ЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ С ГЕТЕРОСКЕДАСТИЧНЫМИ ОСТАТКАМИ

Несмещенность

$$M(b_j) = b_j^{ген}$$

Эффективность

$$M(b_j - b_j^{ген})^2 = \sigma_{b_j}^2 = \min \sigma_{b_j}^2$$

Состоятельность

$$b_j \xrightarrow[n \to \infty]{} b_j^{ген}$$

Задачи регрессионного анализа:

- **Построение модели**
- **Исследование случайных отклонений**

$$\varepsilon_i = y_i - \hat{y}_i$$

Пять предпосылок МНК:

1. Случайный характер остатков

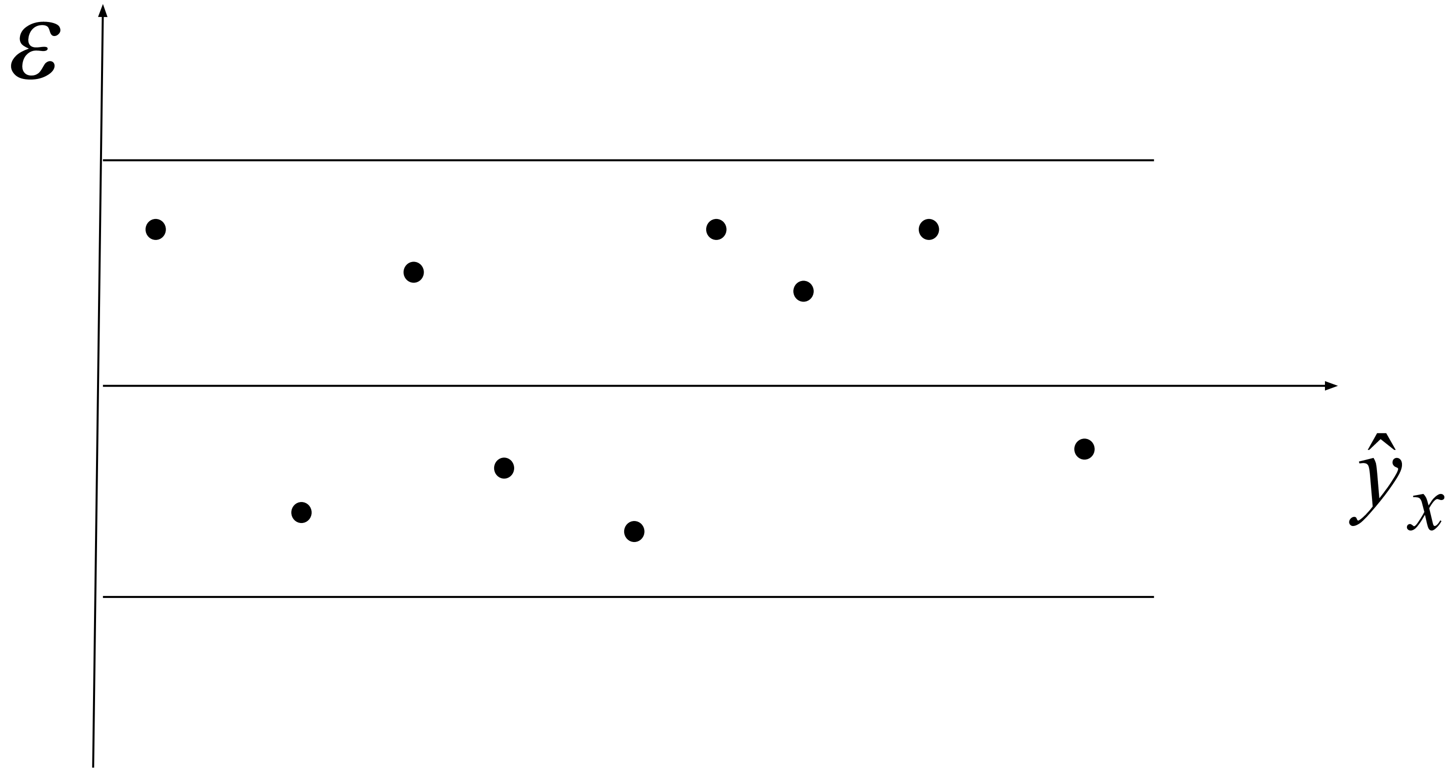
2. Нулевая средняя величина остатков, не зависящая от x_i

