

## Eksamensopgave 2: OLS og Misspecifikation

Betragt følgende to modeller for bankansattes løn:

$$salary = \beta_0 + \beta_1 educ + \beta_2 salbegin + \beta_3 male + \beta_4 minority + u \quad (1)$$

og

$$\log(salary) = \beta_0 + \beta_1 educ + \beta_2 \log(salbegin) + \beta_3 male + \beta_4 minority + u \quad (2)$$

hvor *salary* er årsløn (i 1000 US dollars), *educ* er uddannelse målt i antal år, *salbegin* er startlønnen (i 1000 US dollars) for personens første stilling i samme bank, *male* er en dummy-variabel for køn og *minority* er en dummy-variabel der angiver om man tilhører en minoritet.

Datasættet *data2*, som er tilgængelig på Moodle, indeholder disse variable målt for 450 bankansatte.

Nedenfor er der en række opgaver der skal løses. I forbindelse med de enkelte opgaver forventes det at der redegøres for den relevante teori. Det er altså ikke tilstrækkeligt blot at præsentere et "facit" for hver opgave.

### Opgaver

1. Estimer de to modeller vha. OLS. Kommenter på outputtet, sammenlign og fortolk resultaterne.
2. Udfør grafisk modelkontrol af de to modeller. Hvilken model vil du foretrække?
3. Undersøg om de to modeller er misspecificerede vha. RESET-testet.
4. Forklar hvorfor det kunne være relevant at medtage  $educ^2$  som forklarende variabel i de to modeller. Estimer de to modeller igen hvor  $educ^2$  inkluderes (med tilhørende koefficient  $\beta_5$ ), kommenter kort på outputtet og udfør RESET-testet igen.
5. Test hypotesen  $H_0: \beta_1 = \beta_5 = 0$  i begge modeller (fra spørgsmål 4).
6. Kunne der være problemer med målefejl i de to modeller? I hvilke tilfælde vil det udgøre et problem?