Самарский государственный экономический университет
Кафедра математической статистики

ГЕТЕРОСКЕДАСТИЧНОСТЬ И ЕЕ ПОСЛЕДСТВИЯ

Обнаружение гетероскедастичности. Устранение гетероскедастичности.

Презентация лекции по курсу “Эконометрика” доцента кафедры математической статистики СГЭУ,
к.ф.-м.н., Ширяевой Людмилы Константиновны
E-mail: Shiryeva_LK@mail.ru

План

- Гетероскедастичность и ее последствия
- Методы обнаружения гетероскедастичности.
- Методы устранения гетероскедастичности. Обобщенный метод наименьших квадратов

Гетероскедастичность и ее последствия

- Свойства эмпирических коэффициентов регрессии напрямую зависят от свойств случайной компоненты *ε.*
- Для получения статистически надежных эмпирических коэффициентов регрессии необходимо следить за выполнимостью условий Гаусса-Маркова.
- При нарушении условий Гаусса-Маркова МНК может давать эмпирические коэффициенты регрессии с плохими статистическими свойствами.

Гетероскедастичность и ее последствия

- **Согласно второму условию Гаусса-Маркова, дисперсия случайного фактора должна быть одинаковой для всех наблюдений, т.е. $D(\varepsilon_i)= D(\varepsilon_j)$.**
- **Выполнение этого условия называется *гомоскедастичностью*, а его нарушение - *гетероскедастичностью*.**

Иллюстрация гомоскедастичности

Вероятность того, что случайная ошибка примет какое-–либо значение одинакова дл всех наблюдений

