Økonometri I: Ugeseddel 4

Levi van Boekel

Uge 10, 2025

Opgave 1

(A)

Hvis den sande model er:

$$y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i,$$

Men vi i stedet kun estimerer:

$$y = \beta_0 + \beta_1 x_{1i} + \underbrace{\tilde{u}}_{\beta_2 x_2 + u}$$

Her indeholder fejlledet både det oprindelige fejlled og den forklarende variabel, der nu ikke længere medtages. Vi forventer en negativ bias, fordi $\beta_2 < 0$ og x_2 afhænger positivt af x_1. Vi bruger nu formlen fra opgaven:

$$\begin{aligned}
plim(\tilde{\beta}_1) - \beta_1 &= \frac{cov(x_1, \beta_2 x_2 + u)}{var(x_1)} \\
&= \frac{\text{cov}(x_1, \beta_0) + \text{cov}(x_1, \beta_1 x_2) + \text{cov}(x_1, u)}{var(x_1)} \\
&= \frac{\beta_1 \, \text{cov}(x_1, \overbrace{\rho x_1 + x_2^\star}^{x_2})}{var(x_1)} \\
&= \frac{\rho \beta_1 var(x_1) + \beta_1 \text{cov}(x_1, x_2^\star)}{var(x_1)} = \rho \beta_1 \\
&= 0.05 \cdot (-3) = -1.5
\end{aligned}$$

(B)

MLR 1 er opfyldt, fordi modellen er lineær i parametrene (der er ingen parametre, der er ganget sammen, opløftet i hinanden eller lign).

MLR 2 er ligeledes opfyldt, idet alle observationer trækkes via STATA, hvor der er uafhængighed imellem dem.

MLR 3 er ligeledes opfyldt, fordi x_1 og x_2 genereres sådan i STATA, at der ikke er nogen lineæresammenhænge imellem dem (den ene variable dannes som normalfordeling og den anden som summen af normalfordeling og uniform fordeling). Der er heller ingen af de forklarende variable, der har den samme værdi for alle observationer.

MLR 4 er ligeledes opfyldt, fordi den middelværdien af fejlledet blot er:

$$E(u) = \frac{a+b}{2} = \frac{0}{2} = 0$$

Hvorfor den betingede middelværdi af fejlledet ligeledes er 0. Da MLR 1 til MLR 4 er opfyldt, OLS estimatoren middelret.

(C)

MC simulationer kan bruges til at sammenligne variansen af to middelrette estimatorer ved at generere data på baggrund af gæt for β_1 og β_2, og specificere en DGB proces for variablene og fejlleddet. Herefter kan vi på baggrund af det generede data bestemme estimatorerne i f.eks. 1000 iterationer (dvs. vi generer data 1000 gange og lader STATA finde estimatoren).

Opgave 2

(A)

Vi starter med at indse, at en standardnormalfordeling kan rykkes til $N(a, b)$ ved:

$$Z = a + \sqrt{b} \cdot N(0,1)$$

Fordi $E(Z) = a$ og $var(Z) = \sqrt{b}^2 1 = b$. Derfor bruger vi $x_1 = 25 + 5 \cdot rnormal()$. Herefter indser vi, at en standard uniform fordeling rykkes til $U(a, b)$ ved:

$$Y = a + (b - a)\,U.$$

Derfor bruger vi $u = -50 + 100 \cdot runiform()$ og $x_2^\star = 10 + 20 \cdot runiform()$.

```
cap log close
clear all
eststo clear
log using "PS1.txt", text replace

```

```
*PRE*

***********Directory Settings************
global path "/Users/levivanboekel/Desktop/ku/ ko /ugesedler"
global outtex "/Users/levivanboekel/Desktop/ku/ ko /ugesedler/a1.tex"

******************************************

***********Data************
cd "/Users/levivanboekel/Desktop/ku/ ko /ugesedler"
******************************************

clear all

global numobs = 50

global beta_0 =1
global beta_1=2
global beta_2=2
global rho = 0.5


*DEFINE PROGRAM THAT SPECIFIES THE DGP
program olsdata, rclass
  drop _all

  *SET NUMBER OF CURRENT OBSERVATIONS
  set obs $numobs

  *DATA GENERATING PROCESS
  generate rho = $rho
  generate beta_0 = $beta_0
  generate beta_1 = $beta_1
  generate beta_2 = $beta_2
  generate x1 = 25 + 5*rnormal()
  generate x2star = 10 + 20*runiform()
  generate x2 = $rho *x1 + x2star
  generate u = -50 + 100*runiform()
  generate y = $beta_0 + $beta_1 * x1 + $beta_2 * x2 + u

  *CALCULATE OLS ESTIMATES AND SAVE RESULTS
  regress y x1
  return scalar betahat_slr = _b[x1]
  return scalar se_slr    = _se[x1]
  regress y x1 x2
  return scalar betahat_mlr = _b[x1]
  return scalar se_mlr    = _se[x1]
```

```
end

olsdata
```