3. Гомоскедастичность

4. Отсутствие автокорреляции остатков

5. Остатки подчиняются нормальному распределению

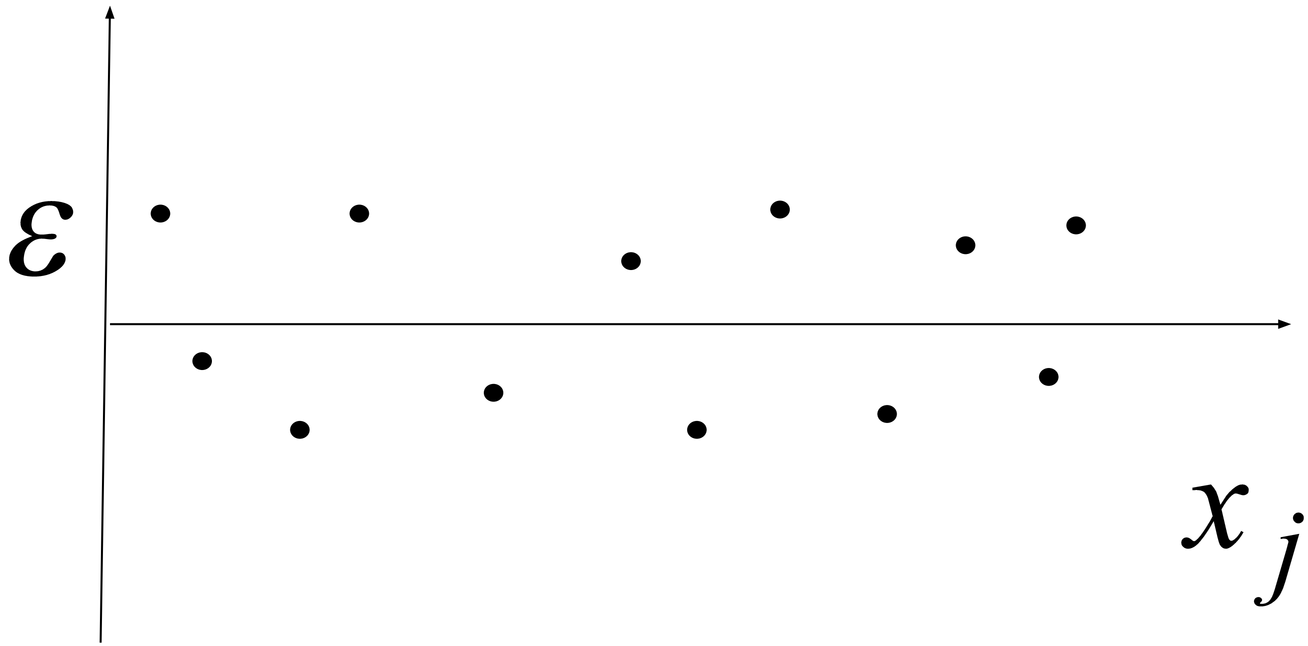
1. Случайный характер остатков

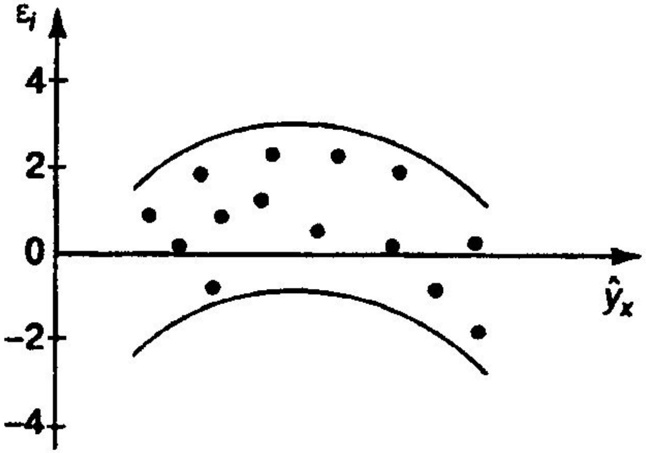


2. Нулевая средняя величина остатков

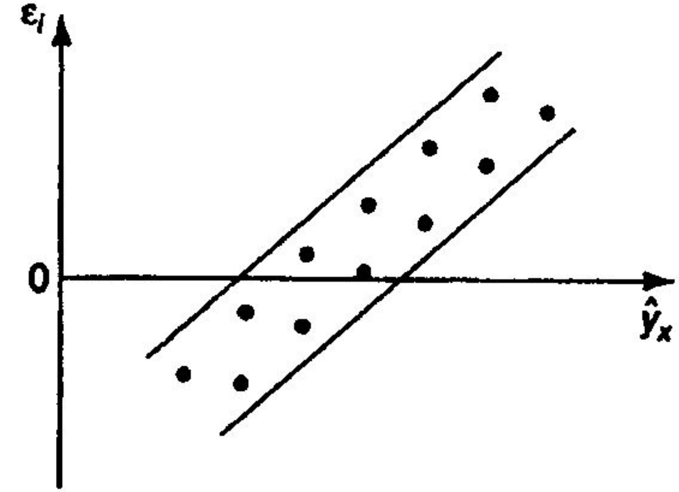
$$\sum(y-\hat{y}_x)=0$$

2. Независимость величины остатков от величины фактора *x*

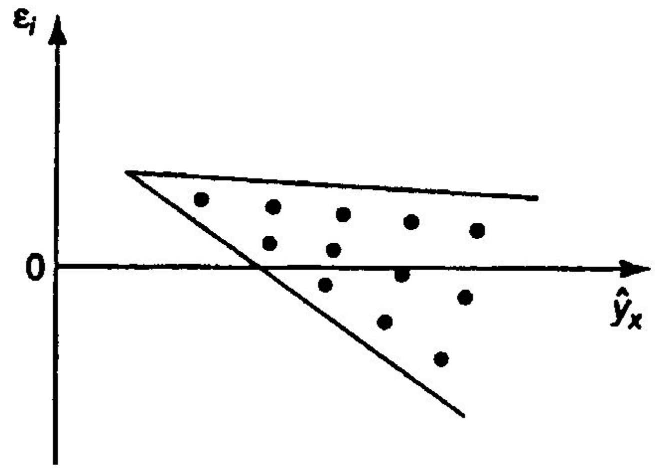




ε_i
4
2
0
−2
−4
$\hat{y}_x$



ε_i
0
$\hat{y}_x$



ε_i
0
$\hat{y}_x$

- **нарушение третьей предпосылки МНК.**
- **неправильная спецификация модели.**
- **наличие систематической погрешности модели.**

3. Гомоскедастичность дисперсии остатков

1. Оценки коэффициентов - несмещенные и линейные.

2. Оценки коэффициентов не эффективные

3. Дисперсии оценок будут рассчитываться со смещением.

4. Ненадежные интервальные оценки коэффициентов.

