

Эконометрика, 2019-2020, 3 модуль
Семинары 1-2
14.01.2020, 21.01.2020
Группы Э_Б2017_Э_3
Семинарист О.А.Демидова

Гетероскедастичность, взвешенный МНК

Учебник Демидова и Малахов, задачи 10.1-10.5, задание 10.1-10.4

1) Задача 10.3. (Учебник Демидова, Малахов)

Предположим, что для модели парной регрессии $\sigma_{ui}^2 = \sigma^2 X_i^4$. Как избавиться от проблемы гетероскедастичности ошибок?

2) Задача 10.4. (Gujarati, 4 изд., с.429, задача 11.6)

По данным с 1946 г. по 1975 г. Е.А. Hanushek и J.E. Jackson (Statistical methods for social scientists, Academic, NY, 1977, с.160) оценили коэффициенты уравнений регрессий (в скобках указаны оценки стандартных отклонений):

$$\hat{C}_t = 26.19 + 0.6248 GNP_t - 0.4398 D_t$$

(2.73) (0.006) (0.0736)

$$\left(\frac{\hat{C}}{GNP} \right)_t = 25.92 \frac{1}{GNP_t} + 0.6246 - 0.4315 \left(\frac{D}{GNP} \right)_t$$

(2.22) (0.00597) (0.0597)

где С – агрегированные частные потребительские расходы,
GNP (Gross National Product) – валовой национальный продукт,
D – национальные расходы на оборону.

С какой целью оценили второе уравнение? Какое при этом было сделано предположение о дисперсии ошибок?

Можно ли сравнивать R^2 в двух регрессиях? Ответ обоснуйте.
Дайте интерпретацию полученным результатам.

2) Учебник Демешев и Борзых

- 8.2** В модели $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ присутствует гетероскедастичность вида $\text{Var}(\varepsilon_i) = \sigma^2 x_i^2$. Как надо преобразовать исходные регрессоры и зависимую переменную, чтобы устранить гетероскедастичность?
- 8.3** В модели $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ присутствует гетероскедастичность вида $\text{Var}(\varepsilon_i) = \lambda |x_i|$. Как надо преобразовать исходные регрессоры и зависимую переменную, чтобы устранить гетероскедастичность?
- 8.4** Известно, что после деления каждого уравнения регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ на x_i^2 гетероскедастичность ошибок была устранена. Какой вид имела дисперсия ошибок, $\text{Var}(\varepsilon_i)$?
- 8.5** Известно, что после деления каждого уравнения регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ на $\sqrt{x_i}$ гетероскедастичность ошибок была устранена. Какой вид имела дисперсия ошибок, $\text{Var}(\varepsilon_i)$?

8.7 По наблюдениям $x = (1, 2, 3)'$, $y = (2, -1, 3)'$ оценивается модель $y = \beta_1 + \beta_2 x + \varepsilon$. Ошибки ε гетероскедастичны и известно, что $\text{Var}(\varepsilon_i) = \sigma^2 \cdot x_i^2$.

1. Найдите оценки $\hat{\beta}_{ols}$ с помощью МНК и их ковариационную матрицу.
2. Найдите оценки $\hat{\beta}_{gls}$ с помощью обобщенного МНК и их ковариационную матрицу.

8.8 Для линейной регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ была выполнена сортировка наблюдений по возрастанию переменной x . Исходная модель оценивалась по разным частям выборки:

Выборка	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	RSS
$i = 1, \dots, 30$	1.21	1.89	2.74	48.69
$i = 1, \dots, 11$	1.39	2.27	2.36	10.28
$i = 12, \dots, 19$	0.75	2.23	3.19	5.31
$i = 20, \dots, 30$	1.56	1.06	2.29	14.51

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

8.12 Рассмотрим линейную регрессию $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ по 50 наблюдениям. При оценивании с помощью МНК были получены результаты: $\hat{\beta}_1 = 1.21$, $\hat{\beta}_2 = 1.11$, $\hat{\beta}_3 = 3.15$, $R^2 = 0.72$.