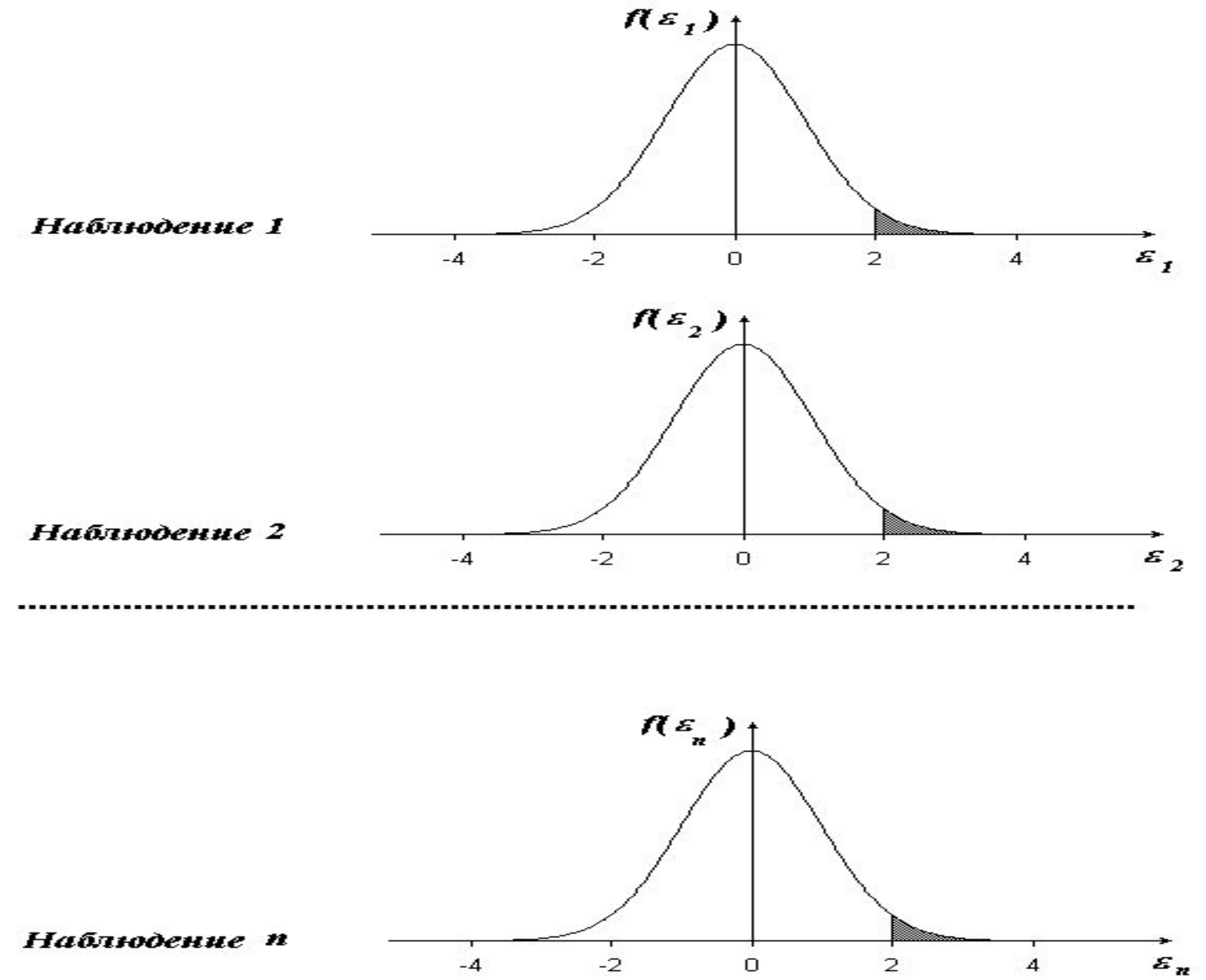
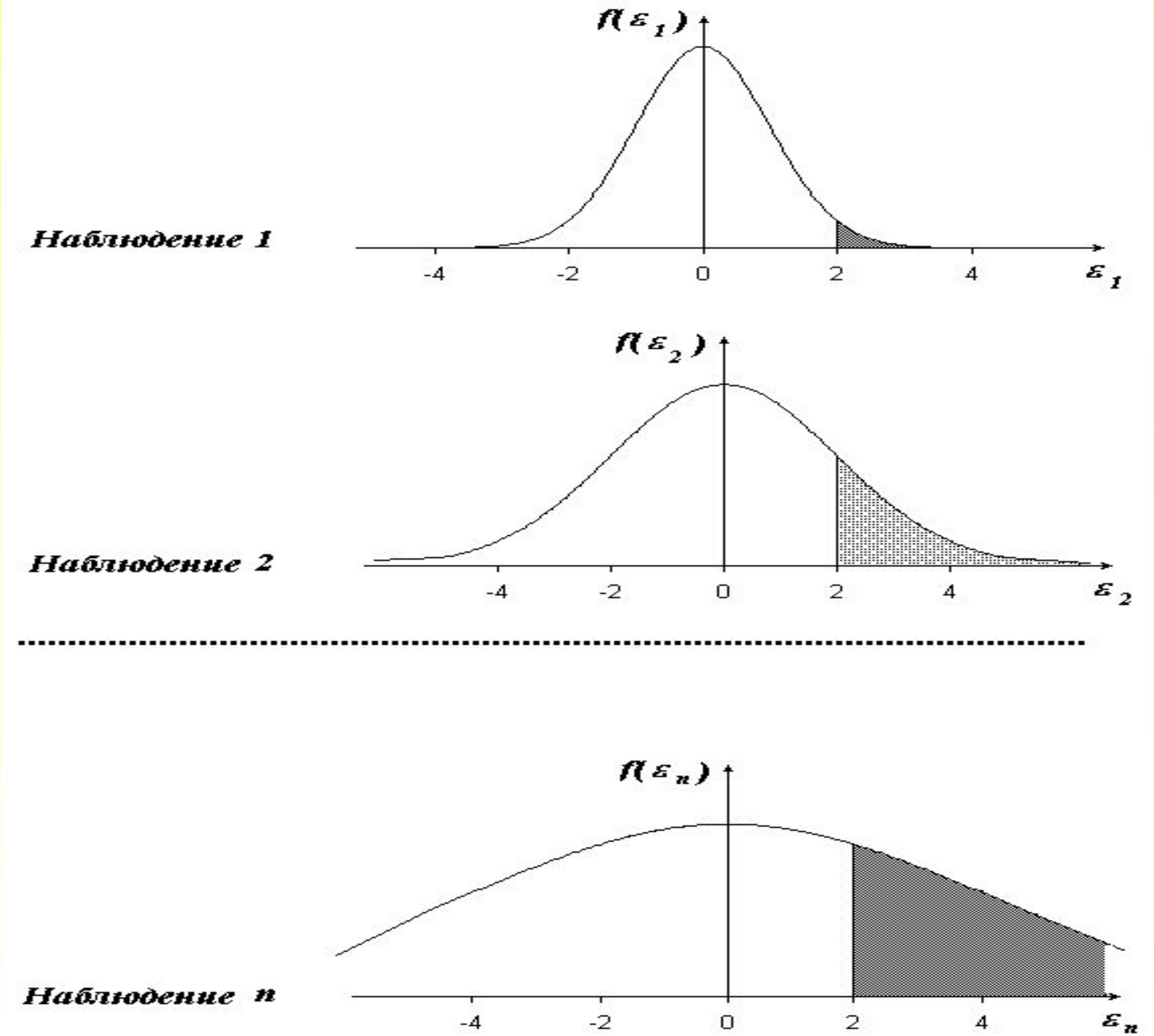


