Lecture 6: Heteroskedasticity

Jan Zouhar

Introductory Econometrics

Summary of MLR Assumptions

MLR.1 (linear in parameters)

MLR.2 (random sampling)

- the basic framework (we have to start somewhere)

MLR.3 (no perfect collinearity)

- a technical assumption that allows us to estimate the model

MLR.4 (zero conditional mean of *u*)

- the key one for causal work, cannot be tested statistically, has to be argued from the economic theory

- MLR.1 though MLR.4 already give us *unbiasedness* of OLS
- typically, we want more than this
- we want to know we're using the best estimator – the BLUE one
- for this, we needed the assumption of constant error variance:

MLR.5 (homoskedasticity)

Summary of MLR Assumptions

- with MLR.1 through MLR.5, we know OLS is BLUE
- we also know the variance and the asymptotic sampling distribution of the OLS estimator (we use this to compute standard errors and carry out t-tests and F-tests)
- the important questions for this lecture:
  - what happens if MLR.5 is violated in my equation?
  - can I test MLR.5 statistically?
- then we had another one:

MLR.6 (normality)

  - this completes CLRM
  - we needed MLR.6 for small-sample properties of OLS
  - this is a technical thing, we won't be bothered with it anymore

How do I Find Out That MLR.5 Is Violated?

- there's a bunch of statistical tests to find out; all of them have their limitations
- we won't cover the theory behind them here (see Wooldridge, Chapter 8 for a thorough discussion)
- for now, just note that they all use the information about u that is contained in the *residuals* from OLS regression
- therefore, you always have to run the OLS regression first
- after you do so, *Gretl* offers you some of the most widely-used tests in Tests → Heteroskedasticity
  - in any of the tests, just look at the final p-value
  - the hypotheses are always like this:

$$\mathrm{H}_0\text{: } \textit{homoskedasticity}$$

$$\mathrm{H}_1\text{: } \textit{heteroskedasticity}$$

  - therefore, p-values less than 0.05 indicate a problem with heteroskedasticity

How do I Find Out That MLR.5 Is Violated?

$$\hat{u}^2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$
$$+ \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_3^2$$
$$+ \beta_7 x_1 x_2 + \beta_8 x_2 x_3 + \beta_9 x_1 x_3 + error$$

Gretl: *White's test (squares only)*

Breusch & Pagan (1979), Koenker (1981)

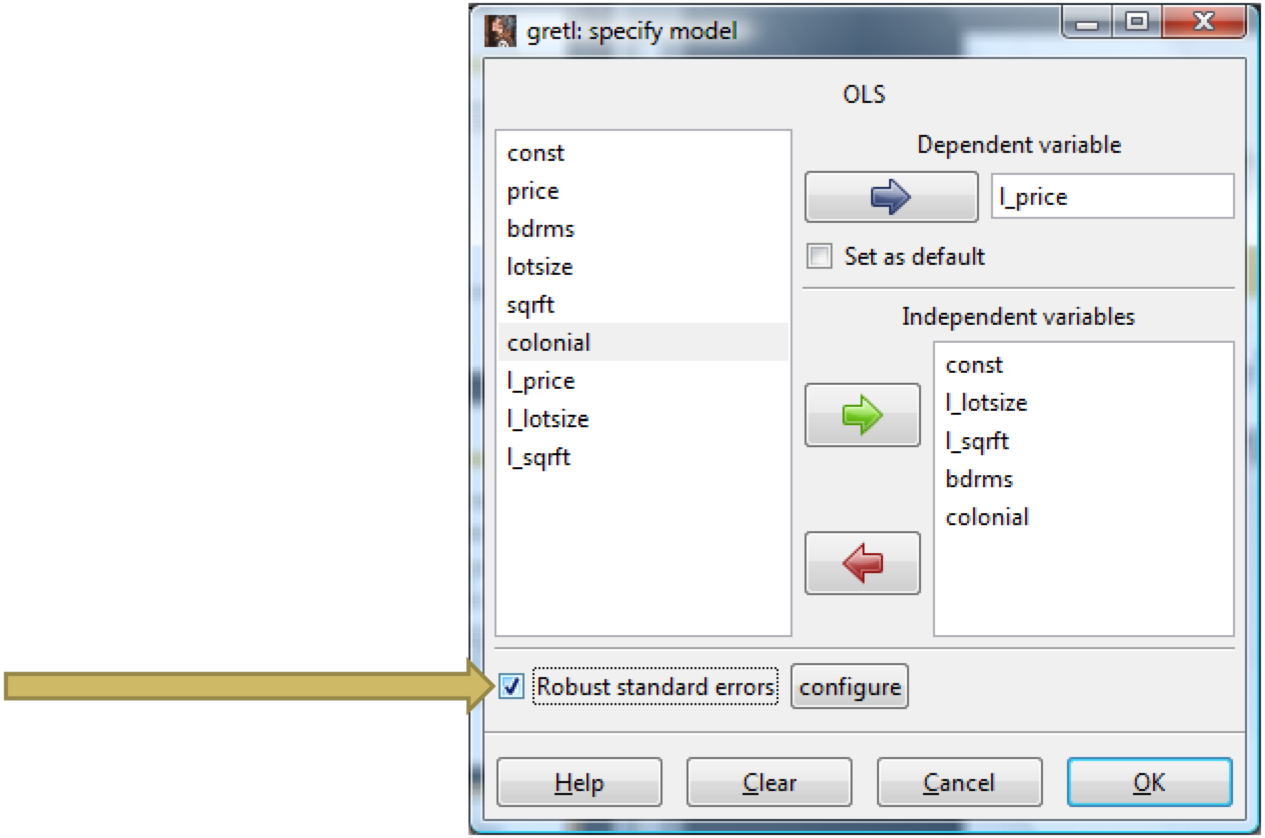
White (1980)

