

Økonometri I Forår 2021 Obligatorisk opgave

1 Deadline: 15. marts, 23:59

Christian Birk, Carl Buan
Peter Ravn

March 2021

Lavet af:

Navn og studie ID, person 1: Carl Buan - tqn889

Navn og studie ID, person 2: Christian Birk - ldg790

Navn og studie ID, person 3: Peter Langgård Ravn - pmv384

Problem 1

I nedenstående figur ser vi en deskriptiv tabel over vores data. Vi ser, at der er 98 kommuner, og for hver kommune er der data for skatteprovenu (taxrev), skatteprocent (taxrate) og indbyggertal (pop). Først ser vi at der er en høj varians i 'taxrev' og 'pop', da standardafvigelsen er stor. Eftersom gennemsnittet for 'taxrev' og 'pop' er højere end medianen, så vil begge variable være højskæve. Det vil sige, at der er større outliers til højre for gennemsnittet.

For 'taxrate' er standardafvigelsen meget lille, hvilket vil sige, at kommunerne har meget den samme skatteprocent. Det ses, at skatteprocenten ligger lige omkring 25 pct.

(1)								
	count	min	max	mean	p25	p50	p75	sd
taxrev	98	211.23	44170.34	4477.34	2414.52	3317.85	4848.51	5251.18
taxrate	98	22.80	27.80	25.21	24.80	25.30	25.70	0.91
pop	98	1969.00	528208.00	56475.89	29638.00	43475.00	59978.00	62925.30
N	98							

Problem 2

1

Regressionsmodellen er en log-level-model, som viser sammenhængen mellem skatteprovenu og skatteprocenten. Når vi har med en log-level-model at gøre, så vil én enhedsændring i 'taxrate' give en $\delta_1 \cdot 100$ pct.-vis ændring i 'taxrev'. Det vil altså sige, at når skatteprocenten stiger med 1, så skal skatteprovenu stige med $\delta_1 \cdot 100$ pct.

2

Man må regne med, at en højere skatteprocent vil føre til højere skatteprovenu. Ifølge ovenstående svar, så vil højere skatter medføre større skatteprovenu, hvilket giver $\delta_1 > 0$. Som laffer-kurven også indikerer, så må δ_1 have et maksimum, da en skatteprocent på 100pct. alt andet lige må give et provenu på nul, da ingen da vil arbejde. Hvis skatteprocenten bliver tilpas høj, så må man derfor regne med, at $\delta_1 < 0$.

3

Vi estimerer vores model ved brug af OLS:

Source	SS	df	MS	Number of obs	=	98
Model	1.62662646	1	1.62662646	F(1, 96)	=	2.82
Residual	55.4155865	96	.577245693	Prob > F	=	0.0965
				R-squared	=	0.0285
				Adj R-squared	=	0.0184
Total	57.0422129	97	.588064051	Root MSE	=	.75977

Intaxrev	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
taxrate	-.1426169	.0849586	-1.68	0.096	-.3112584	.0260246
_cons	11.69821	2.143025	5.46	0.000	7.444336	15.95208

Mod vores forventning så ser vi, at $\delta_1 < 0$. Det betyder, at en højere skatteprocent alt andet lige vil føre til et lavere skatteprovenu. Kommunerne kan ifølge vores model sænke skatterne og få en større indtjening. Dette resultat kunne tyde på, at vi ligger højt på Laffer-kurven.

Vi ser også på vores output, at 0 er indeholdt i 95pct. konfidensintervallet, og vi ser at R^2 er helt nede på 0,0285. Modellen kan kun forklare 2,85pct. af variationen i skatteprovenuet. Vi kan altså ikke på et 5pct. signifikansniveau afvise, at der ikke skulle være nogen signifikant forskel på skatteprovenuet i de forskellige kommuner.

4

Fortolkningen af β_1 er præcis den samme som δ_1 , dog holder vi nu populationen konstant, hvilket skyldes β_2 . Vi kan variere på populationsstørrelsen via β_2 , så vores nye model ser bort for populationsstørrelserne. Det vil igen sige, at én enhedsændring i skatteprocenten vil ændre skatteprovenuet med $100 \cdot \beta_1$, når vi har en konstant population.

