

Økonometri I: Ugeseddel 7

Levi van Boekel

Uge 13, 2025

Opgave 1

1.1

Hvis variansen af fejleddet *ikke* er konstant (dvs. MLR 5 ikke er opfyldt), men MLR1 til MLR 4 er opfyldt, vil det gælde, at:

- OLS **stadigvæk** er middelret og konsistent
- Variansformlen for OLS estimatoren **ikke længere er gyldig**, dvs. standardfejl og konfidensinterval er heller ikke længere gyldige
- T-test og F-test **ikke længere er gyldige**
- OLS **ikke længere** er BLUE eller asymptotisk efficient

1.2

Vi kan vælge at fortolke fejleddet som en konjunktoreffekt. Hvis MLR4 er opfyldt, svarer det til, at konjunktoreffekten ikke samvarierer med nogle af de andre forklarende variable, dvs. at den er ens i de forskellige bydele og ikke afhænger af m^2 , værelser eller toiletter. MLR4 siger imidlertid ikke noget om variansen af effekten, hvorfor man sagtens kan have en ikke-konstant varians, selvom MLR4 er opfyldt. Vi kan her forestille os, at konjunktoreffekten varierer mere for dyre boliger, men at selve effekten er ukorreleret med de andre, uafhængige, forklarende variable (og den forventede værdi af effekten er 0).

1.3

Hvis den betingede varians er proportionel med m^2_i kan vi benytte WLS, hvor vi dividerer hele regressionsligningen med $\frac{1}{\sqrt{m^2_i}}$. Det ses nemlig nedenfor, at vi hermed opnår konstant varians på

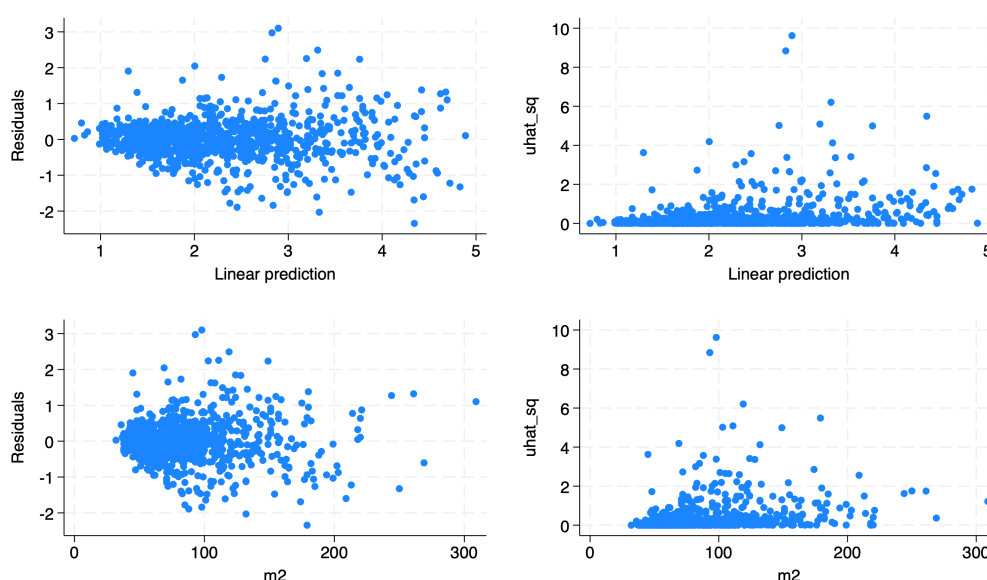
fejlladet.

$$\begin{aligned} \text{Var}(u_i | x) &= \text{Var} \left[\frac{1}{\sqrt{m2_i}} \cdot u_i \mid x \right] \\ &= \frac{1}{m2_i} \cdot \sigma^2 m2_i = \sigma^2 \end{aligned}$$

Opgave 2

2.1 & 2.2

De fire scatterplots er vist nedenfor:



Figur 1: Grafisk test af heteroskedasticitet

Selvom der er tegn på en ikke-konstant varians, er der umiddelbart ikke noget systematisk mønster i fejllederne. For at afgøre homoskedasticitet kræves derfor yderligere test.

