ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Эконометрика

Лекция 9
Гетероскедастичность

Вакуленко Е.С.

д.э.н., доцент департамента прикладной экономики

evakulenko@hse.ru

Москва, 2022

План

- Гетероскедастичность:
  - Признаки;
  - Тесты;
  - Коррекция.

Предположения о случайном члене (условия Гаусса-Маркова)

Для всех наблюдений $i = 1, \dots, n$

1. $E(\varepsilon_i) = 0$ – несистематичная ошибка
2. $Var(\varepsilon_i) = \sigma^2$ – гомоскедастичность
3. $Cov(\varepsilon_i, \varepsilon_j) = 0$, для всех $i \neq j$
4. Случайный член должен быть распределен независимо от объясняющих переменных

Предположения о случайном члене (условия Гаусса-Маркова)

Для всех наблюдений $i = 1, \dots, n$

1. $E(\varepsilon_i) = 0$ – несистематичная ошибка
2. $Var(\varepsilon_i) = \sigma^2$ – гомоскедастичность
3. $Cov(\varepsilon_i, \varepsilon_j) = 0$, для всех $i \neq j$
4. Случайный член должен быть распределен независимо от объясняющих переменных

Что произойдет, если будет нарушена 2-ая предпосылка?

Гетероскедастичность

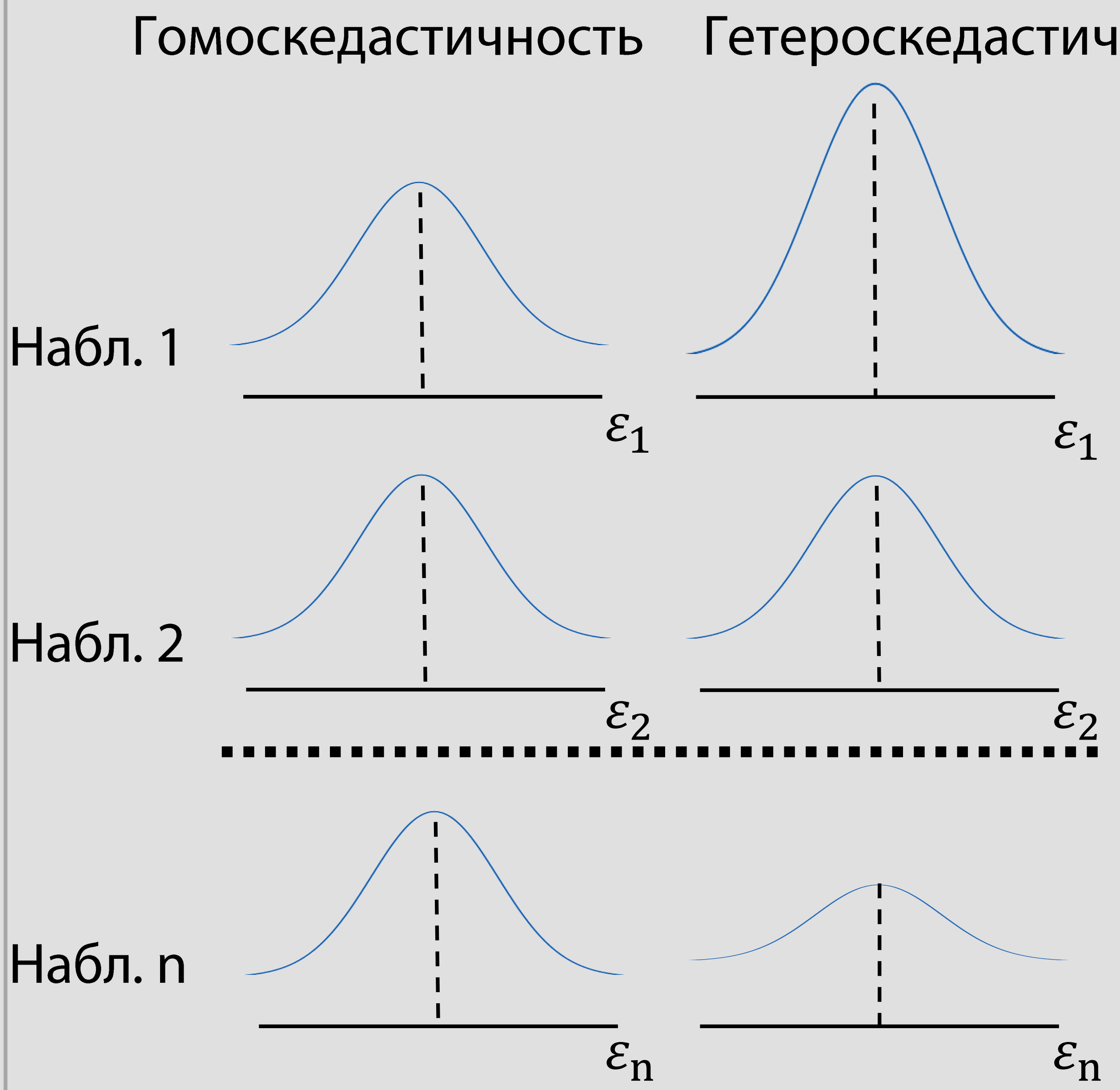
Гетероскедастичность – явление, характерное для перекрестных выборок и заключающееся в различии дисперсий ошибок регрессии, т.е. для регрессии

$$Y_i = \alpha + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i, i = 1, \ldots, n$$

не выполнено одно из условий теоремы Гаусса-Маркова:

