Гетероскедастично сть

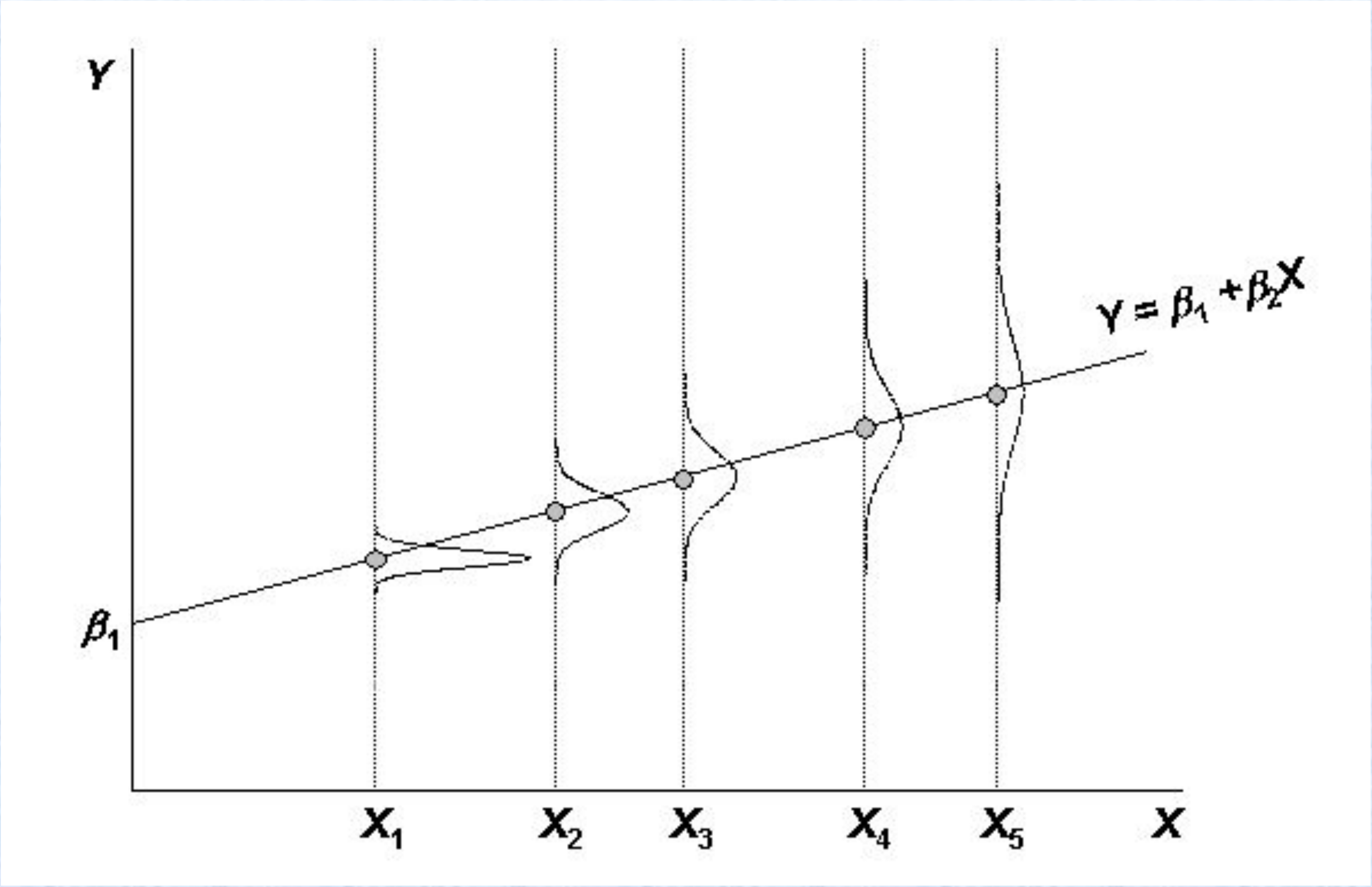
ОПРЕДЕЛЕНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

Гетероскедастичность – это неоднородность наблюдений. Она характеризуется тем, что не выполняется предпосылка 2^0 использования МНК:

$$2^0.\ D[\varepsilon] = \sigma^2 = const$$

Выполнимость предпосылки 2^0 называется гомоскедастичностью.