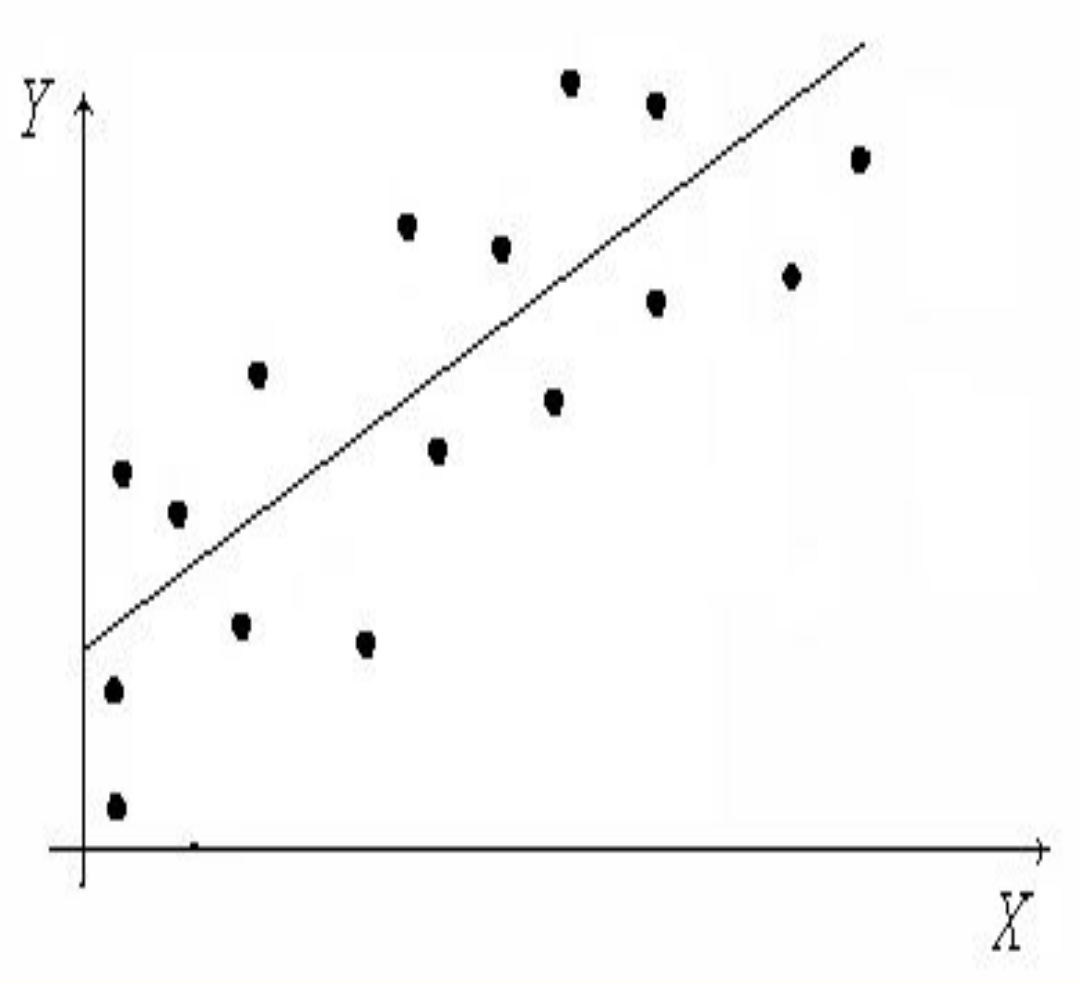
Иллюстрация гетероскедастичности

Вероятность того, что случайная ошибка примет какое-либо значение неодинакова для всех наблюдений



Гомоскедастичность означает “одинаковый разброс”.