What Should I Do If MLR.5 Is Violated?

- basically, there are two different approaches
  1. try and come up with a more sophisticated method than OLS (and, hopefully, a BLUE one)
     - one such method is the **generalized least squares** estimator (GLS), see Wooldridge, Chapter 8
  2. use OLS to estimate the model, but calculate the standard errors (and the resulting t-ratios and F-statistics) in a different way
     - the idea here is that even without MLR.5, OLS has many favorable properties (*unbiasedness* and some others)
     - the only thing that doesn't really work is the estimate of σ (with heteroskedasticity, there is no "universal" σ in the first place)
     - we needed this for standard errors and p-values, so we'll have to calculate these differently
- we won't cover the theory here (see Wooldridge, Chapter 8 for a thorough discussion)
- fortunately, all of this can be done in *Gretl* very easily

Heteroskedasticity-Robust Inference with OLS

- I'll start with the second approach
- I estimate the equation using OLS (**Model → Ordinary least squares**), but use the **Robust standard errors** option:



Heteroskedasticity-Robust Inference with OLS (cont'd)

- the only thing that differs is the last three columns in the table and the overall F-test, these were calculated differently; the rest is the same

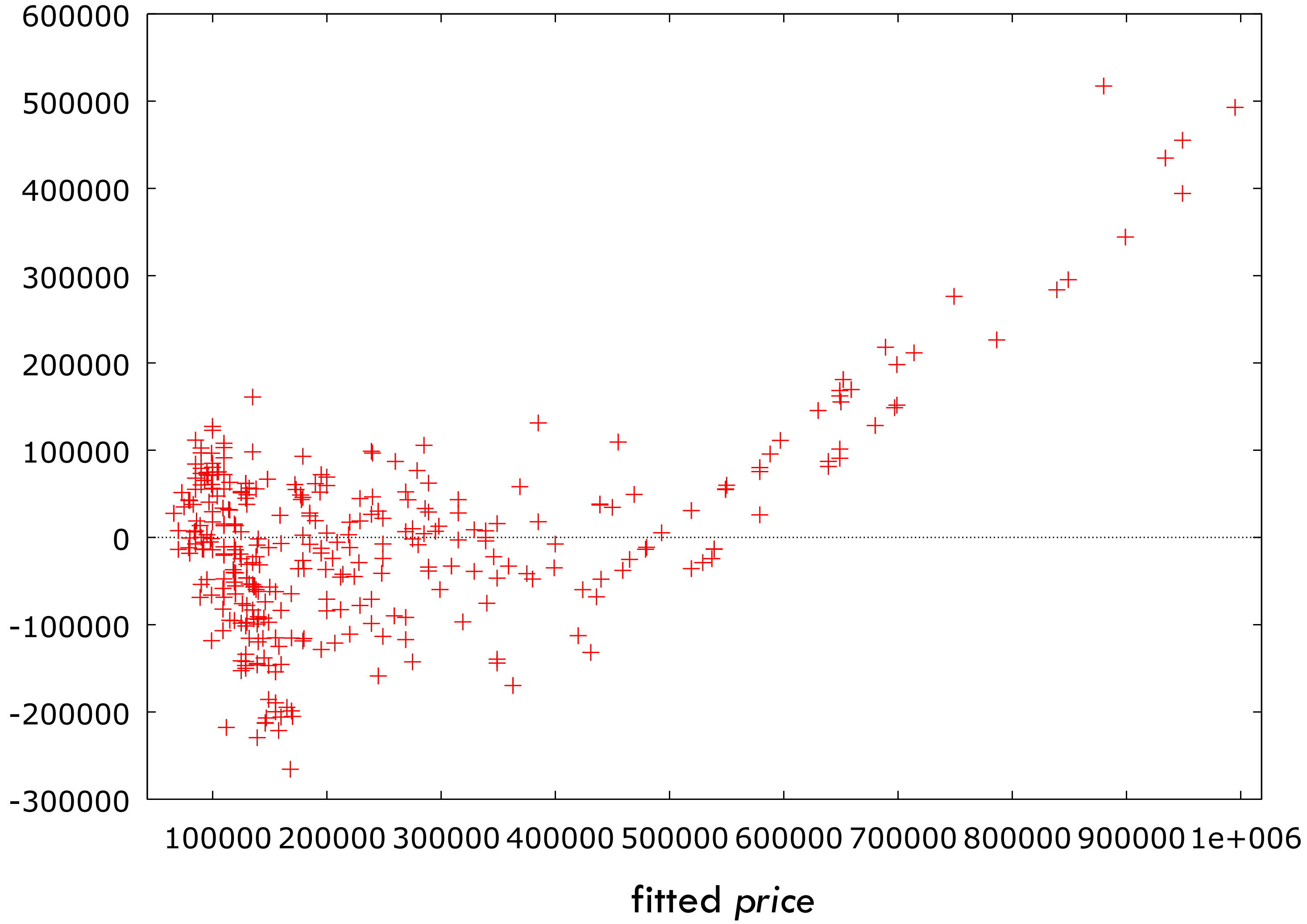
```
gretl: model 1
File  Edit  Tests  Save  Graphs  Analysis  LaTeX

Model 1: OLS, using observations 1-88
Dependent variable: l_price
Heteroskedasticity-robust standard errors, variant HC1

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const      -1.34959      0.811580     -1.663    0.1001
  l_lotsize   0.167819     0.0440357     3.811    0.0003   ***
  l_sqrft     0.707193     0.109045      6.485    6.02e-09 ***
  bdrms       0.0268305    0.0327181     0.8201   0.4145
  colonial    0.0537962    0.0489041     1.100    0.2745

Mean dependent var   5.633180   S.D. dependent var   0.303573
Sum squared resid    2.813624   S.E. of regression   0.184117
R-squared            0.649069   Adjusted R-squared   0.632157
F(4, 83)             34.49699   P-value(F)           6.03e-17
Log-likelihood       26.61940   Akaike criterion    -43.23879
Schwarz criterion   -30.85211   Hannan-Quinn        -38.24851

Log-likelihood for price = -469.1

Excluding the constant, p-value was highest for variable 2 (bdrms)
```

