

# Ekonometrija 1

## Desete vaje:

### *Konstantnost variance slučajne spremenljivke in heteroskedastičnost.*

Na desetih vajah bomo obravnavali eno temeljnih predpostavk klasičnega linearnega regresijskega modela, in sicer homoskedastičnost. Spoznali bomo različne teste za preverjanje te predpostavke ter postopek ocenjevanja regresijskih modelov, kadar varianca slučajne spremenljivke ni konstantna, torej v primeru heteroskedastičnosti.



**Primer 1:** Na podlagi podatkov za 50 ameriških zveznih držav za leto 1985 ocenite naslednjo funkcijo porabe bencina (glej datoteko `bencin.dta`):

$$CON_i = \beta_1 + \beta_2 REGMV_i + \beta_3 TAX_i + u_i.$$

Pri tem imajo posamezne spremenljivke naslednji pomen:

- ♦ *CON*: poraba bencina v posamezni državi (v milijonih sodčkov);
- ♦ *REGMV*: število registriranih motornih vozil (v tisočih);
- ♦ *TAX*: stopnja trošarine za bencin (v centih na galono).

Najprej s pomočjo grafične metode odkrivanja heteroskedastičnosti preverite, ali je v zgornji funkciji porabe bencina mogoče zaznati prisotnost heteroskedastičnosti. Nato opravite naslednje formalne teste, pri katerih predpostavljamo, da heteroskedastičnosti slučajne spremenljivke povzroča pojasnjevalna spremenljivka *REGMV*: Parkov test, Glejserjev test tipa 1 in Goldfeld–Quandtov test. Pojasnite vaše ugotovitve.

Pri vsakem formalnem testu zapišite model v obliki, ki bo zagotavljala, da je predpostavka o homoskedastičnosti ponovno izpolnjena in to tudi dokažite. Ocenite ta model, nato pa zapišite ustrezne (NENALICE) ocene prvotnega regresijskega modela in jih razložite.

### *Izpis rezultatov obdelav v programskem paketu Stata:*

```
. regress con regmv tax
```

Source	SS	df	MS		Number of obs =	50
Model	19504553.9	2	9752276.93		F( 2, 47) =	152.36
Residual	3008446.64	47	64009.5029		Prob > F =	0.0000
Total	22513000.5	49	459448.99		R-squared =	0.8664
					Adj R-squared =	0.8607
					Root MSE =	253

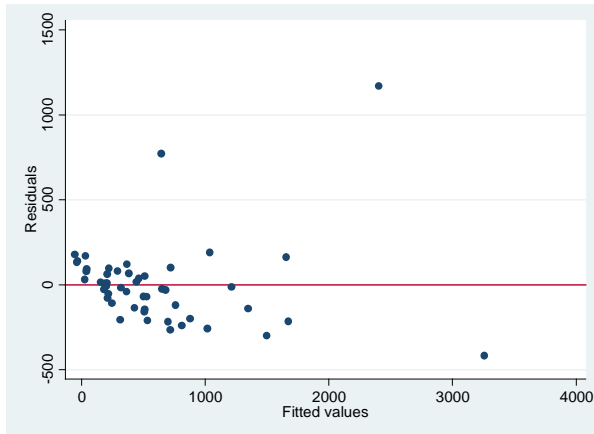
  

con	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
regmv	.1861319	.0117189	15.88	0.000	.1625565 .2097074
tax	-53.59101	16.85588	-3.18	0.003	-87.50067 -19.68134
_cons	551.688	186.2709	2.96	0.005	176.9592 926.4168

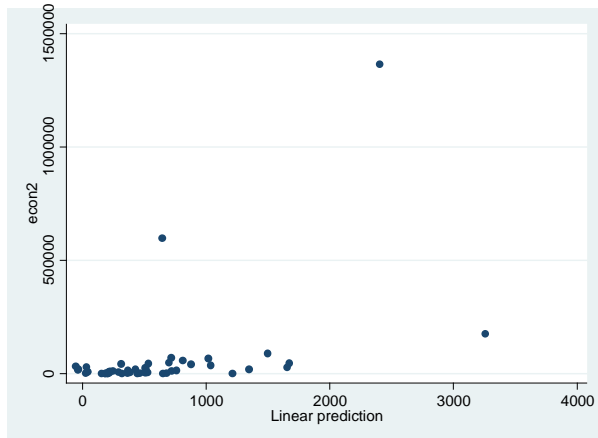
```
. predict econ, resid
. gen econ2=econ^2
. predict confit, xb
```

