantledd lik null er valgt som restriksjon i estimeringen av modellene i Tabell 1.

e) Test av nullrestriksjoner  $\beta_{\text{parcoll}} = \beta_{\text{female}} = \beta_{\text{Black}} = 0$  på modell B med Likelihood-ratio-test.

Bør kort kommentere intuisjonen for testen og beregne observatoren  $LR = 2(L_A - L_B)$  der  $L_A, L_B$  er Log-likelihoodverdien for modellen i henholdsvis Modell A og B i tabellen.

LR=17.9 som klart overstiger kritisk verdi i Kjikvadratfordelinga med antall frihetsgrader (3) pålagt modell A under nullhypotesen

Oppgave 2.

a) Bør kommentere kommentatorens forslag om å ta hensyn til valget av fagforeningsstatus med utgangspunkt i Heckman seleksjonsmodell (Tobit II kap 7.5.1 i Verbeek). Formulere

seleksjonsligning i tillegg til lønnslikninga (strukturelligning av interesse) og presisere at lønnsbare observeres for fagorganiserte. Få fram hvordan modellen kan estimeres med en to-stegsprosedyre gitt forutsetninger om fordelinga til restleddene i struktur og seleksjonslikninga, eller evt. Maximum likelihood. Kandidatene bør også få fram at seleksjonsproblemet gir opphav til skjevhet i ordinær OLS-estimering av lønnslikninga, og gode studenter vil påpeke at dette vil være tilfelle hvis restleddet i seleksjonslikninga og strukturelligning er korrelert, og hvordan dette eventuelt kan testes i 2.steget med nullhypotese om at koeffisienten foran inverse mills-raten (seleksjonsleddet) er lik null. Relevant å resonnerer på at identifikasjonen blir mer troverdig dersom det kan inkluderes en variabel i seleksjonslikninga som kan antas å bare påvirke valget mellom organisering/ikke-organisering uten å påvirke lønna direkte (eksklusjonsforutsetning) siden det da ikke bare er ikke-linearitet i seleksjonsleddet i 2.steget som bidrar til identifikasjon.

b) Her må kandidatene utvide lønnslikninga med dummy for privat eller offentlig sektor. Håndtering av seleksjon i treatment-modeller er behandlet i Verbeek kap. 7.7 s. 265. Bør vise hvorfor naiv estimering av lønnspremien i den utvidede lønnslikninga med OLS gir skjevhet pga endogen seleksjon, og vise hvordan seleksjonsproblemet kan håndteres med en to-stegsprosedyre ala det som er beskrevet s.265 i Verbeek. Bør også her som i spm a) påpeke at identifikasjonen blir mer troverdig dersom det kan inkluderes en variabel i seleksjonslikninga som bare påvirker seleksjonen uten å påvirke lønna direkte.

c) Her er det mulig å demonstrere betydelig innsikt i identifikasjonsproblemet med utgangspunkt i en enkel panelmodell med to 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 lønnspremie i et slikt opplegg krever

1) variasjon i organiseringsstatus for enkeltindivider mellom 2006 og 2010.

2) kausal tolking (troverdig anslag på lønnspremien) i modell med individfaste effekter krever at endring i organiseringsstatus for individer er ukorrelert med den idiosynkratiske

restleddskomponenten. Dvs. at valget mellom organisering/ikke organisering 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 fagforeningsstatus over tid innen individ, uten å påvirke lønna direkte være relevant. Men slik variabel er selvsagt vanskelig å finne i praksis.

### Oppgave 3.

a) Estimering av modellene med ML-metoden forklares, se s. 211 i Verbeek. Likelihood-funksjonen bør formuleres og få fram hvordan koeffisientene i modellene kan estimeres ved å løse førsteordensbetingelsene for maksimum av log-likelihoodfunksjonen mhp de ukjente koeffisientene.

b) Få fram marginaleffekter i de to ikke-lineære modellene i kolonne (2) og (3) og at de blir lik tettheten til de underliggende fordelingsfunksjonene (normal eller logistisk) evaluert i eksplisitte verdier, f.eks utvalgsgjennomsnitt og estimerte koeffisienter) multiplisert med estimert koeffisient, se Verbeek s. 209. Tolking av marginaleffekter er rett fram effekten av en marginal (en enhets) økning i forklaringsvariabelen (h.h.v. GPA og TUCE) på sannsynligheten for å få A eller B på makroeksamen.

I kolonne (1) blir marginaleffektene av GPA og TUCE lik estimerte koeffisienter (0.464 og 0.010), i kolonnene (2) og (3) blir marginaleffektene henholdsvis 0.534 og 0.017 og 0.533 og 0.017.

c) Marginaleffekt av binær forklaringsvariabel, PCI evaluert i GPA=3 og TUCE=29.1, framkommer ved å beregne differansen mellom sannsynligheten for utfall lik 1 (A eller B på markroeksamen) når PCI (eksponert for nytt undervisningsopplegg på innføringskurset) er «slått på» (lik 1) og «slått av» (lik 0)

I logitmodell, kolonne (2) tabell 2.

$$\frac{e^{-13.021+2.826 \cdot 3+0.095 \cdot 29.1+2.3759}}{1+e^{-13.021+2.826 \cdot 3+0.095 \cdot 29.1+2.3759}} - \frac{e^{-13.021+2.826 \cdot 3+0.095 \cdot 29.1}}{1+e^{-13.021+2.826 \cdot 3+0.095 \cdot 29.1}}$$

I probitmodell, kolonne (3) tabell 2.

$$\Phi(-7.452 + 3 \cdot 1.6386 + 1.426) - \Phi(-7.452 + 3 \cdot 1.6386)$$

der  $\Phi$  er fordelingsfunksjonen for standard normalfordelt variabel.

d) Test av nullrestriksjoner på koeffisientene foran GPA og TUCE, gjøres med Likelihood-ratio test, se også oppgave 1, spm e). Basert på opplysningene blir LR=8.664. Kritisk verdi i Kvikvadratfordelinga med to frihetsgrader og 5% signifikansnivå er 5.99, og nullhypotesen forkastes.