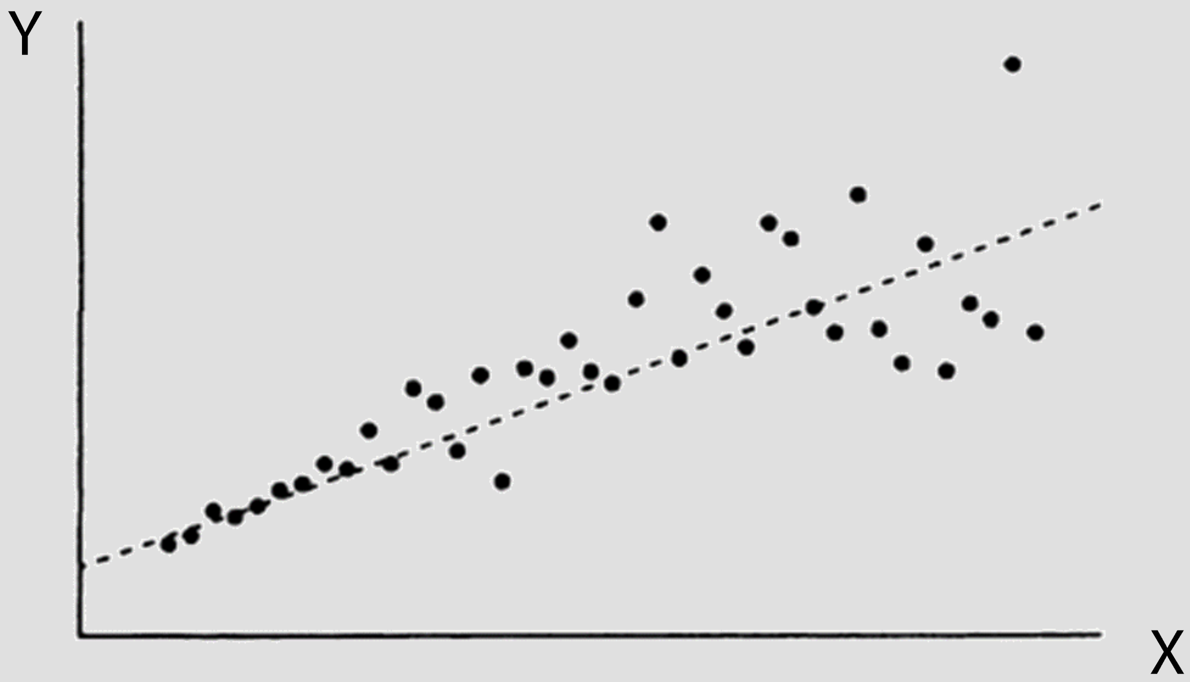
$$\sigma^2_{\varepsilon_i} = \sigma^2 \; \forall i = 1, \ldots, n$$

Оно часто возникает, если анализируемые объекты неоднородны.



В случае гетероскедастич-ности дисперсии случайных составляющих различаются. На графиках видно, что в таком случае распределения случайных ошибок имеют разную форму.

Графическое определение гетероскедастичности



Y – зависимая переменная, X – регрессор.
На графике видно, что с ростом значений X разброс наблюдений по Y возрастает. Это свидетельствует о наличии гетероскедастичности.

Причины гетероскедастичности

Значительное различие переменных для разных наблюдений.

Неоднородность данных.

Последствия гетероскедастичности

Оценки стандартных ошибок коэффициентов регрессии **смещены**;
Оценки МНК коэффициентов регрессии **неэффективны**;
t-статистики коэффициентов регрессии **неадекватны**.

⇩

Неправильно делаем вывод о значимости коэффициентов

Обнаружение гетероскедастичности

Графически, изучение графика остатков

Обнаружение гетероскедастичности

Графически, изучение графика остатков

Тесты

$$\boldsymbol{H_0 : \sigma_i^2 = \sigma^2 \ \forall i} \quad \text{(гомоскедастичность)}$$
$$\boldsymbol{H_1 : \exists i, j : \sigma_i^2 \neq \sigma_j^2} \quad \text{(гетероскедастичность)}$$

Тест Уайта
Тест Глейзера
Тест Голдфельда – Квандта
Тест Бройша - Пагана
И другие

График «остатки-прогнозы»

График «остатки-прогнозы» позволяет выявить и гетероскедастичность, и ошибку функциональной формы. Для построения графика нужно выполнить следующие шаги.

1. Оценить проверяемую регрессионную модель

$$Y_i = \alpha + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i$$

методом наименьших квадратов.

2. Рассчитать прогнозные значения объясняемой переменной   для всех наблюдений.

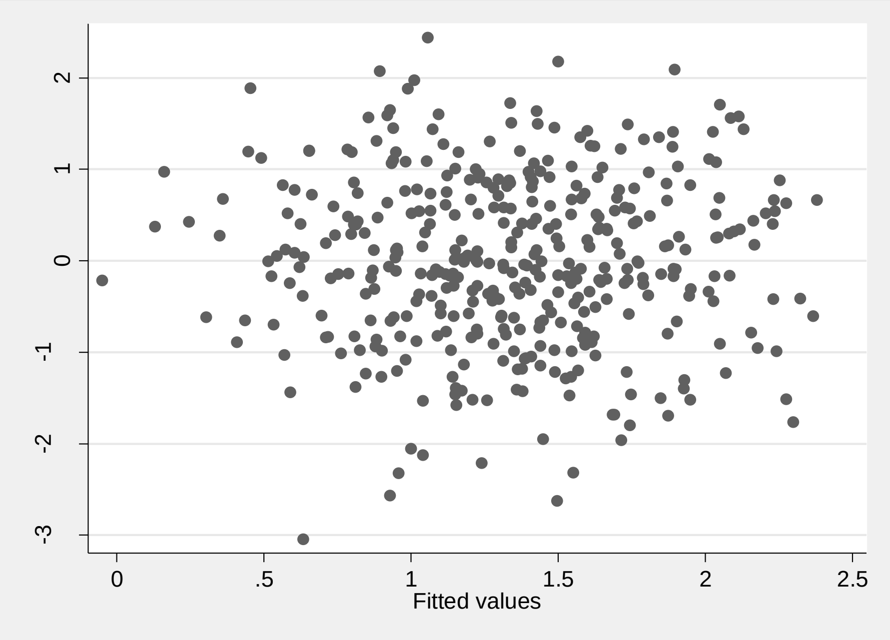
$$\hat{Y}_i = \alpha + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_k x_{ik}$$

3. Рассчитать остатки $e_i = Y_i - \hat{Y}_i$ для всех наблюдений.
4. Построить график рассеяния в осях e и $\hat{Y}$. Обычно значения остатков откладываются по вертикальной оси, а прогнозы — по горизонтальной.