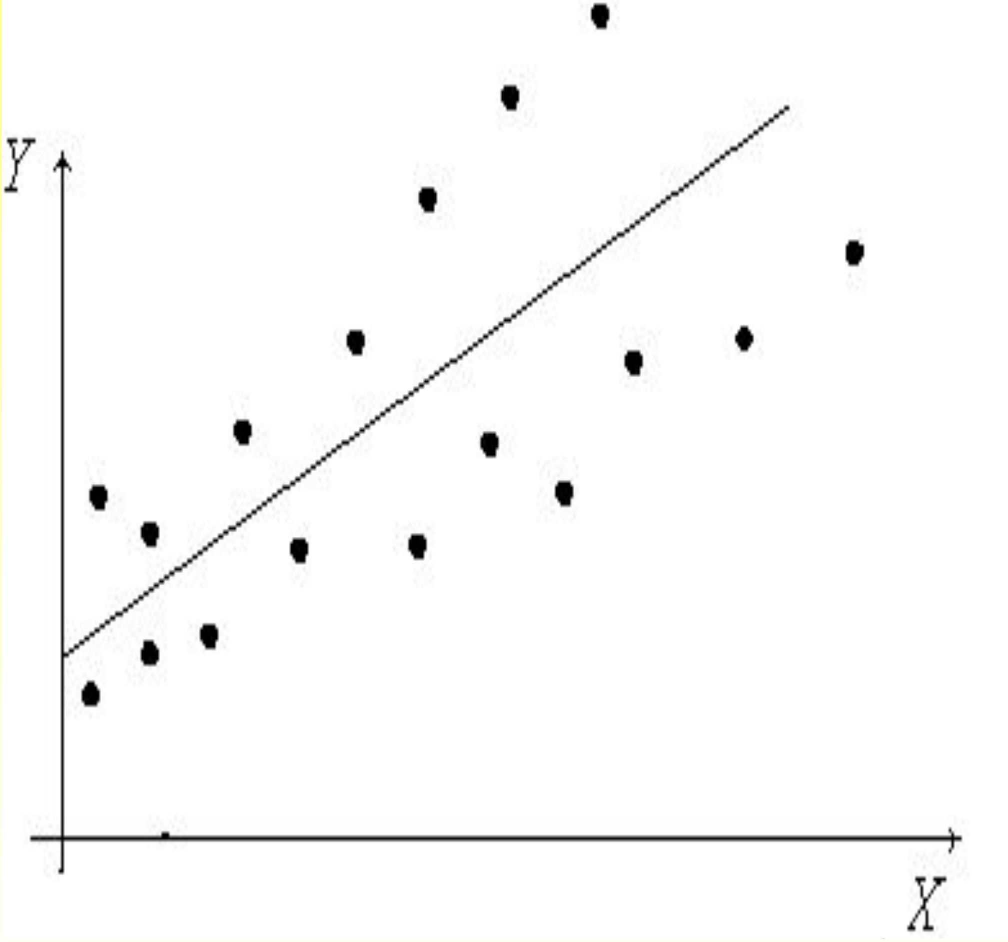
- Типичный вид облака точек в модели с гомоскедастичными остатками



“гетероскедастичное” облако точек

Гетероскедастичность означает “неодинаковый разброс”.

- Типичный вид облака точек в модели с гетероскедастичными остатками



Последствия применения МНК в случае гетероскедастичности

- МНК-оценки не будут являться эффективными;
- формулы для вычисления стандартных ошибок коэффициентов регрессии становятся некорректными;
- дисперсия остатков регрессии становится смещенной оценкой для дисперсии случайной компоненты;
- все выводы, получаемые на основе F – и t -статистик, а также интервальные оценки становятся ненадежными.

Проверка остатков модели на гетероскедастичность

- Первичная проверка на наличие гетероскедастичности осуществляется с помощью визуального анализа поведения остатков регрессии.
- Дальнейшая проверка на наличие гетероскедастичности осуществляется уже с помощью статистических тестов.