5

Source	SS	df	MS	Number of obs	=	98
Model	55.9094061	2	27.954703	F(2, 95)	=	2344.35
Residual	1.13280688	95	.011924283	Prob > F	=	0.0000
				R-squared	=	0.9801
				Adj R-squared	=	0.9797
Total	57.0422129	97	.588064051	Root MSE	=	.1092

Intaxrev	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
taxrate	.0226223	.0124539	1.82	0.072	-.0021019	.0473465
lnpop	.9711011	.0143929	67.47	0.000	.9425275	.9996747
_cons	-2.802166	.375576	-7.46	0.000	-3.547779	-2.056554

Vi ser nu at $\hat{\beta}_1$ er positiv, hvilket også er forventet. Dog ser vi at den er tæt på 0, hvilket ikke er signifikant på en 95 pct. konfidensinterval. Desuden ser vi at populationen har en væsentlig betydning for modellen. Vi ser netop, at én enhedsændring i populationen medfører en 0.97 pct. stigning i skatteprovenu. Yderligere ser vi, at vores R^2 er steget til 98 pct. Det betyder at modellen forklarer 98 pct. af variationen i skatteprovenu, fremfor 2.8 pct. som vores tidligere model viste. Alt i alt, er populationen en væsentlig variabel.

6

Udeladt variabel bias i vores tilfælde er populationen. Som vi fandt ud af i foregående spørgsmål, er populationen en meget relevant variabel for at beskrive skatteprovenu. Denne er korreleret med skatteprocenten. Dermed kan vi nu bestemme bias ud fra følgende formel:

$$\delta_1 = -0.143, \quad \beta_2 = 0.971$$

$$Bias(\tilde{\beta}_1) = \beta_2 * \delta_1 = 0.971 * (-0.143) = -0.1389$$

Vi ender altså med en negativt bias, hvilket betyder at populationen har en negativ effekt på skatteprocenten. Dette giver intuitivt mening, da en højere skatteprocent kan medføre, at befolkningen flytter over til en anden kommune med lavere skatteprocent. Derudover har store kommuner muligvis ikke brug for lige så høj skatteprocent, da de allerede har et højt provenu. Vi forventer at $\beta_2 > 0$, eftersom at større befolkning medfører større skatteprovenu.

Problem 3

1

Følgende udregning er lavet i lyx, og kan derfor se en smule anderledes ud:

Vi skal opstille modellen:

$$y = \beta_1 x_1 + \beta_2 x_2 + u$$

Vi får følgende:

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 0 & x_{11} & x_{12} \\ 0 & x_{21} & x_{22} \\ \dots & \dots & \dots \\ 0 & x_{n1} & x_{n2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \dots & \dots \\ x_{n1} & x_{n2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix}$$

For $\hat{\beta}$ gælder:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1} X'y$$

Vi finder delkomponenterne i udtrykket:

$$X'X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \dots & \dots \\ x_{n1} & x_{n2} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} \\ \sum_{i=1}^n x_{i1}x_{i2} & \sum_{i=1}^n x_{i2}^2 \end{bmatrix}$$

En transponeret matrix regnes som:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$[X'X]^{-1} = \frac{1}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left(\sum_{i=1}^n x_{i1}x_{i2}\right)^2} \begin{bmatrix} \sum_{i=1}^n x_{i2}^2 & -\sum_{i=1}^n x_{i1}x_{i2} \\ -\sum_{i=1}^n x_{i1}x_{i2} & \sum_{i=1}^n x_{i1}^2 \end{bmatrix}$$

$$X'y = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_{i1}y_i \\ \sum_{i=1}^n x_{i2}y_i \end{bmatrix}$$

Nu kan vi sætte det hele sammen:

$$\begin{aligned} [X'X]^{-1} * X'y &= \frac{1}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left(\sum_{i=1}^n x_{i1}x_{i2}\right)^2} \begin{bmatrix} \sum_{i=1}^n x_{i2}^2 & -\sum_{i=1}^n x_{i1}x_{i2} \\ -\sum_{i=1}^n x_{i1}x_{i2} & \sum_{i=1}^n x_{i1}^2 \end{bmatrix} \cdot \begin{bmatrix} \sum_{i=1}^n x_{i1}y_i \\ \sum_{i=1}^n x_{i2}y_i \end{bmatrix} \\ &= \frac{1}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left(\sum_{i=1}^n x_{i1}x_{i2}\right)^2} \begin{bmatrix} \sum_{i=1}^n x_{i2}^2 \sum_{i=1}^n x_{i1}y_i - \sum_{i=1}^n x_{i1}x_{i2} \sum_{i=1}^n x_{i2}y_i \\ -\sum_{i=1}^n x_{i1}x_{i2} \sum_{i=1}^n x_{i1}y_i + \sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}y_i \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \frac{\sum_{i=1}^n x_{i2}^2 \sum_{i=1}^n x_{i1} y_i - \sum_{i=1}^n x_{i1} x_{i2} \sum_{i=1}^n x_{i2} y_i}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left(\sum_{i=1}^n x_{i1} x_{i2} \right)^2} \\ - \frac{\sum_{i=1}^n x_{i1} x_{i2} \sum_{i=1}^n x_{i1} y_i + \sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2} y_i}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left(\sum_{i=1}^n x_{i1} x_{i2} \right)^2} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