График «остатки-прогнозы»

- При **отсутствии гетероскедастичности и верной функциональной форме** не проявляется какая-либо зависимость остатков от прогнозов, он выглядит как бессистемное облако точек.

График «остатки-прогнозы». Пример



Residuals – остатки, Fitted values – прогноз по модели.

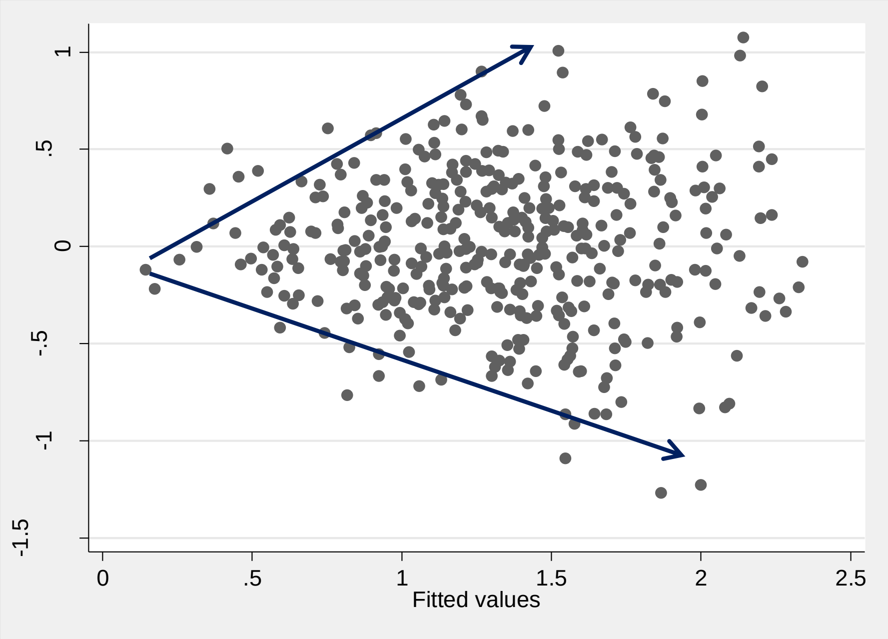
График «остатки-прогнозы» в случае гомоскедастичности и верной функциональной формы.

Источник: Борзых, Вакуленко, Фурманов (2021)

График «остатки-прогнозы»

- При **гетероскедастичности** разброс остатков зависит от прогнозных значений. При этом средний уровень остатков во всех частях графика примерно равен нулю.

График «остатки-прогнозы». Пример



Residuals – остатки, Fitted values – прогноз по модели.

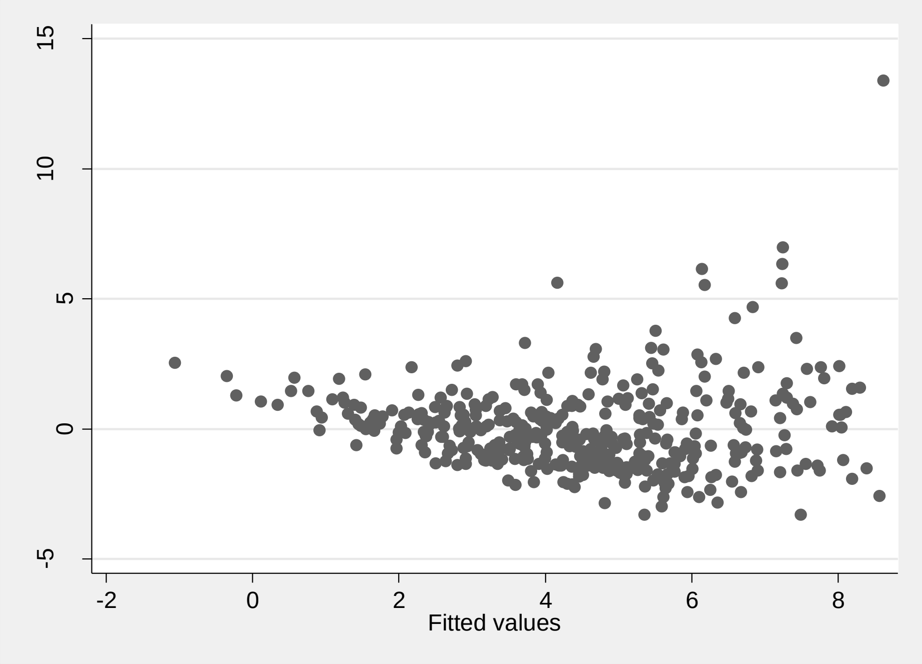
График «остатки-прогнозы» в случае гетероскедастичности и верной функциональной формы.

Источник: Борзых, Вакуленко, Фурманов (2021)

График «остатки-прогнозы»

- В случае **неверно выбранной функциональной формы** средний уровень остатков зависит от прогнозных значений. При этом разброс остатков также может зависеть от прогнозов, хотя это не обязательно.

График «остатки-прогнозы». Пример



Residuals – остатки, Fitted values – прогноз по модели.

График «остатки-прогнозы» в случае ошибочно выбранной функциональной формы.

Источник: Борзых, Вакуленко, Фурманов (2021)

Тесты на гетероскедастичность

Тест Уайта

- Оцениваем коэффициенты основной регрессии $Y = \alpha+\beta_1 X_1+\ldots+\beta_k X_k+\varepsilon$.
- Сохраняем остатки регрессии e_i, $i=1,\ldots,n$.
- Оцениваем регрессию квадратов остатков на все регрессоры, их квадраты, попарные произведения и константу:

$$e^2 = \alpha_1 + \sum_{l=2}^{k} \alpha_l X_l + \sum_{l=2}^{k} \beta_{l2} X_l^2 + \sum_{\substack{l,j=2 \\ l<j}}^{k} \gamma_{lj} X_l X_j + u$$

- В последней оцененной регрессии находим коэффициент множественной детерминации $\mathbf{R^2}$
- Вычисляем тестовую статистику по формуле $\mathbf{nR^2}$. При выполнении нулевой гипотезы тестовая статистика имеет распределение χ^2_{m-1}, где m – число оцениваемых в последней регрессии коэффициентов.