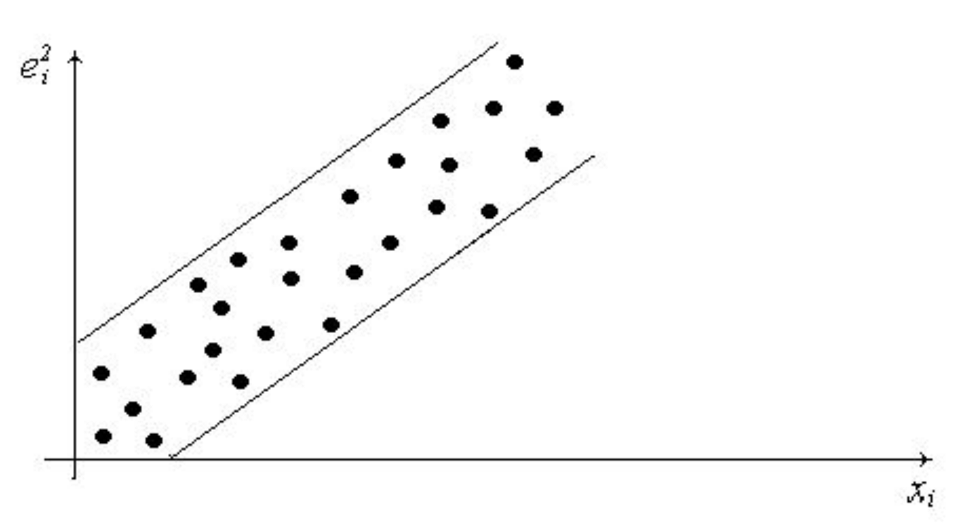
Методы обнаружения гетероскедастичности

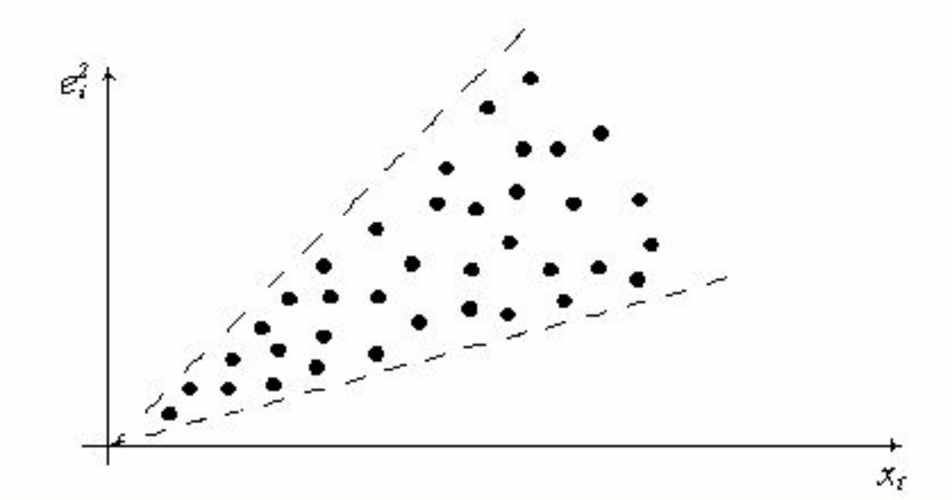
Тесты на гетероскедастичность

- **графический тест**
  - графический анализ остатков
- **статистические тесты**
  - тест Спирмена
  - тест Квандта
  - тест Глейзера

Графический анализ остатков. Гетероскедастичность.