2.3

Vi starter med at estimere nedenstående

$$\hat{u}^2 = \delta_0 + \delta_1 KbhK + \delta_2 KbhN + \delta_3 KbhV + \delta_4 m2_i + \delta_5 m2_i^2 + \delta_6 rooms + \delta_7 toilets + v$$

Nullhypotesen for homoskedasticitet er nu, at alle hældningskoefficienter er nul (svarende til en konstant varians), mens alternativhypotesen er, at mindst et $\delta_j \neq 0$ (svarende til en ikke-konstant varians, og dermed heteroskedasticitet). Da:

$$\mathcal{H}_0 : \delta_1 = \dots = \delta_7 = 0 \quad \text{mod} \quad \delta_j \neq 0 \text{ for mindst et } j = 1, \dots, 7$$

Vi noterer, at $R_u^2 = 0.0905$. Dermed bliver LM og F-test størrelsen:

$$F = \frac{\frac{R_u^2}{k}}{\frac{1-R_u^2}{n-k-1}} = \frac{\frac{0.0905}{7}}{\frac{1-0.0905}{987-7-1}} \approx 13.93$$

$$LM = R_u^2 n = 0.0905 \cdot 988 \approx 89.4$$

Det kritiske område for F-testen findes ved:

```
1 display invF(7, 988-7-1, 0.95) = 2.01
```

Mens det kritiske område for LM-testet findes ved:

```
1 display invchi2(7, 0.95) = 14.06
```

I begge tilfælde ligger vores test-statistisk udenfor det kritiske område, hvorfor nulhypotese kan afvises og vi derfor har tegn på heteroskedasticitet (på et 5 pct. signifikansniveau).

2.4

Vi estimerer nu

$$\hat{u}^2 = \delta_0 + \delta_4 m2_i + \delta_5 m2_i^2 + v$$

Nullhypotesen for homoskedasticitet er nu, at alle hældningskoefficienter er nul (svarende til en konstant varians), mens alternativhypotesen er, at mindst et $\delta_j \neq 0$ (svarende til en ikke-konstant varians, og dermed heteroskedasticitet). Da:

$$\mathcal{H}_0 : \delta_1 = \delta_2 = 0 \quad \text{mod} \quad \delta_j \neq 0 \text{ for mindst et } j = 1, 2$$

Vi noterer, at $R_u^2 = 0.0795$ og får dermed $LM = 0.0795 \cdot 988 \approx 78.546$. Det kritiske område er:

```
1 display invchi2(7, 0.95) = 5.99
```

Da test-statistikken er udenfor det kritiske område, afvises nullhypotesen. Vi kan dermed – på et 95 pct. signifikansniveau – afvise, at $m2$ ikke forklarer fejlleddet i anden.

NB: Vi kan også bare teste $\hat{u}^2 = \delta_0 + \delta_4 m2_i + v$ og bestemme t-test statistikken til $t = 9.03$, hvormed nullhypotesen også afvises.

2.5

Variansudtrykket vi bruger nu, er netop konsistent med det sidste test, hvor vi afviste, at $m2$ ikke kunne forklare variansen. De to regressionsresultater er vist nedenfor.

Tabel 1: Regressionsresultater for model (2) med OLS og WLS

	(2) OLS	(2) WLS
	y_scale	y_scale
KbhK	0.192*** (0.0479)	0.207*** (0.0434)
KbhV	-0.0646 (0.0563)	-0.0457 (0.0503)
KbhN	-0.213*** (0.0489)	-0.180*** (0.0414)
m2	0.0371*** (0.00206)	0.0359*** (0.00224)
m2_sq	-0.0000670*** (0.00000780)	-0.0000647*** (0.00000977)
rooms	-0.0547 (0.0286)	-0.0209 (0.0279)
toilets	0.0641 (0.0884)	0.0478 (0.107)
_cons	-0.193 (0.137)	-0.197 (0.150)
N	988	988
R^2	0.674	0.662

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Begge estimater er konsistente (under antagelse MLR1 til MLR4), mens kun WLS resultaterne er efficiente, da MLR5 ikke er overholdt. Vi ser, ligeledes at standardfejlene er mindre for de fleste variable, mens der er nogle, der er større i WLS. Det skyldes, at efficiensen er et teoretisk begreb, og der i en given stikprøve kan være afvigelser.