Тест Уайта

- Сравниваем полученное значение тестовой статистики с критическим χ^2_{m-1} при заданном уровне значимости α. Если значение тестовой статистики превышает критическое, то нулевая гипотеза о гомоскедастичности отвергается.
- Если гетероскедастичность выявлена, то тест Уайта не дает указания на функциональную форму гетероскедастичности.

Тест Глейзера

Тесты Глейзера выявляют значимость вспомогательных регрессионных зависимостей модуля остатков от функциональных форм каждого регрессора

$$|e| = \alpha + \beta X_j + u, |e| = \alpha + \beta \sqrt{X_j} + u, |e| = \alpha + \beta / X_j + u$$
$$j \in \{1, \ldots, k\}$$

Если коэффициент β значим хотя бы в одной из трех регрессий (значимость коэффициента проверяется как обычно с помощью t – статистики), то имеет место гетероскедастичность.

Тест Бройша-Пагана

- Тест Бройша - Пагана применяется в тех случаях, когда априорно предполагается, что дисперсии остатков зависят от некоторых дополнительных переменных и состоит из следующих шагов:
- Оцениваем коэффициенты основной регрессии

$$\boldsymbol{Y = \alpha + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon}$$

- Сохраняем остатки регрессии e_i, $i=1,\ldots,n$.
- Строим оценку для дальнейшей нормировки квадратов остатков регрессии:

$$\boldsymbol{\hat{\sigma}^2 = \frac{\sum\limits_{i=1}^{n} e_i^2}{n}}$$

Тест Бройша-Пагана

- Оцениваем параметры регрессии
$$\frac{e^2}{\hat{\sigma}^2} = \gamma_0 + \gamma_1 Z_1 + \ldots + \gamma_p Z_p + u$$
и вычисляем для нее **ESS**.
- При выполнении нулевой гипотезы тестовая статистика ESS/2 имеет распределение χ_p^2.
- Сравниваем полученное значение тестовой статистики с критическим χ_p^2 при заданном уровне значимости α. Если значение тестовой статистики превышает критическое, то нулевая гипотеза о гомоскедастичности отвергается.

Пример. Затраты пользователей в мобильном приложении

Данные содержат информацию о 150 пользователях некоторого мобильного приложения:

Expend — затраты пользователя на покупки в мобильном приложении;

Time — среднее время, которое пользователь проводил в приложении, минуты;

Age1 — 1 для пользователей от 18 до 21 года, 0 иначе;

Age2 — 1 для пользователей от 22 до 25 лет, 0 иначе;

Age3 — 1 для пользователей от 26 до 29 лет, 0 иначе;

Age4 — 1 для пользователей от 30 до 34 лет, 0 иначе;

Age5 — 1 для пользователей от 35 лет и старше, 0 иначе;

MPrice — рыночная стоимость используемой модели смартфона.

Источник: Борзых, Вакуленко, Фурманов (2021)

Пример. Затраты пользователей в мобильном приложении

Изучается зависимость затрат в мобильном приложении от прочих характеристик.

- Оцените уравнение:

$$\text{Expend}_i = \alpha + \beta_1 \text{Time}_i + \beta_2 \text{MPrice}_i + \beta_3 Age1_i + \beta_4 Age2_i + \\ + \beta_5 Age3_i + \beta_6 Age4_i + \varepsilon_i$$

- Протестируйте наличие гетероскедастичности.

Источник: Борзых, Вакуленко, Фурманов (2021) / m.alterainvest.ru

Оценка регрессии в Gretl

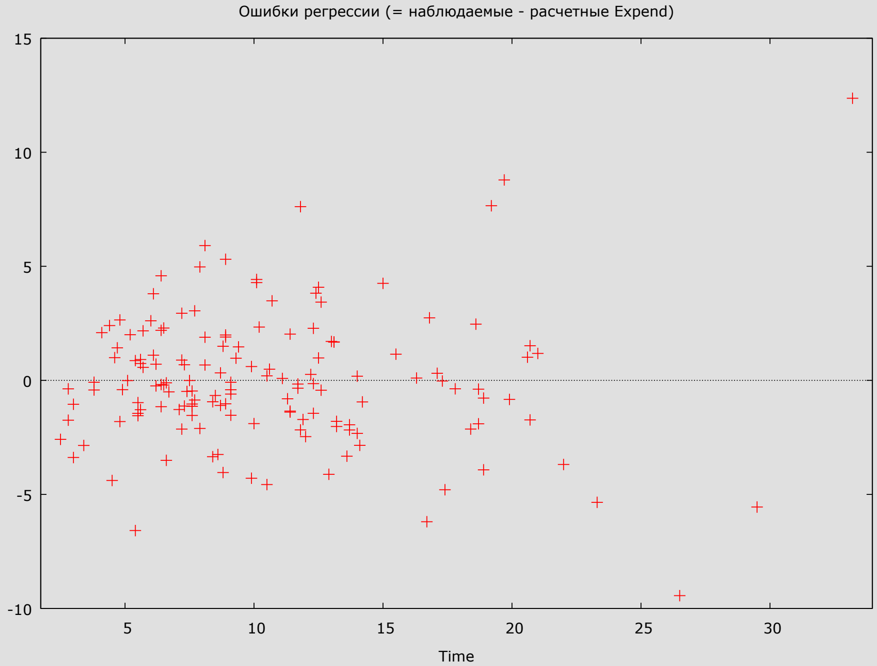
$$\widehat{\text{Expend}}_i = 3.36 + 0.50 \cdot \text{Time}_i + 0.005 \cdot \text{MPrice}_i + \\ +1.27 \cdot Age1_i + 5.86 \cdot Age2_i + \\ +5.38 \cdot Age3_i + 0.27 \cdot Age4_i$$