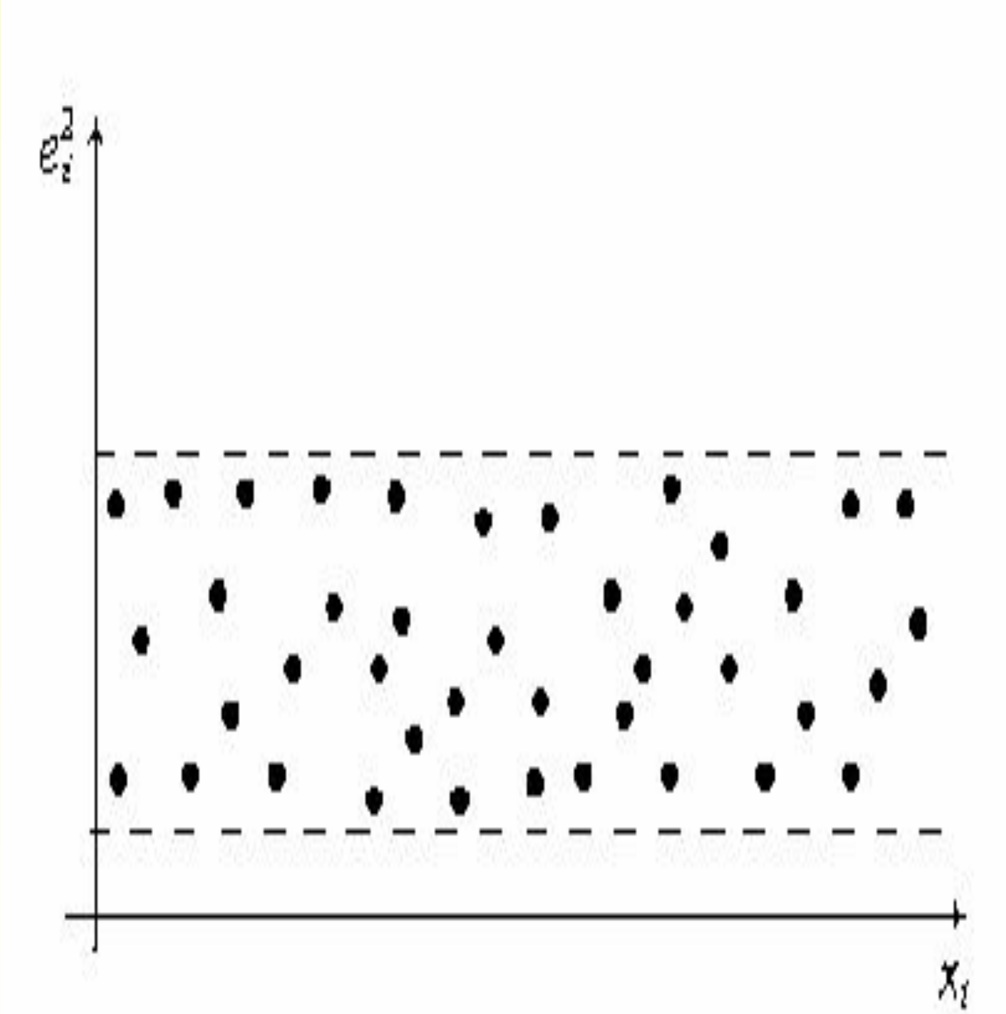
- Если все отклонения расположены внутри расширяющейся или наклонной полосы, то это свидетельствует в пользу гетероскедастичности





Графический анализ остатков. Гомоскедастичность.

- Если все отклонения равномерно заполняют некоторую полосу постоянной ширины, то это свидетельствует в пользу гомоскедастичности.



Сравнительный анализ статистических тестов

Наименование	**Основная идея теста**	**Статистика**
Тест Спирмена	Проверка корелированности остатков регрессии *e* и значений *X*	*t*- статистика
Тест Квандта	Проверка значимости отличий ESS для верхней и нижней третей упорядоченной выборки	*F*- статистика
Тест Глейзера	Подбор формы связи между остатками регрессии е и значениями X	*t*- статистика и *F*- статистика

Общая схема проведения любого статистического теста

$\hat{E}_{\text{íàáë}}$

$\boldsymbol{K}$

$K_{\text{крит}}$

$\boldsymbol{H_0}$

Тест ранговой корреляции Спирмена

- Тест Спирмена проверяет ***коррелированность*** модулей остатков регрессии со значениями объясняющей переменной.

- При использовании этого теста предполагается, что дисперсия случайной ошибки либо уменьшается, либо увеличивается по мере увеличения X.

- При этом точки (x_i,e_i) могут располагаться внутри либо расширяющейся, либо – наклонной полосы (см. слайд 12).

- Теснота взаимосвязи между модулями остатков регрессии $|e_i|$ и значениями x_i оценивается с помощью *выборочного рангового коэффициента корреляции* r_{xe}.

- Если связь между абсолютными величинами остатков регрессии и значениями может быть признана статистически значимой, то принимается гипотеза о наличии гетероскедастичности.