ИЛЛЮСТРАЦИЯ ГЕТЕРОСКЕДАСТИЧНОСТИ



ГЕТЕРОСКЕДАСТИЧНОСТЬ ОШИБОК

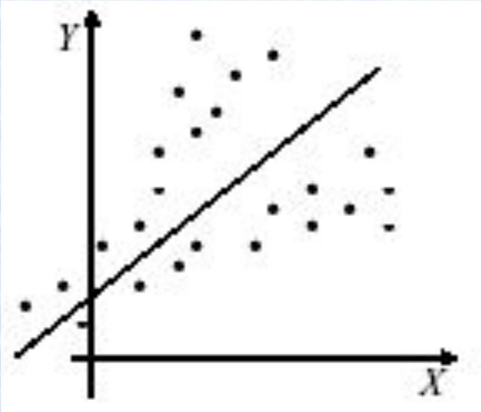
Причиной непостоянства дисперсии σ_{ε}^{2}

эконометрической модели часто является ее зависимость

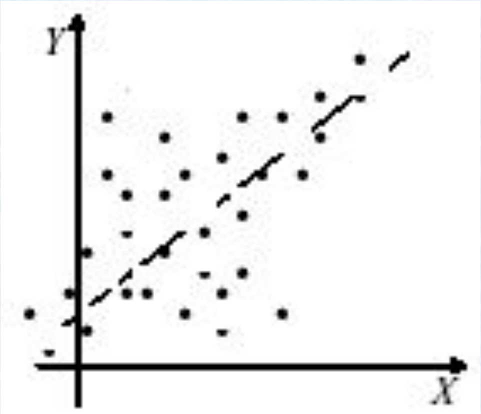
от масштаба рассматриваемых явлений.

В модель ошибка входит как аддитивное слагаемое. В то же время часто она имеет относительный характер и определяется по отношению к измеренному уровню рассматриваемых факторов.

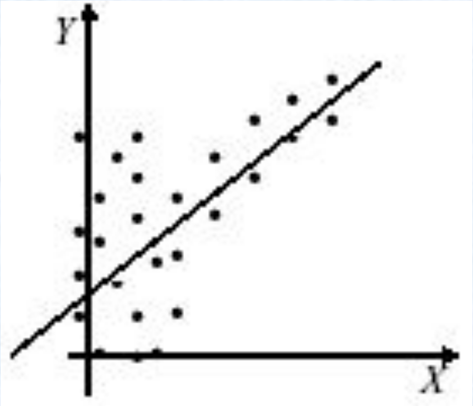
Примеры моделей с гетероскедастичным случайным членом



а)



б)



в)

а) Дисперсия σ_ε^2 растет по мере увеличения значений объясняющей переменной *X*

б) Дисперсия σ_ε^2 имеет наибольшие значения при средних значениях *X*, уменьшаясь по мере приближения к крайним значениям

в) Дисперсия ошибки наибольшая при малых значениях *X*, быстро уменьшается и становится однородной по мере увеличения *X*

ИСТИННАЯ И ЛОЖНАЯ ГЕТЕРОСКЕДАСТИЧНОСТЬ

1. Истинная гетероскедастичность

Вызывается непостоянством дисперсии случайного члена, ее зависимостью от различных факторов

2. Ложная гетероскедастичность

Вызывается ошибочной спецификацией модели регрессии.