4. Отсутствие автокорреляции остатков

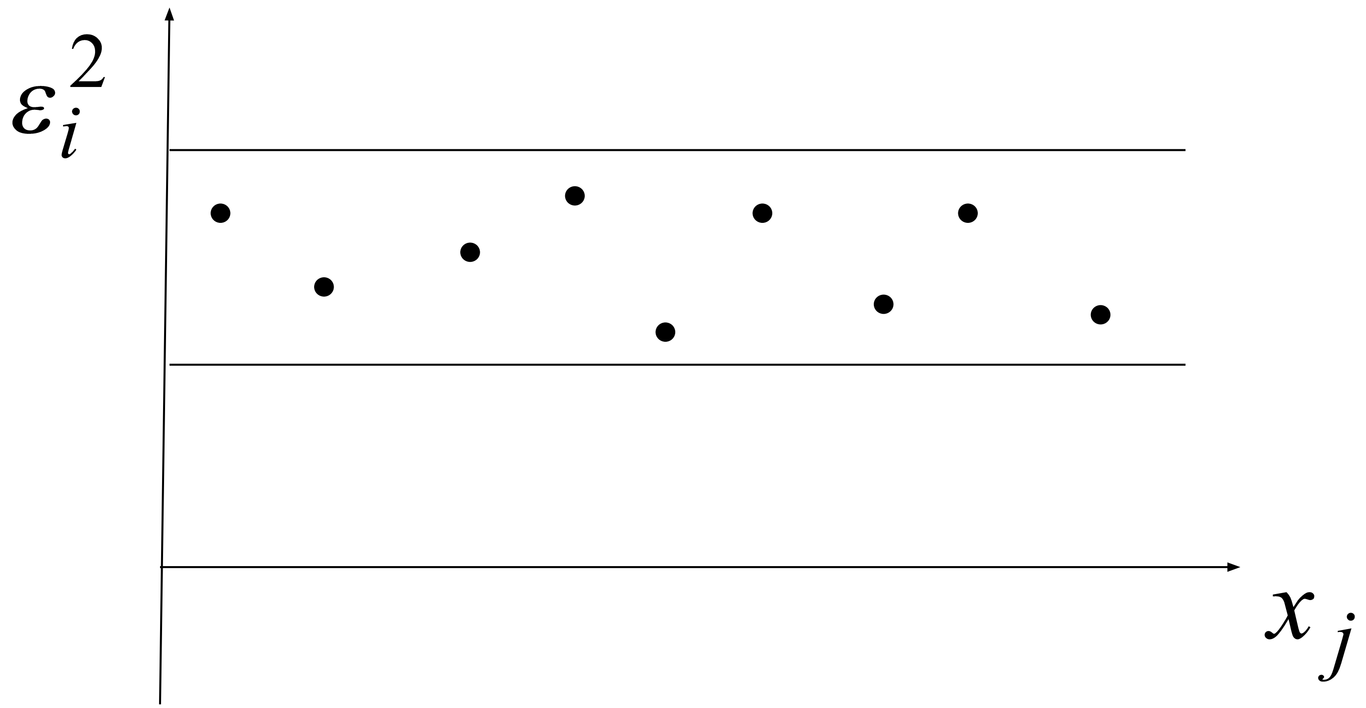
Коэффициент корреляции

$$r_{\varepsilon_i \varepsilon_{i-1}} = \frac{\operatorname{cov}(\varepsilon_i, \varepsilon_{i-1})}{\sigma_{\varepsilon_i} \cdot \sigma_{\varepsilon_{i-1}}}$$

$$r_{\varepsilon_i \varepsilon_{i-1}} > 0 \Longrightarrow$$ **Остатки автокоррелированы**

Линейные регрессионные модели с гетероскедастичными остатками

Графический анализ остатков



Линейные регрессионные модели с гетероскедастичными остатками

<u>Тест Гольдфельда-Кванта</u>

$$\sigma_i = \sigma(x_i) \propto x_i$$

$$\sigma_i^2 = \sigma^2 \cdot x_i^2$$

Линейные регрессионные модели с гетероскедастичными остатками

<u>Тест Гольдфельда-Кванта</u>

1. *Все наблюдения упорядочиваются по доминирующему фактору*

2. *Упорядоченная совокупность делится на три группы:*

$$k,\ (n-2\cdot k),\ k$$

3. *По первой и третьей группе оцениваются регрессии*

Тест Гольдфельда-Кванта

4. $H_0 : S_1 = S_3 \qquad H_1 : S_1 \neq S_3$

5. $$F_{факт} = \frac{S_3 /(k-m-1)}{S_1/(k-m-1)} = \frac{S_3}{S_1}.$$

$$F_{факт} = \frac{S_3}{S_1} > F_{табл} \Longrightarrow$$ **Остатки гетероскедастичны**

$$df_1 = df_2 = k - m - 1$$

Тесты Парка, Глэйзера, Уайта

1.Парка

$$\ln \varepsilon_i^2 = a + b \cdot \ln x_{ji} + v_i$$

$$x_{ji} - i-е\ значение\ j-го\ фактора$$

$$v_i - случайный\ остаток$$

Тесты Парка, Глэйзера, Уайта

1.Глэйзера

$$|\varepsilon_i| = a + bx_{ji}^k + v_i \ ,$$

$$k - \text{какое} - \text{либо} \ \text{число}, \ \text{например},$$

$$k = -1; -0{,}5; 0{,}5; 1$$

2.Уайта

$$\varepsilon_i^2 = a + b_{11}x_{1i} + b_{12}x_{1i}^2 + b_{21}x_{2i} + b_{22}x_{2i}^2 + c_{12}x_{1i}x_{2i} + v_i$$

Тест ранговой корреляции Спирмена

1. Ранжируются значения модулей остатков и значения выбранного фактора

2. Вычисляется коэффициент Спирмена

$$r_{x,\varepsilon} = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)}$$

d - разность между рангами i-го остатка и i–го фактора

Тест ранговой корреляции Спирмена

3.