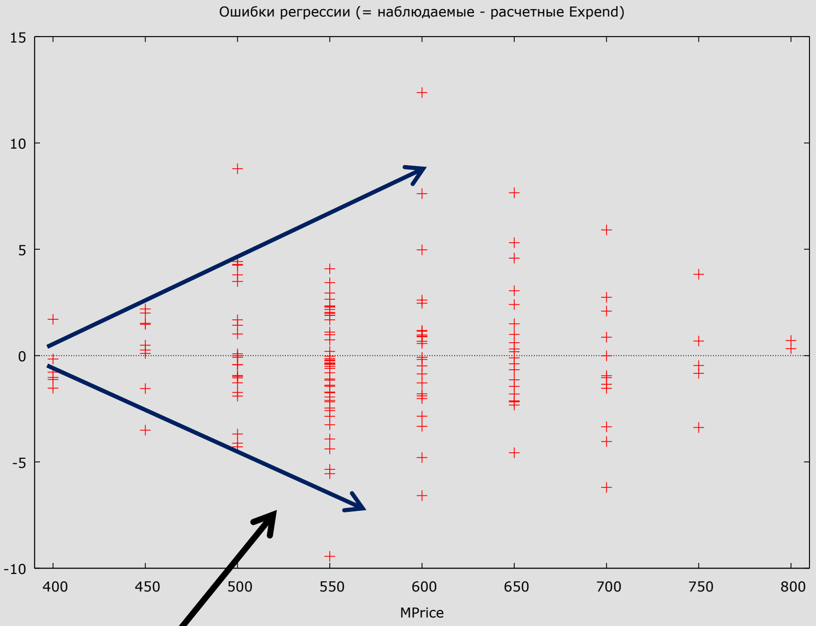
gretl: модель 1

Файл Правка Тесты Сохранить Графики Анализ LaTeX

```
Модель 1: МНК, использованы наблюдения 1-149
Зависимая переменная: Expend

              Коэффициент   Ст. ошибка   t-статистика   P-значение
  -----------------------------------------------------------------
  const       3,36816       1,84825        1,822        0,0705     *
  Time        0,502159      0,0451424     11,12         4,61e-021  ***
  MPrice      0,00454373    0,00299223     1,519        0,1311
  Age1        1,26542       1,46240        0,8653       0,3883
  Age2        5,86393       0,790107       7,422        9,71e-012  ***
  Age3        5,37991       0,797982       6,742        3,64e-010  ***
  Age4        0,273717      0,632526       0,4327       0,6659

Среднее зав. перемен    13,22081   Ст. откл. зав. перемен  4,811603
Сумма кв. остатков      1266,340   Ст. ошибка модели       2,986283
R-квадрат               0,630420   Испр. R-квадрат         0,614804
F(6, 142)               40,36993   Р-значение (F)          2,15e-28
Лог. правдоподобие     -370,8473   Крит. Акаике            755,6947
Крит. Шварца            776,7223   Крит. Хеннана-Куинна    764,2378

Исключая константу, наибольшее p-значение получено для переменной 7 (Age4)
```

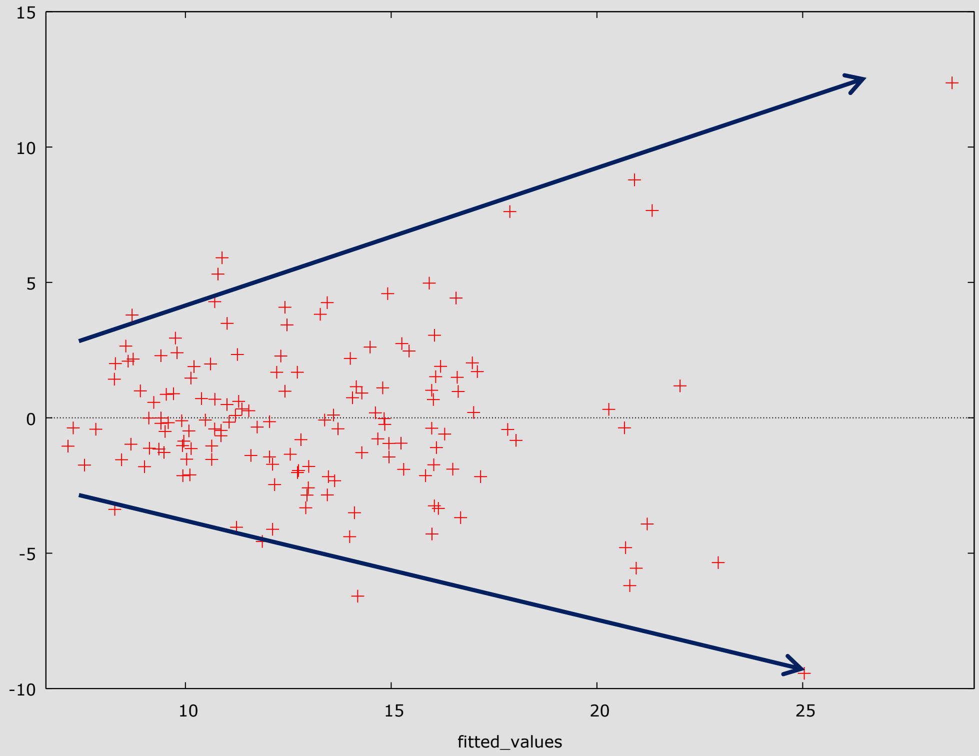
Графики остатков





Есть гетероскедастичность!

График «остатки-прогнозы». Пример



Residuals – остатки, Fitted values – прогноз по модели.

Вывод: есть гетероскедастичность, т.к. растет разброс с ростом прогнозных значений. Средний уровень примерно равен нулю.