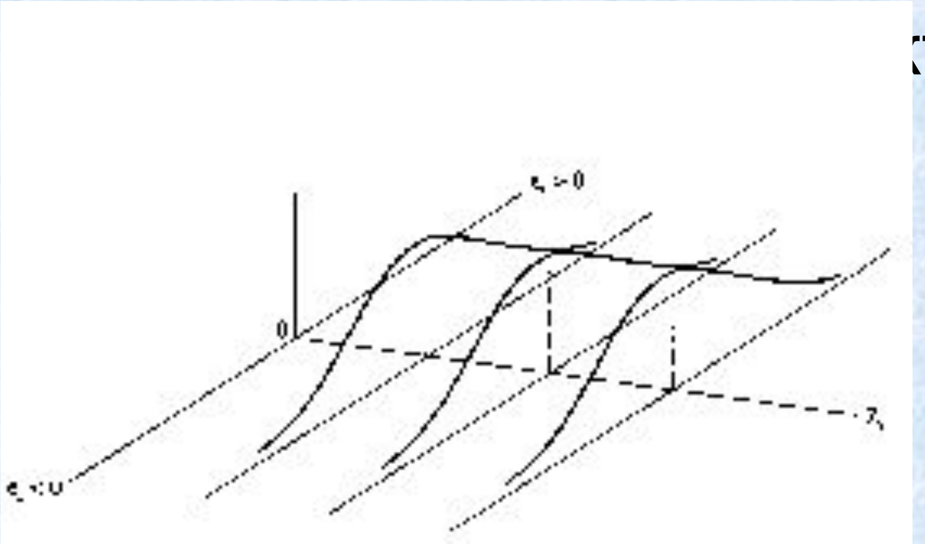
Источники гетероскедастичности – 1

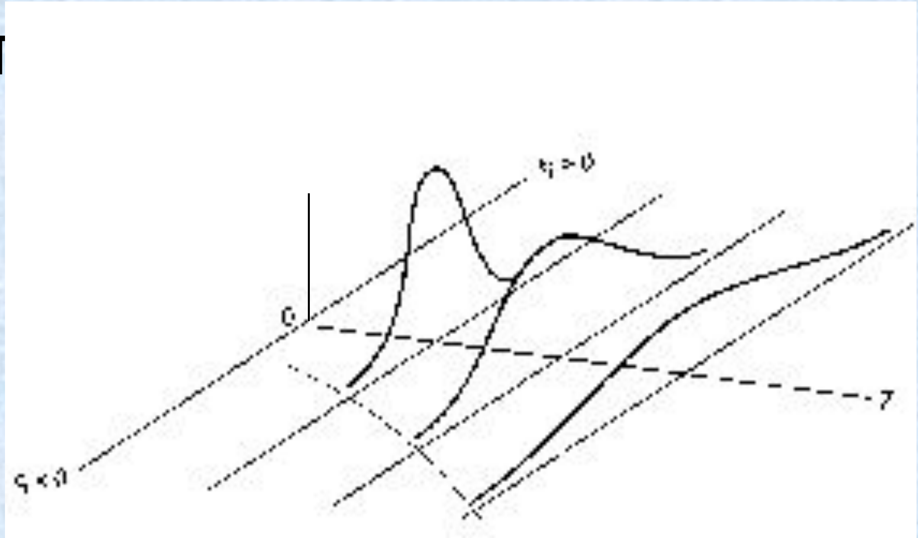
Истинная гетероскедастичность возникает в перекрестных выборках при зависимости масштаба изменений зависимой переменной от некоторой переменной, называемой фактором пропорциональности (Z).

Источники гетероскедастичности – 1

Наиболее распространенный случай истинной

гетероскедастичности – 1: дисперсия растет с





Источники гетероскедастичности – 2

Истинная гетероскедастичность возникает также и

во *временных рядах*, когда зависимая переменная

имеет большой интервал качественно

неоднородных значений или высокий темп

изменения (инфляция, технологические сдвиги,

изменения в законодательстве, потребительские

предпочтения и т.д.).

Гетероскедастичность простейшего вида

Мы в дальнейшем будем рассматривать, главным
образом, только гетероскедастичность простейшего
вида:

$$Var(\varepsilon_i) = \sigma_i^2 \cdot Z_i^2$$

СЛЕДСТВИЯ ГЕТЕРОСКЕДАСТИЧНОСТИ

1. Истинная гетероскедастичность не приводит к смещению оценок коэффициентов регрессии
2. Стандартные ошибки коэффициентов (вычисленные в предположении. гомоскедастичности) будут занижены. Это приведет к завышению *t*-статистик и даст неправильное (завышенное) представление о точности оценок.

ОБНАРУЖЕНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

Обнаружение гетероскедастичности в каждом конкретном

случае – довольно сложная задача.

Для знания $\sigma^2_{\varepsilon_i}$ необходимо знать распределение случайной

величины $Y/X=x_i$. На практике часто для каждого

конкретного значения x_i известно лишь одно y_i, что не

позв

$Y/$

Не существует какого-либо однозначного метода определения гетероскедастичности.

ОБНАРУЖЕНИЕ ГЕТЕРОСКЕДАСТИЧНОСТИ

Тесты:

1. Тест ранговой корреляции Спирмена.
2. Тест Парка.
3. Тест Глейзера.
4. Тест Голдфелда-Квандта.
5. Тест Уайта.
6. Тест Бреуша-Пагана.