Оценена также вспомогательная регрессия: $\hat{\varepsilon}_i^2 = \delta_1 + \delta_2 x_i + \delta_3 z_i + \delta_4 x_i^2 + \delta_5 z_i^2 + \delta_6 x_i z_i + u_i$. Результаты оценивания следующие: $\hat{\delta}_1 = 1.50$, $\hat{\delta}_2 = -2.18$, $\hat{\delta}_3 = 0.23$, $\hat{\delta}_4 = 1.87$, $\hat{\delta}_5 = -0.56$, $\hat{\delta}_6 = -0.09$, $R_{aux}^2 = 0.36$

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

- 8.15** Рассматривается модель $y_t = \beta_1 + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\text{Var}(\varepsilon_t) = t$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.
- 8.16** Рассматривается модель $y_t = \beta_1 + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\text{Var}(\varepsilon_t) = t^2$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.
- 8.17** Рассматривается модель $y_t = \beta_1 x_t + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\text{Var}(\varepsilon_t) = t$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.
- 8.18** Рассматривается модель $y_t = \beta_1 x_t + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\text{Var}(\varepsilon_t) = t^2$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.
- 8.23** Рассмотрим модель регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$, в которой ошибки ε_i независимы и имеют нормальное распределение $N(0, \sigma^2)$. Для $n = 200$ наблюдений найдите
1. вероятность того, что статистика Уайта окажется больше 10;
 2. ожидаемое значение статистики Уайта;
 3. дисперсию статистики Уайта.
- 8.24** Найдите число коэффициентов во вспомогательной регрессии, необходимой для выполнения теста Уайта, если число коэффициентов в исходной регрессии равно k , включая свободный член.

8.25 По 35 наблюдениям сотрудники НИИ оценили уравнение регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ и рассчитали остатки ε_i . После того они приступили к диагностике возможных недостатков модели, обнаружили гетероскедастичность и решили её побороть.

1. Самый младший научный сотрудник выдвинул предположение, что стандартное отклонение случайной составляющей может быть выражено так: $\sigma_{\varepsilon,i} = ax_i$, где a — неизвестный коэффициент. Каким образом нужно преобразовать исходное уравнение регрессии, чтобы избавиться от гетероскедастичности?
2. Профессор решил перепроверить результаты и оценил регрессию:

$$\hat{\varepsilon}_i^2 = -0.3 + 0.08x_i - 0.01x_i^2, R^2 = 0.15.$$

Свидетельствует ли полученный профессором результат о наличии гетероскедастичности?

Упражнение 10.1. (Учебник Демидова, Малахов)

В файле `clothing.dta` содержатся данные о продажах одежды в 400 немецких магазинах одежды, позаимствованные с сайта издательства книги Марно Вербика “A guide to modern econometrics”, <http://wileyurope.com/go/verbeek2ed>.

Переменные:

`tsales` – среднегодовые продажи в гильденах,
`sales` – продажи в расчете на квадратный метр,
`margin` – маржинальная валовая прибыль,
`pown` – количество собственников (менеджеров),
`nfull` – количество полностью занятых,
`part` – количество частично занятых,
`paux` – количество временно работающих,
`hoursw` – общее число отработанных часов,
`hourspw` – количество отработанных часов в расчете на одного работающего,
`inv1` – капиталовложения в помещения,
`inv2` – капиталовложения в автоматизацию,
`ssize` – размер магазина в m^2 ,
`start` – год открытия магазина.

Оцените коэффициенты уравнения регрессии

$$sales = \beta_0 + \beta_1 hoursw + \beta_2 ssize + u.$$

- 1) Проведите тесты Уайта, Бройша – Пагана, Голдфелда-Квандта, Глейзера на выявление гетероскедастичности возмущений.
- 2) Если гетероскедастичность будет выявлена, то проведите коррекцию.

Методические рекомендации

- 1) Открыв файл `clothing.dta` в статистическом пакете STATA, оцените необходимую регрессию с помощью команды:

```
reg sales hoursw ssize
```

2) Для проведения теста Уайта после оценки регрессии необходимо набрать в командном окне:

```
estat imtest, white
```

3) Для проведения теста Бройша – Пагана необходимо набрать в командном окне

```
estat hettest, rhs mtest
```

4) Для проведения теста Глейзера необходимо сохранить остатки регрессии с помощью команды

```
predict res, resid
```

Создать новые переменные можно $|e|$, \sqrt{hoursw} , $\frac{1}{hoursw}$ с помощью команд:

```
gen modres = abs(res)
gen sqrthoursw = sqrt(hoursw)
gen invhoursw = 1/hoursw
```

Оценить коэффициенты трех регрессий:

$$|e| = \alpha + \beta hoursw + u, \quad |e| = \alpha + \beta \sqrt{hoursw} + u, \quad |e| = \alpha + \frac{\beta}{hoursw} + u$$