Пример. Тест Уайта

gretl: LM тест (гетероскедастичность)

```
Тест Вайта (White) на гетероскедастичность
МНК, использованы наблюдения 1-149
Зависимая переменная: uhat^2

              Коэффициент    Ст. ошибка   t-статистика  P-значение
  ----------------------------------------------------------------
  const        16,1320       46,3287        0,3482      0,7282
  Time         -3,88298       1,79171      -2,167       0,0320     **
  MPrice        0,0204611     0,140559      0,1456      0,8845
  Age1         23,1203       60,3161        0,3833      0,7021
  Age2        -59,9675       33,7890       -1,775       0,0783     *
  Age3         -1,85795      25,9664       -0,07155     0,9431
  Age4         -8,28616      19,0699       -0,4345      0,6646
  sq_Time       0,116911      0,0239315     4,885       2,96e-06   ***
  X2_X3         0,00212073    0,00293021    0,7237      0,4705
  X2_X4        -0,498185      1,40204      -0,3553      0,7229
  X2_X5         2,68503       0,625638      4,292       3,42e-05   ***
  X2_X6         1,21693       0,617389      1,971       0,0508     *
  X2_X7         0,0902781     0,546874      0,1651      0,8691
  sq_MPrice    -3,03191e-05   0,000110421  -0,2746      0,7841
  X3_X4        -0,0335656     0,100113     -0,3353      0,7380
  X3_X5         0,0703062     0,0559757     1,256       0,2113
  X3_X6        -0,00637376    0,0425871    -0,1497      0,8813
  X3_X7         0,0116110     0,0289921     0,4005      0,6894

  Неисправленный R-квадрат = 0,629135

Тестовая статистика: TR^2 = 93,741168,
p-значение = P(Хи-квадрат(17) > 93,741168) = 0,000000
```

$p - value = 0$, следовательно, H_0 отвергается, есть проблема гетероскедастичности

Пример. Тест Бройша-Пагана

gretl: LM тест (гетероскедастичность)

```
Тест Бриша-Пэгана (Breusch-Pagan) на гетероскедастичность
МНК, использованы наблюдения 1-149
Зависимая переменная: Масштабированное uhat^2

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const      -2,99682      1,09602        -2,734       0,0070     ***
  Time        0,180988     0,0267695       6,761       3,29e-010  ***
  MPrice      0,00286214   0,00177440      1,613       0,1090
  Age1       -1,01215      0,867205       -1,167       0,2451
  Age2        1,60452      0,468535        3,425       0,0008     ***
  Age3        1,14598      0,473205        2,422       0,0167     **
  Age4        0,0887373    0,375089        0,2366      0,8133

  Объясненная сумма квадратов = 219,862

Тестовая статистика: LM = 109,931079,
p-значение = P(Хи-квадрат(6) > 109,931079) = 0,000000
```

$p - value = 0$, следовательно, H_0 отвергается, есть проблема гетероскедастичности

Гетероскедастичность.
Способы устранения

Способы устранения гетероскедастичности

Если бы дисперсии всех ошибок были заранее известны, то для устранения гетероскедастичности достаточно было бы оценить исходное уравнение, поделенное почленно на стандартные отклонения ошибок

$$\frac{Y_i}{\sigma_i} = \alpha \frac{1}{\sigma_i} + \beta_1 \frac{X_{1i}}{\sigma_i} + \ldots + \beta_k \frac{X_{ki}}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}, \quad i = 1, \ldots, n$$

В таком уравнении все ошибки имели бы одну и ту же дисперсию, равную единице, поэтому оценки МНК коэффициентов были бы несмещенными и эффективными.

Взвешенный МНК

В реальных условиях дисперсии ошибок заранее неизвестны, и их надо оценивать. Тогда для устранения гетероскедастичности надо делить исходное уравнение на оценки стандартных отклонений

$$\frac{Y_i}{\hat{\sigma}_i} = \alpha \frac{1}{\hat{\sigma}_i} + \beta_1 \frac{X_{1i}}{\hat{\sigma}_i} + \ldots + \beta_k \frac{X_{ki}}{\hat{\sigma}_i} + \frac{\varepsilon_i}{\hat{\sigma}_i}, \quad i = 1, \ldots, n$$

В этом и заключается суть взвешенного метода наименьших квадратов.

Где взять эти оценки стандартных отклонений?

Эти оценки можно получить, например, из тестов Глейзера. Если, скажем, оказалась значима зависимость $|e| = \alpha + \beta X_j + u$, то можно все исходное уравнение поделить на $\hat{\sigma} = \hat{\alpha} + \hat{\beta} X_j$.

Взвешенный метод наименьших квадратов

Обозначим преобразованные переменные звездочками:

$Y_i^* = Y_i / \hat{\sigma}_i,\ X_{ji}^* = X_{ji} / \hat{\sigma}_i$, где $j = 1, \ldots, k, i = 1, \ldots, n.$

Тогда МНК-оценки коэффициентов регрессии с преобразованными данными имеют вид:

$$\hat{\beta}^* = \left(X^{*\prime} X^* \right)^{-1} \left(X^{*\prime} Y^* \right)$$

Эту оценку можно выразить через исходные данные:

$$\hat{\beta}^* = \left(X' \Sigma^{-1} X \right)^{-1} \left(X' \Sigma^{-1} Y \right)$$

где $\Sigma = \begin{pmatrix} \hat{\sigma}_1^2 & \ldots & 0 \\ 0 & \ddots & 0 \\ 0 & \ldots & \hat{\sigma}_n^2 \end{pmatrix}$. Это и есть оценки взвешенного метода наименьших квадратов.

Взвешенный метод наименьших квадратов -2

Ковариационная матрица оценок коэффициентов в таком случае имеет вид:

$$V\left(\hat{\beta}^{*}\right)=\left(X'\Sigma^{-1}X\right)^{-1}$$

В силу теоремы Гаусса-Маркова соответствующие оценки дисперсий оценок коэффициентов являются эффективными.

Обобщенный метод наименьших квадратов

В случае, когда ковариационная матрица Σ может быть недиагональной (при автокорреляции), но про нее известно, что это положительно определенная матрица (назовем ее матрицей Ω), то оценка коэффициентов линейной регрессионной модели обобщенным методом наименьших квадратов (ОМНК, GLS) будет выглядеть так:

$$\hat{\beta}_{\text{омнк}}=(X'\Omega^{-1}\mathrm{X})^{-1}(X'\,\Omega^{-1}\,\mathrm{Y}).$$

Таким образом, взвешенный метод наименьших квадратов – частный случай обобщенного метода наименьших квадратов. Только в случае взвешенного метода матрица $\Omega = \Sigma$ является диагональной.