ТЕСТ РАНГОВОЙ КОРРЕЛЯЦИИ СПИРМЕНА

При использовании данного теста предполагается, что дисперсии отклонений остатков будут монотонно изменяться (увеличиваться или уменьшаться) с увеличением фактора пропорциональности Z.

Поэтому значения e_i и z_i будут коррелированы (возможно, нелинейно!).

ТЕСТ РАНГОВОЙ КОРРЕЛЯЦИИ СПИРМЕНА. Алгоритм применения

1. Рассчитываются ранги (порядковые номера) значений фактора пропорциональности $z_i = x_{ik}$.
2. Рассчитывается уравнение

$$\hat{y}_i = b_0 + \sum_{j=1}^{m} b_j x_{ij}$$

и вычисляются остатки $e_i = y_i - \hat{y}_i, \quad i = \overline{1,n}$.
3. Рассчитываются ранги остатков e_i.

ТЕСТ РАНГОВОЙ КОРРЕЛЯЦИИ СПИРМЕНА. Алгоритм применения

4. Рассчитывается коэффициент ранговой корреляции Спирмена

$$\rho_{z/e} = 1 - \frac{6\sum D_i^2}{n(n^2-1)},$$

D_i – разность рангов *z* и *e*.

5. Рассчитывают статистику $u = \rho_{z/e} \cdot \sqrt{n-1}$, распределенную нормально $N(0,1)$ при отсутствии гетероскедастичности.

ТЕСТ ПАРКА

Здесь предполагается, что дисперсии $\sigma_{\varepsilon_i}^2$ связаны

с фактором пропорциональности *Z* в виде:

$$\sigma_{\varepsilon_i}^2 = \sigma^2 z_i^{\beta} e^{\eta_i} \quad \overset{\ln}{\Rightarrow} \quad \ln \sigma_{\varepsilon_i}^2 = \ln \sigma^2 + \beta \ln z_i + \eta_i$$

Т.к. дисперсии $\sigma_{\varepsilon_i}^2$ неизвестны, то их заменяют оценками квадратов отклонений e_i^2.

ТЕСТ ПАРКА.
Алгоритм применения

1. Строится уравнение регрессии: $\hat{y}_i = b_0 + \sum_{j=1}^{m} b_j x_{ij}$

и вычисляются остатки $e_i = y_i - \hat{y}_i, \quad i = \overline{1,n}$.

2. Выбирается фактор пропорциональности *Z* и оценивают вспомогательное уравнение регрессии:

$$\ln(e_i^2) = \alpha_0 + \alpha_1 \ln z_i + \eta_i, \quad i = \overline{1,n}$$

3. Проверяют значимость коэффициента при $\ln z_i$

ТЕСТ ГЛЕЙЗЕРА

Здесь предполагается, что дисперсии $\sigma_{\varepsilon_i}^2$ связаны

с фактором пропорциональности *Z* в виде:

$$\sigma_{\varepsilon_i} = \alpha + \beta z_i^{\gamma} + \eta_i$$

Т.к. средние квадратические отклонения σ_{ε_i} неизвестны, то их заменяют модулями оценок отклонений $|e_i|$.

ТЕСТ ГЛЕЙЗЕРА.
Алгоритм применения

1. Строится уравнение регрессии $\hat{y}_i = b_0 + \sum_{j=1}^{m} b_j x_{ij}$

и вычисляются остатки $e_i = y_i - \hat{y}_i, \quad i = \overline{1,n}$.

2. Выбирается фактор пропорциональности *Z* и оценивают вспомогательное уравнение регрессии: $|e_i| = \alpha_0 + \alpha_1 z_i^{\gamma} + \eta_i, \quad i = \overline{1,n}$

Изменяя *γ*, строят несколько моделей: $\gamma = \square\ , -1, -0{,}5, 0{,}5, 1, \square$

3. Статистическая значимость коэффициента α_1 в каждом случае

означает наличие гетероскедастичности.

4. Если для нескольких моделей будет получена значимая

оценка α_1, то характер гетероскедастичности определяют по

наиболее значимой из них.

ТЕСТЫ ПАРКА и ГЛЕЙЗЕРА. Выводы

Отметим, что как в тесте Парка, так и в тесте Глейзера для отклонений η_i может нарушаться условие гомоскедастичности.

