

## Eksamenssæt 3: Instrumentvariable

### Opgave 1 - Estimer modellen vha. OLS og kommenter på resultaterne

```
model = lm(learnings ~ educ + exp + male + ethblack + ethhisp, data)
summary(model)

##
## Call:
## lm(formula = learnings ~ educ + exp + male + ethblack + ethhisp,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.075849 -0.280064 -0.001448  0.307748  1.984409
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  0.3962265   0.1735078   2.2836    0.022801 *
## educ         0.1242201   0.0094515  13.1429 < 0.00000000000000022 ***
## exp          0.0338820   0.0050456   6.7152    0.00000000004986 ***
## male         0.2934491   0.0458032   6.4067    0.00000000033631 ***
## ethblack     -0.1956696   0.0712545  -2.7461    0.006243 **
## ethhisp      -0.0974063   0.1003417  -0.9707    0.332132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.51025 on 514 degrees of freedom
## Multiple R-squared:  0.35393,    Adjusted R-squared:  0.34764
## F-statistic: 56.316 on 5 and 514 DF,  p-value: < 0.000000000000000222
```

hispanics er insignifikant med p-værdi på 33,2% Resten er på 1% eller lavere Relativ lav  $R^2$  F-test med meget lav p-værdi - Variable er “jointly significant”

### Opgave 2 - Hvorfor kunne vi være bekymrede for at uddannelse er endogen?

Uddannelse vil være endogen hvis den er korreleret med en udeladt variabel som derfor skaber en bias. Denne udeladte variabel kunne eksempelvis være “ability”, som påvirker uddannelsesniveaueet positivt og dermed skaber en bias.

### Opgave 3 - Er siblings, meduc og feduc brugbare som instrumenter?

Hvis de nævnte variable er korreleret med uddannelse, mens de ikke er korreleret med den udeladte variabel, som i dette tilfælde er “ability”, vil de være egnet som instrumenter. Det kan formentlig antages, at forældre uddannelse eller antal søskende uddannelse ikke har indflydelse på “ability”, hvorfor denne betingelse til instrumentvariablen er opfyldt. Samtidig er forældres uddannelsesniveau formentlig delvist korreleret med den pågældendes uddannelse, mens antal søskende ikke i samme grad antages at være korreleret med uddannelsesniveauet. Derfor vil forældres uddannelse formentlig være bedre instrumenter end antal søskende.

### Opgave 4 - Test om uddannelse er endogen

Testen for endogenitet laves vha. den reduceret ligning, som er variabelen mistænkt for endogenitetsproblemer regressed på de øvrige uafhængige variable og instrumentvariablene. Heri er variabelen eksogen, altså ukorreleret med det oprindelige fejledd ( $u$ ), hvis og kun hvis fejleddet fra den reducerede ligning ( $v$ ) er ukorreleret med ( $u$ ). Fejleddet fra den reducerede ligning ( $v$ ) er dog ikke observeret, hvorfor residualet bruges som proxy. Derfor inkluderes ( $v$ ) i den oprindelige regression, hvorefter en t-test bruges til teste hvorvidt den tilhørende estimator  $\delta$  er signifikant. Hvis det findes, at  $\delta$  ikke kan siges at være lig 0 er variabelen endogen. Modsat vil variabelen antages at være eksogen hvis nulhypotesen  $H_0 : \gamma = 0$  ikke kan afvises.

```
red_model = lm(educ ~ exp + male + ethblack + ethhisp + siblings + meduc + feduc, data)
v = resid(red_model)
```

```
endo_model = lm(learnings ~ educ + exp + male + ethblack + ethhisp + v, data)
summary(endo_model)
```

```
##
## Call:
## lm(formula = learnings ~ educ + exp + male + ethblack + ethhisp +
##      v, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.121193 -0.279066 -0.003089  0.298275  2.074787
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) -0.064700   0.339003  -0.1909      0.84871
## educ         0.153036   0.020516   7.4593 0.00000000000003742 ***
## exp          0.037628   0.005567   6.7591 0.000000000000378421 ***
## male         0.290479   0.045775   6.3458 0.00000000004868537 ***
## ethblack     -0.157544   0.075122  -2.0972      0.03647 *
## ethhisp      -0.069476   0.101739  -0.6829      0.49499
## v            -0.036550   0.023106  -1.5818      0.11430
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.50951 on 513 degrees of freedom
## Multiple R-squared:  0.35706,    Adjusted R-squared:  0.34955
## F-statistic: 47.484 on 6 and 513 DF,  p-value: < 0.000000000000000222
```

Opgave 5 - Estimer modellen vha. 2SLS hvor du gør brug af de tre beskrevne instrumenter. Sammenlign med resultaterne i spørgsmål 1.

```
educ_fitted = fitted(red_model)

linearHypothesis(red_model, c("meduc=0", "feduc=0", "siblings=0")) #Test at IVs er signifikante
```

```
## Linear hypothesis test
##
## Hypothesis:
## meduc = 0
## feduc = 0
## siblings = 0
##
## Model 1: restricted model
## Model 2: educ ~ exp + male + ethblack + ethhisp + siblings + meduc + feduc
##
##      Res.Df      RSS Df Sum of Sq      F           Pr(>F)
## 1      515 2914.55
## 2      512 2297.79  3    616.752 45.8087 < 0.000000000000000222 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sls = lm(learnings ~ educ_fitted + exp + male + ethblack + ethhisp, data)
#summary(sls)
screenreg(list(OLS = model, two_SLS = sls), digits = 4)
```

```
##
## =====
##              OLS              two_SLS
## -----
## (Intercept)    0.3962 *      -0.0647
##                (0.1735)      (0.3762)
## educ           0.1242 ***
##                (0.0095)
```

```
## exp          0.0339 ***    0.0376 ***
##              (0.0050)      (0.0062)
## male         0.2934 ***    0.2905 ***
##              (0.0458)      (0.0508)
## ethblack     -0.1957 **    -0.1575
##              (0.0713)      (0.0834)
## ethhisp      -0.0974       -0.0695
##              (0.1003)      (0.1129)
## educ_fitted          0.1530 ***
##                  (0.0228)
## -----
## R^2           0.3539       0.2065
## Adj. R^2      0.3476       0.1988
## Num. obs.    520          520
## =====
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

*#Nedenstående er 2SLS lavet i R*

```
#sls_r = ivreg(learnings ~ educ + exp + male + ethblack + ethhisp | meduc + feduc + siblings + exp + ma
#summary(sls_r)
```

**Opgave 6 - Udfør overidentifikationstestet. Hvad konkluderer du?**

**Opgave 7 - Udfør hele analysen igen hvor du kun bruger meduc og feduc som instrumenter. Ændrer det på dine konklusioner?**