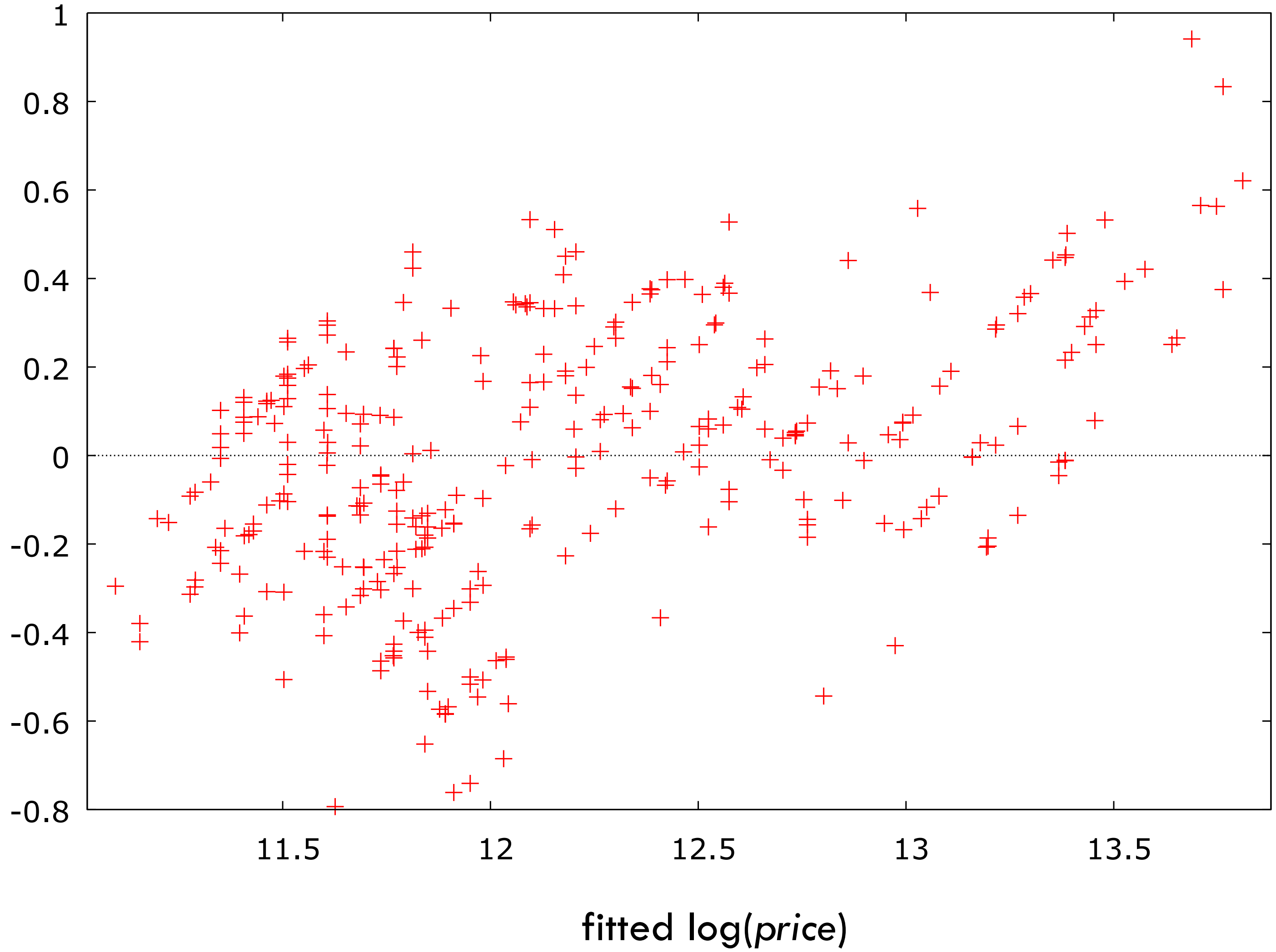
Dependent variable: l_price

	Ordinary SE	*HC1 SE*
const	12.6**	12.6**
	(0.0428)	**(0.0409)**
km1000	-0.00148**	-0.00148**
	(0.000264)	**(0.000272)**
age	-0.110**	-0.110**
	(0.00695)	**(0.00679)**
combi	0.0899**	0.0899**
	(0.0235)	**(0.0278)**
diesel	0.165**	0.165**
	(0.0241)	**(0.0236)**
LPG	0.0521	0.0521
	(0.0610)	**(0.0809)**
octavia	0.564**	0.564**
	(0.0250)	**(0.0206)**
superb	1.07**	1.07**
	(0.0510)	**(0.0480)**

```
Dependent variable: price
...

Breusch-Pagan test for heteroskedasticity (robust variant) -
  Null hypothesis: heteroskedasticity not present
  Test statistic: LM = 44.4887
  with p-value = P(Chi-square(5) > 44.4887) = 1.84309e-008

White's test for heteroskedasticity -
  Null hypothesis: heteroskedasticity not present
  Test statistic: LM = 65.9639
  with p-value = P(Chi-square(16) > 65.9639) = 5.02484e-008
```

```
Dependent variable: l_price
...

Breusch-Pagan test for heteroskedasticity (robust variant) -
  Null hypothesis: heteroskedasticity not present
  Test statistic: LM = 15.5747
  with p-value = P(Chi-square(5) > 15.5747) = 0.00816946

White's test for heteroskedasticity -
  Null hypothesis: heteroskedasticity not present
  Test statistic: LM = 29.9919
  with p-value = P(Chi-square(16) > 29.9919) = 0.0180442
```