Однако, во многих случаях используемые в тестах модели являются достаточно хорошими для определения гетероскедастичности.

ТЕСТ БРЕУШ-ПАГАНА

Тест применим в предположении, что:

Дисперсии $\sigma_{\varepsilon_i}^2$ зависят от некоторых дополнительных переменных $Z_j, \quad j = \overline{1, p}$:

$$\sigma_{\varepsilon_i}^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j Z_{ij}, \quad i = \overline{1, n}$$

ТЕСТ БРЕУШ-ПАГАНА. Алгоритм применения

1. Строится уравнение регрессии: $\hat{y}_i = b_0 + \sum_{j=1}^{m} b_j x_{ij}$

и вычисляются остатки: $e_i = y_i - \hat{y}_i, \quad i = \overline{1,n}$

2. Вычисляют оценку дисперсии остатков:

$$\tilde{\sigma}_e^2 = \frac{\sum e_i^2}{n}$$

3. Строят вспомогательное уравнение регрессии:

$$\frac{e_i^2}{\tilde{\sigma}_e^2} = \alpha_0 + \sum_{j=1}^{p} \alpha_j z_{ji} + \eta_i, \quad i = \overline{1,n}$$

ТЕСТ БРЕУШ-ПАГАНА. Алгоритм применения

4. Для вспомогательного уравнения регрессии определяют

объясненную часть вариации *RSS*.

5. Находим тестовую статистику:

$$BP = \frac{RSS}{2}$$

6. Если верна гипотеза H_0: гомоскедастичность остатков, то

статистика BP имеет распределение χ^2_p. Т.е. о наличии

гетероскедастичности остатков на уровне значимости α

свидетельствует: $BP > \chi^2_{\alpha;\,p}$

ТЕСТ БРЕУШ-ПАГАНА. Замечания

При $p = 1$ гетероскедастичность может быть скорректирована:

$$\hat{y}_i = b_0 + \sum_{j=1}^{m} b_j x_{ij} \quad \Rightarrow \quad \hat{y}_i = \frac{b_0}{z_{i1}} + \sum_{j=1}^{m} b_j \frac{x_{ij}}{z_{i1}}$$

При $p > 1$ не существует естественного преобразования, корректирующего гетероскедастичность

ТЕСТ ГОЛДФЕЛДА-КВАНДТА

<u>В этом тесте предполагается</u>:

1. Стандартные отклонения остатков σ_{ε_i} пропорциональны фактору пропорциональности *z*, т.е.

$$\sigma_{\varepsilon_i}^2 = \sigma^2 z_i^2, \quad i = \overline{1,n}$$

2. Случайный член *ε* имеет нормальное распределение и отсутствует автокорреляция остатков (предпосылка 3^0).

ТЕСТ ГОЛДФЕЛДА-КВАНДТА. Алгоритм применения

1. Выделяют фактор пропорциональности $Z = X_k$.
Данные упорядочиваются в порядке возрастания
величины Z.
2. Отбрасывают среднюю треть упорядоченных
наблюдений. Для первой и последней третей
строятся две отдельные регрессии, используя ту же
спецификацию модели регрессии.
3. Количество наблюдений в этих подвыборках
должно быть одинаково. Обозначим его l.

ТЕСТ ГОЛДФЕЛДА-КВАНДТА. Алгоритм применения

4. Берутся суммы квадратов остатков для регрессий по
первой трети RSS_1 и последней трети RSS_3. Рассчитывают
их отношение:

$$GQ = \frac{RSS_3}{RSS_1}$$

5. Используем *F*-тест для проверки гомоскедастичности.
Если статистика *GQ* удовлетворяет неравенству

$$GQ > F_{\gamma; l-m-1; l-m-1}$$

то гипотеза гомоскедастичности остатков отвергается

ТЕСТ ГОЛДФЕЛДА-КВАНДТА. Замечание

Тест Голдфелда-Квандта применим и для случая обратной пропорциональности:

$$\sigma_{\varepsilon_i}^2 = \frac{\sigma^2}{z_i^2}, \quad i = \overline{1,n}$$

При этом используется та же процедура, но тестовая

статистика равна:

$$GQ = \frac{RSS_1}{RSS_3}$$

ТЕСТ УАЙТА

Предполагается, что дисперсии $\sigma^2_{\varepsilon_i}$ связаны с объясняющими переменными $X_j, \ j=\overline{1,m}$ в виде:

$$\sigma^2_{\varepsilon_i} = f(X_{i1}, X_{i2}, \ldots, X_{im}) + \eta_i, \quad i=\overline{1,n}$$

где $f(\cdot)$ – квадратичная функция от аргументов.