Теорема Айткена

Если

- модель линейна по параметрам и правильно специфицирована
  $y=X\beta+\varepsilon$,
- матрица X — детерминированная матрица, имеющая максимальный ранг k,
- $E(\varepsilon)=0$,
- $V(\varepsilon)=\Omega$ — произвольная положительно определенная и симметричная матрица,

то оценка вектора коэффициентов модели

$\hat{\beta}_{\text{омнк}}=(X'\Omega^{-1}X)^{-1}(X'\,\Omega^{-1}\,Y)$ является:

- несмещенной
- и эффективной, то есть имеет наименьшую ковариационную матрицу в классе всех несмещенных и линейных по $\boldsymbol{y}$ оценок.

Доступный обобщенный метод наименьших квадратов

Поскольку Ω обычно неизвестна, принято вместо нее использовать оценку $\hat{\Omega}$. В таком случае оценки коэффициентов доступным обобщенным методом наименьших квадратов (feasible GLS) будут такими:

$$\hat{\beta}_{FGLS} = \left(X'\hat{\Omega}^{-1} X \right)^{-1} \left(X'\hat{\Omega}^{-1} Y \right)$$

Другие способы устранения

- Можно использовать оценки стандартных ошибок в форме Уайта, позволяющие учесть гетероскедастичность.
- Другим способом устранения гетероскедастичности может быть изменение функциональной формы регрессионной зависимости, например, логарифмирование.

Робастные стандартные ошибки в форме Уайта

Наиболее распространенным способом коррекции гетероскедастичности в общем виде является использование оценок в форме Уайта для дисперсии коэффициентов:

$$\hat{V}\left(\hat{\beta}\right) = \frac{1}{n}\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i'\right)\left(\frac{1}{n}X'X\right)^{-1}$$

где x_i' – i-я строка матрицы X, i =1, …, n.

Диагональные элементы в матрице $\hat{V}(\hat{\beta})$ являются оценками дисперсий оценок коэффициентов $\widehat{\beta_1}, \ldots, \widehat{\beta_k}$.

Оценки Уайта

Оценки Уайта (HC0: heteroskedasticity consistent, состоятельные при гетероскедастичности) являются состоятельными, но на малых выборках даже они не полностью корректируют смещение оценок стандартных ошибок коэффициентов.

Другие поправки

При наличии в выборке выбросов – наблюдений, сильно отличающихся от остальных, используют поправки:

- HC2
- HC3

Исследования показывают, что на конечных выборках HC2 и HC3 дают более точные результаты, чем HC0.

Поправки на гетероскедастичность

Пусть e_i^2 - квадрат остатков регрессии, а оценка ковариационная матрица коэффициентов регрессии

$X'\sum X$

Тогда поправки:

HC0: $\sum = \mathrm{diag}\left\{e_i^2\right\}$ (White, 1980)

HC2: $\sum = \mathrm{diag}\left\{\dfrac{e_i^2}{1-h_i}\right\}$, где h_i – это диагональный элемент матрицы проектор $P_X = X\left(X'X\right)^{-1}X'$ (MacKinnon and White, 1985)

HC3: $\sum = \mathrm{diag}\left\{\left(\dfrac{e_i}{1-h_i}\right)^2\right\}$ (Davidson and MacKinnon, 1993)

Пример. Затраты пользователей в мобильном приложении

Данные содержат информацию о 150 пользователях некоторого мобильного приложения:

Expend — затраты пользователя на покупки в некотором приложении;

Time — среднее время, которое пользователь проводил в приложении, минуты;

Age1 — 1 для пользователей от 18 до 21 года, 0 иначе;

Age2 — 1 для пользователей от 22 до 25 лет, 0 иначе;

Age3 — 1 для пользователей от 26 до 29 лет, 0 иначе;

Age4 — 1 для пользователей от 30 до 34 лет, 0 иначе;

Age5 — 1 для пользователей от 35 лет и старше, 0 иначе;

MPrice — рыночная стоимость используемой модели смартфона.

Источник: Борзых, Вакуленко, Фурманов (2021)

Пример. Затраты пользователей в мобильном приложении

Изучается зависимость затрат в мобильном приложении от прочих характеристик.

- Оцените уравнение:

$$\text{Expend}_i = \alpha + \beta_1 \text{Time}_i + \beta_2 \text{MPrice}_i + \beta_3 Age1_i + \beta_4 Age2_i + \\ + \beta_5 Age3_i + \beta_6 Age4_i + \varepsilon_i$$

- Примените различные поправки на гетероскедастичность (HC0, HC2, HC3) и сравните результат.

Источник: Борзых, Вакуленко, Фурманов (2021) / m.alterainvest.ru

Оценка регрессии в Gretl

$$\widehat{\text{Expend}}_i = 3.36 + 0.50 \cdot \text{Time}_i + 0.005 \cdot \text{MPrice}_i +$$
$$+1.27 \cdot Age1_i + 5.86 \cdot Age2_i +$$
$$+5.38 \cdot Age3_i + 0.27 \cdot Age4_i$$

gretl: модель 1

Файл Правка Тесты Сохранить Графики Анализ LaTeX

```
Модель 1: МНК, использованы наблюдения 1-149
Зависимая переменная: Expend
Пропущены из-за совершенной коллинеарности: Age5

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const      3,36816       1,84825        1,822        0,0705     *
  Time       0,502159      0,0451424     11,12         4,61e-021  ***
  MPrice     0,00454373    0,00299223     1,519        0,1311
  Age1       1,26542       1,46240        0,8653       0,3883
  Age2       5,86393       0,790107       7,422        9,71e-012  ***
  Age3       5,37991       0,797982       6,742        3,64e-010  ***
  Age4       0,273717      0,632526       0,4327       0,6659

Среднее зав. перемен    13,22081   Ст. откл. зав. перемен  4,811603
Сумма кв. остатков      1266,340   Ст. ошибка модели       2,986283
R-квадрат               0,630420   Испр. R-квадрат         0,614804
F(6, 142)               40,36993   Р-значение (F)          2,15e-28
Лог. правдоподобие     -370,8473   Крит. Акаике            755,6947
Крит. Шварца            776,7223   Крит. Хеннана-Куинна    764,2378

Исключая константу, наибольшее p-значение получено для переменной 7 (Age4)
```

Но мы помним, что есть проблема гетероскедастичности!