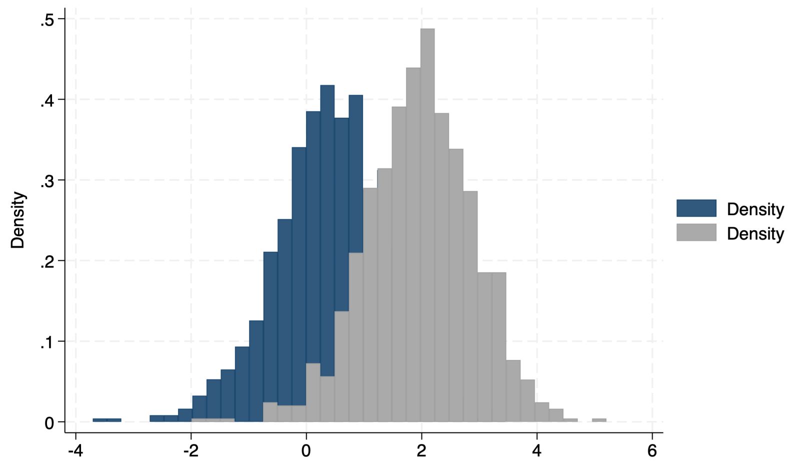
(B)

Resultaterne af MC simulationen er:

Tabel 1: Resultater af MC $n = 50$

	mean	sd	min	max	count
betahat_slr	.4900462	1.002646	-3.70497	3.446828	1000
se_slr	.9697383	.1261159	.6689053	1.57614	1000
betahat_mlr	1.966271	.9253823	-1.99193	5.202202	1000
se_mlr	.9179777	.1138071	.6226666	1.384858	1000

Vi ser, at $\beta^1_{SLR} - \beta^1_{MLR} = -1.5$, hvorfor vi umiddelbart kan bekræfte det teoretiske resultat vedrørende bias. Et histogram for fordelingen af de to estimatorer er vist nedenfor:



Figur 1: Histogram af β^1_{SLR} og β^1_{MLR} $n = 50$

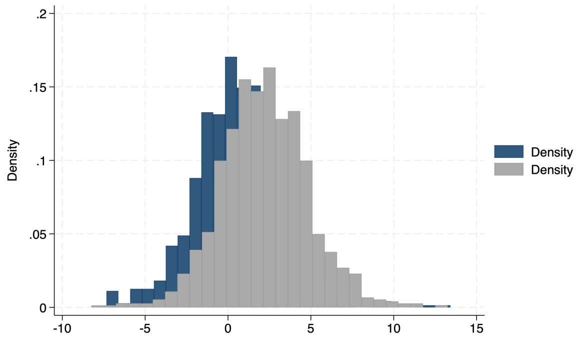
(C)

Resultaterne af MC simulationen for $n = 10$ er:

Tabel 2: Resultater af MC for $n = 10$

	mean	sd	min	max	count
betahat_slr	.7477973	2.667016	-7.333603	13.42256	1000
se_slr	2.413528	.8551656	.7495107	10.12099	1000
betahat_mlr	2.216393	2.649948	-8.246656	13.25287	1000
se_mlr	2.371828	.9261441	.6016496	12.28955	1000

Her er histogrammet for fordelingerne:



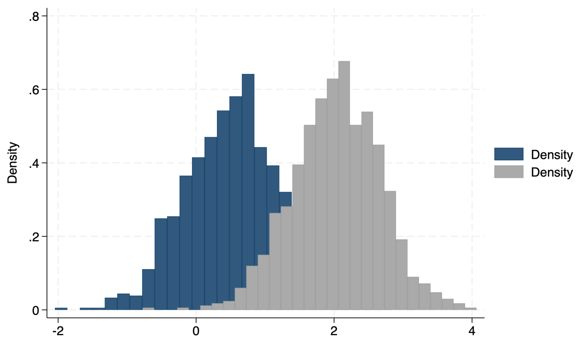
Figur 2: Histogram af β^1_{SLR} og β^1_{MLR} ved $n = 10$

Resultaterne af MC simulationen for $n = 100$ er:

Tabel 3: Resultater af MC for n=100

	mean	sd	min	max	count
betahat_slr	.5326974	.6882016	-2.046489	3.193244	1000
se_slr	.6827607	.0630891	.5243197	.8897234	1000
betahat_mlr	2.032808	.6408089	-.7744344	4.065322	1000
se_mlr	.6393965	.0562104	.4838415	.8145825	1000

Her er histogrammet for fordelingerne:



Figur 3: Histogram af β^1_{SLR} og β^1_{MLR} ved $n = 100$

Vi ser, at estimaterne går mod deres sande værdi, når n bliver større, hvorfor estimatorerne er konsistente.