$$H_0 : r_{x,\varepsilon} = 0$$

$$H_1 = r_{x,\varepsilon} \neq 0$$

$$t = \frac{r_{x,\varepsilon}\sqrt{n-2}}{\sqrt{1-r_{x,\varepsilon}^2}}$$

$$t_{факт} > t_{табл} \longrightarrow$$ **остатки гетероскедастичны**

$$df = n - 2$$

Обобщенный метод наименьших квадратов

$$y_i = a + b \cdot x_i + \varepsilon_i \qquad \sigma_{\varepsilon_i}$$

$$\frac{y_i}{\sigma_{\varepsilon_i}} = \frac{a}{\sigma_{\varepsilon_i}} + b \cdot \frac{x_i}{\sigma_{\varepsilon_i}} + \frac{\varepsilon_i}{\sigma_{\varepsilon_i}}$$

$$D\left(\frac{\varepsilon_i}{\sigma_{\varepsilon_i}}\right) = \frac{1}{\sigma_{\varepsilon_i}^2} \cdot \sigma^2(\varepsilon_i) = \frac{\sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_i}^2} = 1$$

Обобщенный метод наименьших квадратов

$$\sigma^2_{\varepsilon_i} = \sigma^2 \cdot K_i$$

$$y_i = a + bx_i + \varepsilon_i \qquad \frac{y_i}{\sqrt{K_i}} = \frac{a}{\sqrt{K_i}} + b \cdot \frac{x_i}{\sqrt{K_i}} + \frac{\varepsilon_i}{\sqrt{K_i}}$$

$$\sigma^2_{\varepsilon_i} = \sigma^2(\varepsilon_i) = \sigma^2 \cdot K_i$$

$$\sigma^2\left(\frac{\varepsilon_i}{\sqrt{K_i}}\right) = \frac{1}{K_i} \cdot \sigma^2(\varepsilon_i) = \frac{1}{K_i} \cdot \sigma^2 \cdot K_i = \sigma^2 \Rightarrow const$$

Обобщенный метод наименьших квадратов

$$S(a,b)=\sum_{i=1}^{n}\frac{1}{K_i}\left(y_i-a-bx_i\right)^2$$

$$b=\frac{\sum\frac{1}{K}\cdot x\cdot y}{\sum\frac{1}{K}\cdot x^2}$$

$$b=\frac{\sum x\cdot y}{\sum x^2}$$

Обобщенный метод наименьших квадратов

$$y_i = a + b_1 \cdot x_{1_i} + b_2 \cdot x_{2_i} + \varepsilon_i$$

$$K_i^2 \qquad\qquad \sigma^2_{\varepsilon_i} = \sigma^2 \cdot K_i^2$$

$$\frac{y_i}{K_i} = \frac{a}{K_i} + b_1 \cdot \frac{x_{1_i}}{K_i} + b_2 \cdot \frac{x_{2_i}}{K_i} + \frac{\varepsilon_i}{K_i}$$

Обобщенный метод наименьших квадратов

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_m \cdot x_m + \varepsilon$$

$$K_i = x_{1_i} \qquad \sigma^2_{\varepsilon_i} = \sigma^2 \cdot x^2_{1_i}$$

$$\frac{y_i}{x_{1_i}} = a \cdot \frac{1}{x_{1_i}} + b_1 + b_2 \cdot \frac{x_{2_i}}{x_{1_i}} + \ldots + b_m \cdot \frac{x_{m_i}}{x_{1_i}} + \frac{\varepsilon_i}{x_{1_i}}$$

4. Автокорреляция

$$r_{\varepsilon_i \varepsilon_{i-1}} = \frac{\operatorname{cov}(\varepsilon_i, \varepsilon_{i-1})}{\sigma_{\varepsilon_i} \cdot \sigma_{\varepsilon_{i-1}}}$$

$$DW = \frac{\sum (\varepsilon_i - \varepsilon_{i-1})^2}{\sum \varepsilon_i^2} \qquad DW \approx 2 \cdot (1 - r_{\varepsilon_i \varepsilon_{i-1}})$$

4. Автокорреляция

$H_0:$ **Автокорреляция остатков отсутствует**

$H_1:$ **Положительная автокорреляция**

$H_1^*:$ **Отрицательная автокорреляция**

Положительная автокорреляция	**Зона неопределенности**	**Автокорреляция отсутствует**	**Зона неопределенности**	**Отрицательная автокорреляция**

$0 \quad d_L \quad d_U \quad 2 \quad 4-d_U \quad 4-d_L \quad 4$