Порядок выполнения теста Спирмена

1) выполняется регрессия переменной Y на переменную X,
2) для каждого i -ого наблюдения вычисляют модуль остатков регрессии $|e_i|$;
3) значения x_i и модули $|e_i|$ ранжируются, т. е. упорядочиваются по возрастанию;
4) вычисляются ранги – порядковые номера значений в ранжированном ряде из значений x_i, и ранги $|e_i|$ – порядковые номера значений в ранжированном ряде, составленном из модулей остатков;
5) для каждого i-ого наблюдения вычисляется значение d_i как разность между рангами x_i, и $|e_i|$ (пусть, например, наблюдаемое значение объясняющей переменной x_{11} является 33-им по величине, т.е. ранг x_{11} равен 33, а $|e_{11}|$ является 5-ым по величине, т.е. ранг $|e_{11}|$ равен 5, тогда $d_{11}=33-5=28$);
6) вычисляется выборочный коэффициент ранговой корреляции по следующей формуле:

$$r_{xe} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

Порядок выполнения теста Спирмена (окончание)

- 7) выдвигаются нулевая и альтернативная гипотезы:
- H_0 :(ранговый коэффициент корреляции для генеральной совокупности $\rho_{xe} = 0$, или гетероскедастичность отсутствует);
- H_1 :(ранговый коэффициент корреляции для генеральной совокупности ρ_{xe} отличен от 0, или гетероскедастичность имеет место);
- 8) статистика для проверки H_0 имеет вид:
- $$t = \frac{r_{xe}\sqrt{n-2}}{\sqrt{1-r_{xe}^2}} \sim t(n\text{-}2)$$
- 9) строится двусторонняя критическая область $|t| > t_{\text{кр.дв}}(\alpha, n-2)$;
- 10) если наблюдаемое значение t-статистики попадает в критическую область, то принимается гипотеза о наличии гетероскедастичности; если же наблюдаемое значение t-статистики попадает в область принятия гипотезы, то принимается гипотеза о наличии гомоскедастичности.
- ***Замечание.*** Если в модели более одной объясняющей переменной, то с помощью t-статистики проверка гипотезы может выполняться для каждой из переменных отдельно..

Тест Голдфелда-Квандта

- Тест Голдфелда-Квандта предполагает, что с ростом x_i
- дисперсия $D(\varepsilon_i)$ либо растет, т.е.

$$\sigma(\varepsilon_i) \approx \lambda x_i$$

- либо падает, т.е.

$$\sigma(\varepsilon_i) \approx \frac{\lambda}{x_i}$$

Порядок выполнения теста Голдфелда-Квандта

1) все наблюдений упорядочиваются по величине объясняющей переменной;

2) упорядоченная выборка разбивается на три подвыборки объемом $[n/3]$;

3) средняя треть наблюдений отбрасывается и оцениваются отдельные регрессии для верхней и нижней подвыборок;

4) вычисляются дисперсии остатков регрессии для верхней (S_1^2) и нижней (S_2^2) подвыборок;

5) выдвигают основную гипотезу

H_0 : модель является гомоскедастичной;

против альтернативной

H_1 :модель является гетероскедастичной;

Порядок выполнения теста Голдфелда-Квандта (окончание)

- гипотеза H_0: проверяется с помощью статистики:

$$F = \frac{S_2^2}{S_1^2} \sim \mathbf{F}([n/3]-m-1;[n/3]-m-1)$$

- при выбранном уровне значимости α строится правосторонняя критическая область , описываемая неравенством: $F > F_{\text{кр}}(\alpha, k_1, k_1)$;
- вычисляется наблюдаемое значение F-критерия;
- далее следует выполнить проверку гипотезы H_0 по стандартной схеме.

Тест Глейзера

- Тест Глейзера позволяет обнаружить гетероскедастичность в случае, когда стандартное отклонение случайной компоненты связано со значением X нелинейной зависимостью:

$$\sigma(\varepsilon_i) \approx \alpha + \beta x_i^{\gamma}$$

Порядок выполнения теста Глейзера

1) по МНК оценивается линейная регрессия $\hat{y} = \hat{b}_0 + \hat{b}_1 x$;

2) оценки $\hat{\sigma}(\varepsilon_i)$ стандартных отклонений случайной компоненты в каждом наблюдении вычисляются как:

$$\hat{\sigma}(\varepsilon_i) \equiv |e_i| = |y_i - \hat{y}_i|$$

3) выбирается набор значений показателя степени γ, например, такой:

$$\gamma = -1; -0{,}75; -0{,}5; -0{,}25; 0{,}25;\ 0{,}5; 0{,}75; 1; 1{,}25; 1{,}5; 1{,}75;\ 2$$

4) для каждого γ строится по МНК регрессионная модель вида: $|\hat{e}_i| = \hat{\alpha} + \hat{\beta} x_i^{\gamma}$

Порядок выполнения теста Глейзера (окончание)

- 5) с помощью t- статистики проверяется статистическая значимость каждого коэффициента $\hat{\beta}$;

6) если оценка $\hat{\beta}$ окажется статистически значима, то имеет место гетероскедастичность.