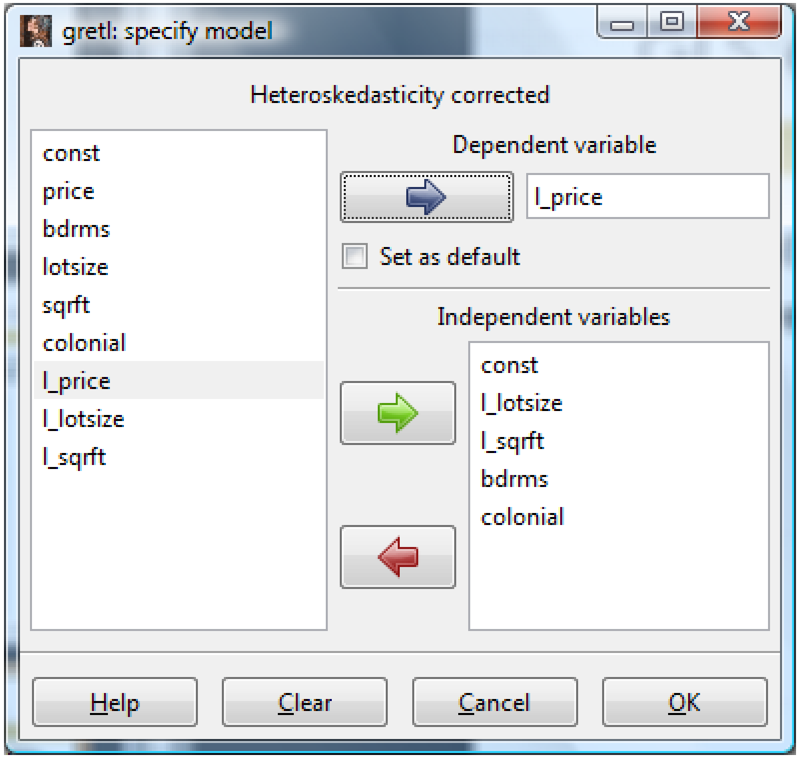

600000
500000
400000
300000
200000
100000
0
-100000
-200000
-300000
100000 200000 300000 400000 500000 600000 700000 800000 900000 1e+006
fitted *price*



1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
11.5
12
12.5
13
13.5
fitted log(*price*)

GLS estimation

- in order to run GLS estimation, use **Other linear models → Heteroskedasticity corrected**)
- the window looks just as with OLS:



GLS estimation (cont'd)

- the *Gretl* output looks a bit different now; the results under the table (including the R-squared) have a slightly different interpretation

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: Heteroskedasticity-corrected, using observations 1-88
Dependent variable: l_price

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       -0.666843     0.631073     -1.057    0.2937
  l_lotsize    0.231019     0.0547308     4.221    6.18e-05 ***
  l_sqrft      0.541765     0.0948327     5.713    1.69e-07 ***
  bdrms        0.0354503    0.0260389     1.361    0.1771
  colonial     0.0353411    0.0477197     0.7406   0.4610

Statistics based on the weighted data:

Sum squared resid    429.6738   S.E. of regression   2.275257
R-squared            0.600115   Adjusted R-squared   0.580844
F(4, 83)             31.13994   P-value(F)           7.83e-16
Log-likelihood      -194.6369   Akaike criterion     399.2738
Schwarz criterion    411.6605   Hannan-Quinn         404.2641

Statistics based on the original data:

Mean dependent var   5.633180   S.D. dependent var   0.303573
Sum squared resid    2.998114   S.E. of regression   0.190058

Excluding the constant, p-value was highest for variable 5 (colonial)
```

Lecture 6:
Heteroskedasticity

Jan Zouhar

Introductory Econometrics