Регрессия с поправкой Уайта (HC0)

gretl: модель 2

Файл Правка Тесты Сохранить Графики Анализ LaTeX

```
Модель 2: МНК, использованы наблюдения 1-149
Зависимая переменная: Expend
Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC0
Пропущены из-за совершенной коллинеарности: Age5

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const      3,36816       1,86338        1,808        0,0728    *
  Time       0,502159      0,0842282      5,962        1,88e-08  ***
  MPrice     0,00454373    0,00250293     1,815        0,0716    *
  Age1       1,26542       0,768035       1,648        0,1016
  Age2       5,86393       0,944152       6,211        5,49e-09  ***
  Age3       5,37991       0,861332       6,246        4,60e-09  ***
  Age4       0,273717      0,469957       0,5824       0,5612

Среднее зав. перемен    13,22081   Ст. откл. зав. перемен  4,811603
Сумма кв. остатков      1266,340   Ст. ошибка модели       2,986283
R-квадрат               0,630420   Испр. R-квадрат         0,614804
F(6, 142)               16,31549   Р-значение (F)          3,09e-14
Лог. правдоподобие     -370,8473   Крит. Акаике            755,6947
Крит. Шварца            776,7223   Крит. Хеннана-Куинна    764,2378

Исключая константу, наибольшее р-значение получено для переменной 7 (Age4)
```

Стандартные ошибки изменились! На 10% стала значима переменная MPrice

Регрессия с поправкой HC2

gretl: модель 3

Файл Правка Тесты Сохранить Графики Анализ LaTeX

```
Модель 3: МНК, использованы наблюдения 1-149
Зависимая переменная: Expend
Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC2
Пропущены из-за совершенной коллинеарности: Age5

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  --------------------------------------------------------------
  const      3,36816       1,95041        1,727        0,0864     *
  Time       0,502159      0,0897509      5,595        1,10e-07   ***
  MPrice     0,00454373    0,00258090     1,761        0,0805     *
  Age1       1,26542       0,835731       1,514        0,1322
  Age2       5,86393       0,984966       5,953        1,96e-08   ***
  Age3       5,37991       0,888075       6,058        1,17e-08   ***
  Age4       0,273717      0,480057       0,5702       0,5695

Среднее зав. перемен    13,22081   Ст. откл. зав. перемен   4,811603
Сумма кв. остатков      1266,340   Ст. ошибка модели        2,986283
R-квадрат               0,630420   Испр. R-квадрат          0,614804
F(6, 142)               14,67925   Р-значение (F)           5,31e-13
Лог. правдоподобие     -370,8473   Крит. Акаике             755,6947
Крит. Шварца            776,7223   Крит. Хеннана-Куинна     764,2378

Исключая константу, наибольшее р-значение получено для переменной 7 (Age4)
```

Стандартные ошибки изменились! На 10% стала значима переменная MPrice

Регрессия с поправкой HC3

gretl: модель 4

Файл Правка Тесты Сохранить Графики Анализ LaTeX

```
Модель 4: МНК, использованы наблюдения 1-149
Зависимая переменная: Expend
Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC3
Пропущены из-за совершенной коллинеарности: Age5

             Коэффициент   Ст. ошибка   t-статистика   P-значение
  ---------------------------------------------------------------
  const      3,36816       2,04449         1,647        0,1017
  Time       0,502159      0,0957246       5,246        5,54e-07  ***
  MPrice     0,00454373    0,00266299      1,706        0,0901    *
  Age1       1,26542       0,913108        1,386        0,1680
  Age2       5,86393       1,02878         5,700        6,68e-08  ***
  Age3       5,37991       0,915768        5,875        2,88e-08  ***
  Age4       0,273717      0,490522        0,5580       0,5777

Среднее зав. перемен    13,22081   Ст. откл. зав. перемен  4,811603
Сумма кв. остатков      1266,340   Ст. ошибка модели       2,986283
R-квадрат               0,630420   Испр. R-квадрат         0,614804
F(6, 142)               13,22979   Р-значение (F)          7,21e-12
Лог. правдоподобие     -370,8473   Крит. Акаике            755,6947
Крит. Шварца            776,7223   Крит. Хеннана-Куинна    764,2378

Исключая константу, наибольшее р-значение получено для переменной 7 (Age4)
```

Стандартные ошибки изменились! На 10% стала значима переменная MPrice

Сравнение стандартных ошибок

В таблице по столбцам указаны стандартные ошибки.

	Без поправок	HC0	HC2	HC3
const	1,848	1,863	1,950	2,045
Time	0,045	0,084	0,090	0,096
MPrice	0,003	0,0025	0,0026	0,0027
Age1	1,462	0,768	0,836	0,913
Age2	0,790	0,944	0,985	1,029
Age3	0,798	0,861	0,888	0,916
Age4	0,633	0,470	0,480	0,491

Литература

Доугерти К. (1992). Введение в эконометрику. М. Инфра-М. Глава 7 (7.1-7.4).

Берндт, Э. Р. Практика эконометрики: классика и современность. М.: ЮНИТИ-ДАНА, 2005. - 863 с. Глава 4 (4.5А).

Newbold P. (1995) Statistics for Business and Economics. 4th ed. London: Prentice-Hall. Глава 13, 14 (14.7).

Вербик М. Путеводитель по современной эконометрике. Научная книга, 2008. Глава 4.1-4.5.

Борзых Д. А., Вакуленко Е. С., Фурманов К. К. Эконометрика: работа с данными на компьютере. Практикум: Элементы теории. Практические задания. Ответы и решения. Издательская группа URSS, 2021. Глава 4 и 5.

Вакуленко Е. С., Ратникова Т. А., Фурманов К. К. Эконометрика (продвинутый курс). Применение пакета Stata. М. : Юрайт, 2020. Глава 7 и 8.6.