. \* Graficna metoda

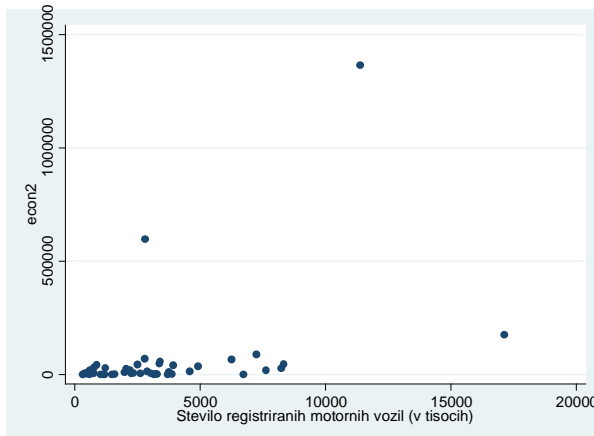
```
. rvfplot, yline(0)
```



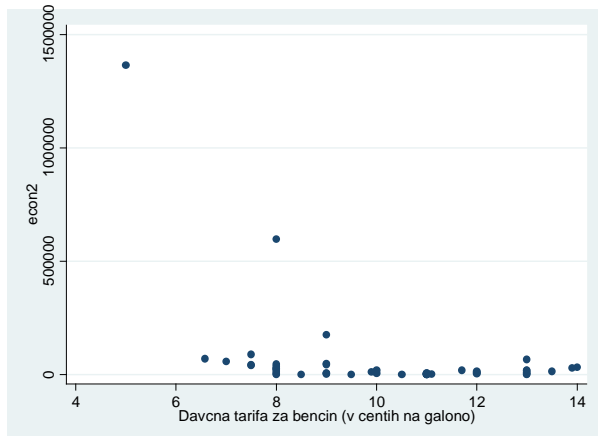
```
. scatter econ2 confit
```



```
. scatter econ2 regmv
```



```
. scatter econ2 tax
```



. \* Parkov test

```
. gen lecon2=log(econ2)
. gen lregmv=log(regmv)
. regress lecon2 lregmv
```

Source	SS	df	MS
Model	41.066836	1	41.066836
Residual	206.772277	48	4.30775578
Total	247.839114	49	5.05794109

```
Number of obs =      50
F( 1, 48) =      9.53
Prob > F      = 0.0033
R-squared     = 0.1657
Adj R-squared = 0.1483
Root MSE     = 2.0755
```

	lecon2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	lregmv	.9519161	.3083036	3.09	0.003	.3320302 1.571802
	_cons	1.650294	2.374469	0.70	0.490	-3.123896 6.424483

```
. gen td=regmv^(_b[lregmv]/2)
```

```
. gen cont=con/td
. gen regmvt=regmv/td
. gen taxt=tax/td
. gen kt=1/td
```

```
. regress cont regmvt taxt kt, nocons
```

Source	SS	df	MS	Number of obs =	50
Model	8725.06082	3	2908.35361	F( 3, 47) =	150.76
Residual	906.687881	47	19.2912315	Prob > F =	0.0000
				R-squared =	0.9059
				Adj R-squared =	0.8999
Total	9631.7487	50	192.634974	Root MSE =	4.3922

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
regmvt	.182248	.0118383	15.39	0.000	.1584325	.2060635
taxt	-22.93641	9.355146	-2.45	0.018	-41.75654	-4.116283
kt	255.3153	99.70358	2.56	0.014	54.73761	455.8931

```
. gen contfit=_b[regmvt]*regmv+_b[taxt]*tax+_b[kt]
```

```
. correlate con contfit
(obs=50)
```

	con	contfit
con	1.0000	
contfit	0.9261	1.0000

```
. scalar r2p=r(rho)^2
. display r2p
.85769622
```

```
. * Glejserjev test, tip 1
```

```
. gen econabs=abs(econ)
```

```
. regress econabs regmv
```

Source	SS	df	MS	Number of obs =	50
Model	517432.009	1	517432.009	F( 1, 48) =	18.43
Residual	1347361.45	48	28070.0303	Prob > F =	0.0001
				R-squared =	0.2775
				Adj R-squared =	0.2624
Total	1864793.46	49	38057.0094	Root MSE =	167.54

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
regmv	.032327	.0075294	4.29	0.000	.0171882	.0474659
_cons	48.25808	33.71507	1.43	0.159	-19.53061	116.0468

```
. drop td cont regmvt taxt kt lecon2 contfit
```

```
. gen td=regmv
```

```
. gen cont=con/td
. gen regmvt=regmv/td
```

```
. gen taxt=tax/td
. gen kt=1/td
```

```
. regress cont regmvt taxt kt, nocons
```

Source	SS	df	MS	Number of obs =	50
Model	2.08732915	3	.695776385	F( 3, 47) =	153.63
Residual	.212852409	47	.004528775	Prob > F =	0.0000
				R-squared =	0.9075
				Adj R-squared =	0.9016
Total	2.30018156	50	.046003631	Root MSE =	.0673

cont	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
regmvt	.1678308	.0136749	12.27	0.000	.1403205 .195341
taxt	-17.389	4.682179	-3.71	0.001	-26.80833 -7.969672
kt	218.5395	48.1033	4.54	0.000	121.7681 315.3108

```
. gen contfit=_b[regmvt]*regmv+_b[taxt]*tax+_b[kt]
```

```
. correlate con contfit
(obs=50)
```

	con	contfit
con	1.0000	
contfit	0.9247	1.0000

```
. scalar r2p=r(rho)^2
. display r2p
.8550451
```

```
. * Goldfeld-Quandt test
```

```
. sort regmv
```

```
. regress con regmv tax if _n<18
```