Замечание. Если для нескольких значений параметра γ получены статистически значимые оценки $\hat{\beta}$, то следует выбрать наилучшую из них (т. е. ту, для которой *t*-статистика максимальна).

Методы устранения гетероскедастичности.

- **Основной метод устранения гетероскедастичности – обобщенный метод наименьших квадратов (ОМНК).**

Суть ОМНК – минимизация суммы квадратов отклонений, в которой каждое наблюдение присутствует с учетом его “веса”.

- **Последствия ОМНК – получение эффективных оценок для коэффициентов регрессии.**

Применение ОМНК

•Пусть в исходной модели: $y_i = b_0 + b_1 x_i + \varepsilon_i \; \left(i = \overline{1,n}\right)$

имеет место гетероскедастичность, т. е. $\sigma^2(\varepsilon_i) \neq \sigma^2(\varepsilon_j),\ i \neq j$

Предположим, что дисперсия случайной компоненты в каждом наблюдении известна.

Разделим каждое наблюдение на соответствующее ему значение стандартного отклонения:

$$\frac{y_i}{\sigma(\varepsilon_i)} = \frac{b_0}{\sigma(\varepsilon_i)} + b_1 \frac{x_i}{\sigma(\varepsilon_i)} + \frac{\varepsilon_i}{\sigma(\varepsilon_i)}$$

Положим $y_i^* = \dfrac{y_i}{\sigma(\varepsilon_i)},\ z_i^* = \dfrac{1}{\sigma(\varepsilon_i)},\ q_i^* = \dfrac{x_i}{\sigma(\varepsilon_i)}$ и $\xi_i = \dfrac{\varepsilon_i}{\sigma(\varepsilon_i)}$.

Тогда преобразованная модель примет вид: $y_i^* = b_0 z_i^* + b_1 q_i^* + \xi_i$

Применение ОМНК (продолжение)

- Легко убедиться, что случайная компонента $\xi_i = \dfrac{\varepsilon_i}{\sigma(\varepsilon_i)}$
- в полученной модели имеет нулевое математическое ожидание и единичную дисперсию для всех наблюдений. Действительно,
- $M(\xi_i) = M\left(\dfrac{\varepsilon_i}{\sigma(\varepsilon_i)}\right) = \dfrac{1}{\sigma(\varepsilon_i)} M(\varepsilon_i) = \dfrac{0}{\sigma(\varepsilon_i)} = 0$ и $D(\xi_i) = D\left(\dfrac{\varepsilon_i}{\sigma(\varepsilon_i)}\right) = \dfrac{1}{\sigma^2(\varepsilon_i)} D(\varepsilon_i) = \dfrac{\sigma^2(\varepsilon_i)}{\sigma^2(\varepsilon_i)} = 1$
- Следовательно, оценки коэффициентов регрессии можно найти по обычному МНК, минимизируя следующую сумму квадратов отклонений

$$S = \sum_{i=1}^{n} \left(\frac{y_i}{\sigma(\varepsilon_i)} - \frac{\hat{b}_0}{\sigma(\varepsilon_i)} - \hat{b}_1 \frac{x_i}{\sigma(\varepsilon_i)} \right)^2 = \sum_{i=1}^{n} \frac{1}{\sigma^2(\varepsilon_i)} \left(y_i - \hat{b}_0 - \hat{b}_1 x_i \right)^2 \to \min$$

Применение ОМНК (окончание)

- ***Замечание.*** Основная трудность в применении обобщенного (взвешенного) МНК состоит в том, что значения $\sigma(\varepsilon_i)$, как правило, неизвестны. На практике неизвестные значения либо заменяют их оценками, либо подбирают некоторую величину, пропорциональную в каждом наблюдении стандартному отклонению $\sigma(\varepsilon_i)$.

Заключение

- Построение любой эконометрической модели должно включать проверку выполнимости второго условия Гаусса-Маркова.
- Проверка осуществляется с помощью статистических тестов.
- При нарушении второго условия Гаусса-Маркова следует предпринять шаги к устранению гетероскедастичности.

Заключение

Всякая наука только тогда достигает своего совершенства , когда она породнится с математикой.

И. Кант