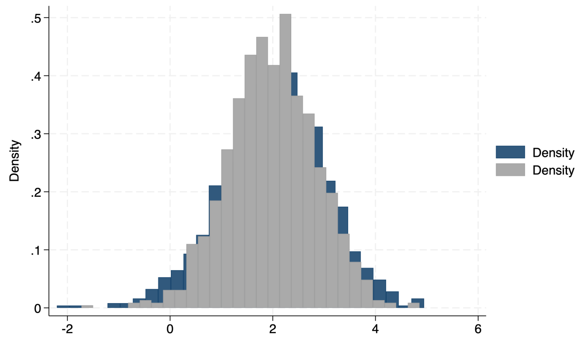
(D)

Vi ser, at estimatoren fortsat er middelret, når $\rho = 0$, da vi her får:

Tabel 4: Resultater af MC for n=50 og $\rho = 0$

	mean	sd	min	max	count
betahat_slr	1.990046	1.002646	-2.20497	4.946828	1000
se_slr	.9697383	.1261159	.6689053	1.57614	1000
betahat_mlr	1.983942	.8508746	-1.727034	4.857749	1000
se_mlr	.8446207	.1060272	.5986493	1.354198	1000

Og:



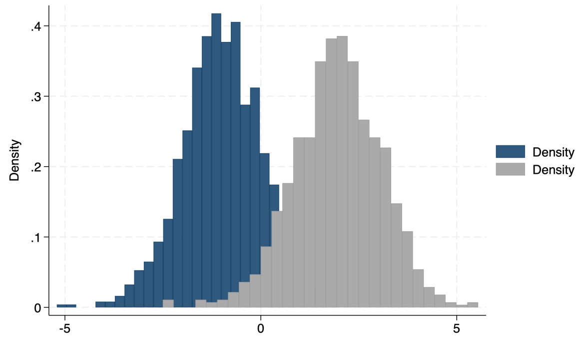
Figur 4: Histogram af β^1_{SLR} og β^1_{MLR} ved $n = 50$ og $\rho = 0$

Årsagen til, at vi nu ikke længere har nogen bias er, at $cov(x_1, x_2)$ er lig 0, fordi vi har sat $\rho = 0$. Hvis vi sætter $\rho = 1$ har vi:

Tabel 5: Resultater af MC for n=50 og $\rho = 1$

	mean	sd	min	max	count
betahat_slr	-1.009954	1.002646	-5.20497	1.946828	1000
se_slr	.9697383	.1261159	.6689053	1.57614	1000
betahat_mlr	1.9486	1.115864	-2.502154	5.546656	1000
se_mlr	1.110899	.1318331	.7185151	1.684046	1000

Og:



Figur 5: Histogram af β^1_{SLR} og β^1_{MLR} ved $n = 50$ og $\rho = 1$

Vi har stadigvæk en middelret estimator, fordi de to variablene x_1 og $x_2^\star$ ikke er lineært afhængige. Variansen er størst i tilfældet af $\rho = 1$, da der her vil være en bias $(= 3)$, mens vi i tilfældet $\rho = 0$ ikke har nogen bias i estimationen af β_1. Vi ville dermed foretrække den multiple regressionsmodel for $\rho = 0$, hvis vi vil have estimatoren med den laveste varians (fordi den simple indeholder $\beta_2 x_2$ i sit fejlled).

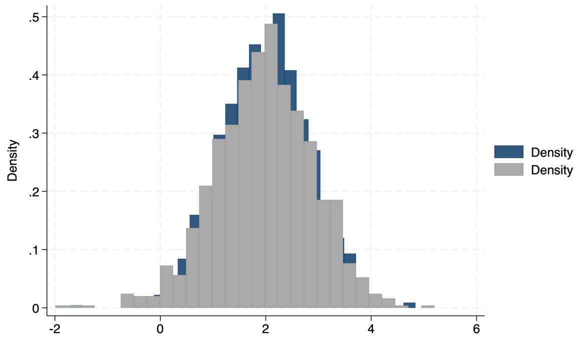
(E)

Hvis $\beta_2 = 0$ og $\rho = 0.5$ vil det være at foretrække at bruge SLR, da vi ikke 'vinder' noget ved at inkludere irrelevante variable, det øger blot estimatorens varians.

Tabel 6: Resultater af MC for n=50

	mean	sd	min	max	count
betahat_slr	1.986588	.8390779	-1.698357	4.839907	1000
se_slr	.8353595	.1033834	.6011781	1.333427	1000
betahat_mlr	1.966271	.9253823	-1.99193	5.202202	1000
se_mlr	.9179777	.1138071	.6226666	1.384858	1000

Og:



Figur 6: Histogram af β^1_{SLR} og β^1_{MLR} ved $n = 50$ og $\rho = 0.5$, samt $\beta_2 = 0$