Source	SS	df	MS	Number of obs =	17
Model	13941.0041	2	6970.50207	F( 2, 14) =	3.37
Residual	28955.937	14	2068.28122	Prob > F =	0.0639
				R-squared =	0.3250
				Adj R-squared =	0.2286
Total	42896.9412	16	2681.05882	Root MSE =	45.478

con	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
regmv	.1025938	.0422156	2.43	0.029	.0120503 .1931373
tax	-8.474985	5.441093	-1.56	0.142	-20.14497 3.194998
_cons	164.6865	57.88825	2.84	0.013	40.5285 288.8444

```
. regress con regmv tax if _n>33
```

Source	SS	df	MS	Number of obs =	17
Model	10024681.3	2	5012340.63	F( 2, 14) =	41.52
Residual	1689986.98	14	120713.355	Prob > F =	0.0000
				R-squared =	0.8557
				Adj R-squared =	0.8351
Total	11714668.2	16	732166.765	Root MSE =	347.44

con	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
regmv	.1858942	.0252233	7.37	0.000	.1317957	.2399928
tax	-112.6418	39.70267	-2.84	0.013	-197.7955	-27.48803
_cons	1084.524	460.8424	2.35	0.034	96.11569	2072.933

```
. display invFtail(17-3,17-3,0.05)
```

```
2.4837257
```

```
. display Ftail(17-3,17-3,1689986.98/28955.937)
```

```
6.042e-10
```

```
. drop td cont regmvt tax kt confit
```

```
. gen td=regmv
```

```
. gen cont=con/td
```

```
. gen regmvt=regmv/td
```

```
. gen taxt=tax/td
```

```
. gen kt=1/td
```

```
. regress cont regmvt taxt kt, nocons
```

Source	SS	df	MS	Number of obs = 50		
Model	2.08732915	3	.695776385	F( 3, 47) =	153.63	
Residual	.212852409	47	.004528775	Prob > F =	0.0000	
				R-squared =	0.9075	
				Adj R-squared =	0.9016	
Total	2.30018156	50	.046003631	Root MSE =	.0673	

cont	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
regmvt	.1678308	.0136749	12.27	0.000	.1403205	.195341
taxt	-17.389	4.682179	-3.71	0.001	-26.80833	-7.969672
kt	218.5395	48.1033	4.54	0.000	121.7681	315.3108

```
. gen confit=_b[regmvt]*regmv+_b[taxt]*tax+_b[kt]
```

```
. correlate con confit
```

```
(obs=50)
```

	con	confit
con	1.0000	
confit	0.9247	1.0000

```
. scalar r2p=r(rho)^2
```

```
. display r2p
```

```
.8550451
```



**Primer 2:** Pri ocenjevanju modela povpraševanja po brezalkoholnih pijačah so na podlagi podatkov za 100 štiričlanskih gospodinjstev ugotovili, da je v modelu prisotna heteroskedastičnost slučajne spremenljivke  $u$ . Zato so ocenili še naslednji model:

$$\widehat{\left(\frac{Q}{RD}\right)} = 1,8 + 2,1 \frac{1}{RD} - 0,5 \frac{P}{RD}; \quad R^2 = 0,955.$$

Pri tem  $Q$  predstavlja količino brezalkoholnih pijač (v litrih),  $RD$  realni razpoložljivi dohodek štiričlanskih gospodinjstev (v 1.000 denarnih enotah) in  $P$  relativne cene brezalkoholnih pijač.

- a) Katero predpostavko o obliki heteroskedastičnosti slučajne spremenljivke  $u$  so upoštevali pri transformaciji izhodiščnega modela? Pokažite, da uporabljena transformacija modela zagotavlja NENALICE regresijskih koeficientov.
  - b) Katera metoda v konkretnem primeru to zagotavlja? Zapišite izhodiščni model povpraševanja po brezalkoholnih pijačah z NENALICE regresijskimi koeficienti in pojasnite vrednost ocenjenega regresijskega koeficienta  $b_2$ .
- 

**Primer 3:** Na podlagi podatkov za 500 podjetij so raziskovalci proučevali odvisnost vrednosti prodaje ( $PROD$ ) od izdatkov za oglaševanje ( $OGL$ ) in izdatkov za raziskave in razvoj ( $RR$ ). Da bi preverili veljavnost ene od temeljnih predpostavk klasičnega regresijskega modela, so ocenili tudi naslednjo regresijsko enačbo:

$$\widehat{e} = 1,70 + 2,00OGL^{0,5} \quad R^2 = 0,892$$

$t_j: (0,5) (4,8)$

- a) Ugotovite, katero predpostavko so preverjali in kateri test so pri tem uporabili.
  - b) Ugotovite, kakšen sklep so sprejeli in ga utemeljite.
  - c) Kako se imenuje metoda ocenjevanja regresijskega modela v primeru, ko je predpostavka iz točke a) kršena? Zapišite proučevani regresijski model v obliki, ki zagotavlja, da bo ta predpostavka ponovno izpolnjena in to tudi dokažite.
-