Можно с помощью команд

```
reg modres hoursw
reg modres sqrthoursw
reg modres invhoursw
```

5) Для проведения теста Голдфелда – Квандта необходимо упорядочить наблюдения по переменной hoursw с помощью команды

```
sort hoursw
```

Для того, чтобы оценить параметры уравнения регрессии по первой и последней трети (приблизительно) наблюдений и сохранить соответствующие RSS, наберите в командном окне

```
reg sales hoursw ssize in 1/133
scalar rss1 = e(rss)
reg sales hoursw ssize in 268/400
scalar rss2 = e(rss)
```

Используя RSS в последних двух оцененных регрессиях, рассчитать тестовую статистику

$$\text{можно по формуле } F = \frac{RSS_2 / (n_2 - k - 1)}{RSS_1 / (n_1 - k - 1)} = \frac{RSS_2}{RSS_1} =$$

С помощью команды:

```
scalar F = rss2/rss1
display F
```

а соответствующее критическое значение F-статистики с помощью команды:

```
display invFtail(130,130,0.05)
```

б) Применим два способа для решения проблемы гетероскедастичности: Первым способом является преобразование исходных переменных.

Для этого создадим новые переменные

$$sales_new = \frac{sales}{hoursw}, \quad cons_new = \frac{1}{hoursw}, \quad size_new = \frac{size}{hoursw}$$

с помощью команд

```
gen sales_new = sales/hoursw
gen cons_new = 1/hoursw
gen size_new = size/hoursw
```

Оценить коэффициенты новой регрессии можно с помощью команды

```
reg sales_new cons_new size_new
```

Нужно быть внимательным при интерпретации оценок коэффициентов новой регрессии!

Вторым способом решения проблемы гетероскедастичности возмущений является использование оценок Уайта (White, 1980) для дисперсий коэффициентов.

Их можно получить с помощью команды `reg sales hoursw ssize, robust`

Проверка нормальности распределения остатков регрессии

По данным файла `clothing.dta`, содержащем данные о продажах одежды в 400 голландских магазинах мужской одежды

1) Оцените коэффициенты уравнения регрессии

$$sales = \beta_0 + \beta_1 hoursw + \beta_2 size + u.$$

2) Постройте гистограмму распределения остатков регрессии и `qqplot`.

С помощью тестов Колмогорова-Смирнова, Харке-Бера, Шапиро-Уилка проверьте гипотезу о нормальности распределения случайных ошибок.

3) Рассчитайте Стьюдентизированные остатки регрессии, постройте их график и с его помощью определите возможные наблюдения, которые являются выбросами. Если такие наблюдения будут выявлены, то оцените регрессию

а) без наблюдений, являющихся выбросами,

б) R-регрессию Хубера.

4) Сравните результаты оценки регрессий, оцененных с помощью МНК (с выбросами и без), с оценками Уайта для стандартных ошибок, R-регрессии Хуберта.

Методические рекомендации

1) Открыв файл `clothing.dta` в статистическом пакете STATA, оцените необходимую регрессию с помощью команды:

```
reg sales hoursw ssize
```

2) Сохраните остатки регрессии с помощью команды

```
predict res, resid
```

3) Постройте их гистограмму с помощью команды `hist res`

`Qq plot` с помощью команды

```
qnorm res
```

4) Проведите тест Колмогорова-Смирнова с помощью команды

```
sum res
```

```
ksmirnov res = normal((res-r(mean))/r(sd))
```

5) Проведите тест Харке-Бера с помощью команды `sktest res, noadjust`

6) Проведите тест Шапиро-Уилка с помощью команды `swilk res`

7) Постройте график стьюдентизированных остатков регрессии с помощью команды

```
predict residst, rstudent
```

```
twoway(scatter residst number, sort)
```

```
list number if abs(residst) > 2
```

Оценить регрессию без наблюдений-выбросов можно с помощью команды

```
reg sales hoursw ssize if abs(residst) < 2
```

8) Оценить R-регрессию Хубера можно с помощью команды

```
rreg sales hoursw ssize
```

9) Сохранить результаты оценки регрессии можно с помощью команды

```
est store reg1 (и т.д.)
```

Для сравнения результатов удобно сформировать общую таблицу с помощью команды

```
est tab reg1 (и т.д.), star (0.1 0.05 0.01) b(%7.3f)
```