2.6

Man får approksimativt det samme svar jf. nedenstående:

Source	SS	df	MS	Number of obs	=	988
-----+-----				F(8, 980)	=	2243.59
Model	.755424207	8	.094428026	Prob > F	=	0.0000
Residual	.041246072	980	.000042088	R-squared	=	0.9482
-----+-----				Adj R-squared	=	0.9478
Total	.796670279	988	.000806346	Root MSE	=	.00649

y_scale2	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
-----+-----						
scale	-.1114099	.1711184	-0.65	0.515	-.4472106	.2243907
KbhK_scale	.2053847	.03941	5.21	0.000	.128047	.2827225
KbhN_scale	-.1631036	.0360365	-4.53	0.000	-.2338212	-.092386
KbhV_scale	-.0460217	.0451429	-1.02	0.308	-.1346096	.0425662
m2_scale	.0323978	.0025041	12.94	0.000	.0274838	.0373117
m2_sq_scale	-.0000527	.0000126	-4.16	0.000	-.0000775	-.0000278
rooms_scale	.0227086	.0270821	0.84	0.402	-.0304369	.0758542
toilets_scale	.0319899	.1292193	0.25	0.805	-.2215884	.2855681

2.7

Vi starter med at teste $\delta_3 = 0$, hvor nullhypotesen og alternativhypotesen blot er:

$$\mathcal{H}_0 : \delta_3 = 0 \quad \text{mod} \quad \mathcal{H}_A : \delta_3 \neq 0$$

Vi ser nedenfor, at t-test statistikken er $t = -1.14$ med tilhørende p-værdi på 0.255, hvilket betyder, at der er ca. 25 pct. sandsynlighed for at få et mere ekstremt udfald. Det er indenfor vores grænse på 5 pct. og kan derfor *ikke* afvise nullhypotesen. Vi kan altså ikke afvise, at der er en statistisk signifikant forskel på prisen mellem Østerbro og Vesterbro, når de øvrige forklarende variable holdes fast.

Linear regression	Number of obs	=	988
	F(7, 980)	=	226.79
	Prob > F	=	0.0000
	R-squared	=	0.6738
	Root MSE	=	.56815

		Robust				
y_scale	Coefficient	std. err.	t	P> t	[95% conf. interval]	
KbhK	.1924627	.0528666	3.64	0.000	.0887179	.2962075
KbhV	-.06464	.0567881	-1.14	0.255	-.1760803	.0468004
KbhN	-.2133712	.0411875	-5.18	0.000	-.2941971	-.1325453
m2	.0371188	.0028324	13.10	0.000	.0315605	.0426771
m2_sq	-.000067	.0000118	-5.66	0.000	-.0000902	-.0000437
rooms	-.0547265	.0334594	-1.64	0.102	-.1203869	.0109339
toilets	.064069	.1328357	0.48	0.630	-.1966062	.3247442
_cons	-.1933774	.1833742	-1.05	0.292	-.5532286	.1664738

Herefter tester vi:

$\delta_3 = 0$, hvor nullhypotesen og alternativhypotesen blot er:

$$\mathcal{H}_0 : \beta_3 = \beta_4 = 0 \quad \text{mod} \quad \mathcal{H}_A : \beta_i \neq 0 \quad \text{for mindst et } i = 3, 4$$

Det gøres i STATA nedenfor, hvor vi igen ser, at P-værdien er over 0.05 som er vores kritiske niveau.

Vi kan altså heller ikke afvise, at værelser og toiletter ikke har nogen forklarende effekt.

```
test rooms toilets
```

```
( 1)  rooms = 0
```

```
( 2)  toilets = 0
```

```
F( 2, 980) = 1.70
```

```
Prob > F = 0.1828
```

2.8

$R_u^2 = 0.0041$ og LM-test statistikken bliver dermed:

$$LM = 0.0041 \cdot 988 = 4.0508$$

Det kritiske område er nu:

```
1 display invchi2(7,0.95)= 14.06
```

Hvorfor LM-test statistikken er indenfor det kritiske område og vi ikke kan afvise nullhypotesen, der er identitisk til den i 2.3. Heteroskedasticitet løses altså her ved at implementere en log-transformation.