Herfra regnes $\hat{\beta}_1$ som:

$$\beta_1 = \frac{\sum_{i=1}^n x_{i2}^2 \sum_{i=1}^n x_{i1} y_i - \sum_{i=1}^n x_{i1} x_{i2} \sum_{i=1}^n x_{i2} y_i}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left(\sum_{i=1}^n x_{i1} x_{i2} \right)^2}$$

$$= \frac{\sum_{i=1}^n x_{i1} y_i - \frac{\sum_{i=1}^n x_{i1} x_{i2}}{\sum_{i=1}^n x_{i2}^2} \sum_{i=1}^n x_{i2} y_i}{\sum_{i=1}^n x_{i1}^2 - \frac{\left(\sum_{i=1}^n x_{i1} x_{i2} \right)^2}{\sum_{i=1}^n x_{i2}^2}}$$

Vi har fundet det, som vi skulle.

2

(a)

3

Vi laver en simpel regression af 'taxrate' på 'ln(pop)'.

Source	SS	df	MS	Number of obs	=	98
Model	3.09262994	1	3.09262994	F(1, 96)	=	3.86
Residual	76.8808485	96	.800842172	Prob > F	=	0.0523
				R-squared	=	0.0387
				Adj R-squared	=	0.0287
Total	79.9734784	97	.82446885	Root MSE	=	.8949

taxrate	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnpop	-.2272655	.1156493	-1.97	0.052	-.4568275	.0022965
_cons	27.62685	1.234119	22.39	0.000	25.17714	30.07656

Koefficienten er -0.227, hvilket vil sige, at der er en negativ sammenhæng mellem skatteprocenten og størrelsen af indbyggerne i kommunerne. En negativ korrelationskoefficient vil føre til en negativ kovarians, og så vil MLR.4 ikke være overholdt i vores simple regressionsanalyse, da vil vi netop have en udeladt variabel bias, fordi befolkningsstørrelsen har korrelation med de andre forklarende variable.

Hvis vi gemmer residualerne og laver en regression af 'ln(taxrev)' på residualerne, så får vi:

Source	SS	df	MS	Number of obs	=	98
Model	.039345082	1	.039345082	F(1, 96)	=	0.07
Residual	57.0028679	96	.593779874	Prob > F	=	0.7974
				R-squared	=	0.0007
				Adj R-squared	=	-0.0097
Total	57.0422129	97	.588064051	Root MSE	=	.77057

ln(taxrev)	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
res1	.0226223	.0878828	0.26	0.797	-.1518236	.1970682
_cons	8.103097	.0778394	104.10	0.000	7.948587	8.257607

Vi ser her, at denne koefficient bliver den samme, som da vi lavede en regression af 'ln(taxrev)' på 'taxrate' og 'ln(pop)'. Dette passer meget fint overens med vores teoretiske estimat, der netop viste, at $\hat{\beta}_1 = \hat{\eta}_1$.

Problem 4

Vi har foretaget en analyse på alle de 98 kommuner i Danmark for at se på sammenhængen mellem skatteprovenu og skatteprocenten med en mulig baggrundsvariabel i befolkningsstørrelse. Hvis vi laver en simpel regressionsanalyse mellem skatteprovenu og skatteprocenten forklares kun 2,9pct. af variationen af skatteprocenten, og vi så, at en én enhedsstigning i skatteprocenten ville give en -14,3pct. fald i skatteprovenu. Økonomisk ville dette betyde, at vi lå til højre for toppunktet på Laffer-kurven . Såfremt vi laver en multipel regressionsanalyse, hvor vi inkluderede befolkningsstørrelsen, så kunne 98 pct af variationen forklares ift. størrelsen af skatteprovenu, og vi kunne se, at én enhedsstigning i skatteprocenten ville give en stigning i provenuet på 2,26pct. Desuden så vi, at MLR.4 ikke ville være overholdt, da der var en negativ korrelation mellem befolkningsstørrelsen og skatteprocenten, hvormed det kan udledes, at befolkningsstørrelsen skal inkluderes.