Т.к. дисперсии $\sigma^2_{\varepsilon_i}$ неизвестны, то их заменяют оценками квадратов отклонений e_i^2.

ТЕСТ УАЙТА.
Алгоритм применения
(на примере трех переменных)

1. Строится уравнение регрессии:

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3}$$

и вычисляются остатки $e_i = y_i - \hat{y}_i, \quad i = \overline{1,n}$.

2. Оценивают вспомогательное уравнение регрессии:

$$e_i^2 = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \alpha_4 X_{i1}^2 + \alpha_5 X_{i2}^2 +$$

$$+ \alpha_6 X_{i3}^2 + \alpha_7 X_{i1} X_{i2} + \alpha_8 X_{i1} X_{i3} + \alpha_9 X_{i2} X_{i3} + \eta_i$$

ТЕСТ УАЙТА.
Алгоритм применения
(на примере трех переменных)

3. Определяют из вспомогательного уравнения тестовую статистику $U = nR^2$

4. Проверяют общую значимость уравнения с помощью критерия χ^2. Если

$$U > \chi^2_{\alpha;k}$$

то гипотеза гомоскедастичности отвергается. Число степеней свободы k равно числу объясняющих Переменных вспомогательного уравнения. В частности, Для рассматриваемого случая $k = 9$.

ТЕСТ УАЙТА.
Замечания

Тест Уайта является более общим чем тест Голдфелда-Квандта.

Неудобство использования теста Уайта:

Если отвергается нулевая гипотеза о наличии гомоскедастичности

$$H_0 : \ \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2,$$

то неясно, что делать дальше.

КОРРЕКЦИЯ ГЕТЕРОСКЕДАСТИЧНОСТИ

1. Использовать обобщенный метод наименьших квадратов.

2. Переопределить переменные.

3. Вычисление стандартных ошибок с поправкой на гетероскедастичность (метод Уайта).

ОБОБЩЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

При нарушении гомоскедастичности и наличии автокорреляции

остатков рекомендуется вместо традиционного МНК

использовать обобщенный МНК. Его для случая устранения

гетероскедастичности часто называют методом

на

Метод применим, если известны дисперсии $\sigma^2_{\varepsilon_i}$ для каждого наблюдения.

Основан на делении каждого наблюдаемого значения на соответствующее ему стандартное отклонение остатков.

МЕТОД ВЗВЕШЕННЫХ НАИМЕНЬШИХ КВАДРАТОВ. Случай парной регрессии

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \overset{\frac{\cdot}{\sigma_i}}{\Longrightarrow} \quad \frac{Y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{X_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}$$

$$\frac{Y_i}{\sigma_i} = Y_i^*, \frac{1}{\sigma_i} = Z_i, \frac{X_i}{\sigma_i} = X_i^*, \frac{\varepsilon_i}{\sigma_i} = \upsilon_i \quad \Rightarrow \quad Y_i^* = \beta_0 Z_i + \beta_1 X_i^* + \upsilon_i$$

Получили уравнение регрессии без свободного члена, но с дополнительной объясняющей переменной *Z* и с «преобразованным» остатком *v*. Можно показать, что для него выполняются предпосылки $1^0 - 5^0$ МНК.

МЕТОД ВЗВЕШЕННЫХ НАИМЕНЬШИХ КВАДРАТОВ. Случай парной регрессии

На практике, значения дисперсии остатков, как

правило, не известны. Для применения метода ВНК

необходимо сделать реалистичные предположения об
этих

значениях. Например:

Дисперсии $\sigma^2_{\varepsilon_i}$ пропорциональны X_i: $\sigma^2_{\varepsilon_i} = \sigma^2 x_i, \quad i = \overline{1,n}$

Дисперсии $\sigma^2_{\varepsilon_i}$ пропорциональны X_i^2: $\sigma^2_{\varepsilon_i} = \sigma^2 x_i^2, \quad i = \overline{1,n}$