

Тема 6. Гетероскедастичность.

1. Гетероскедастичность и ее последствия.
2. Обобщенный метод наименьших квадратов.
3. Проверка выборки на гомоскедастичность.

Гетероскедастичность

- это предположение о неоднородности дисперсий случайных ошибок модели регрессии.

Случайная ошибка модели регрессии - это величина отклонения в модели линейной множественной регрессии:

$$\varepsilon = y - \tilde{y}$$

где ε - остатки модели регрессии.

Гомоскедастичность

- это предположение о постоянстве дисперсии случайной ошибки *e* для всех *i* наблюдений модели регрессии.

В соответствии с третьей предпосылкой МНК требуется, чтобы дисперсия остатков была ***гомоскедастичной***.

Это значит, что для каждого значения фактора x_j остатки e_i имеют одинаковую дисперсию.

Если это условие применения МНК не соблюдается, то имеет место ***гетероскедастичность***.

Последствия гетероскедастичности остатков модели регрессии:

- 1) оценки нормальной линейной модели регрессии остаются несмещенными и состоятельными, но теряется эффективность;
- 2) появляется вероятность неверного вычисления оценок стандартных ошибок коэффициентов модели регрессии, что может привести к утверждению неверной гипотезы о значимости коэффициентов регрессии и значимости модели регрессии в целом. Обнаружить гетероскедастичность остатков модели регрессии можно путем проверки гипотез.

При малом объеме выборки, что наиболее характерно для эконометрических исследований, для оценки гетероскедастичности может использоваться *метод Гольдфельда-Квандта*.

Для того, чтобы оценить нарушение гомоскедастичности, необходимо провести *параметрический тест*, который включает в себя несколько этапов:

1 этап.

Упорядочение n наблюдений по мере возрастания переменной x .

2 этап.

Исключение из рассмотрения C центральных наблюдений;
при этом $(n-C):2 > p$, где p - число оцениваемых параметров.
Из экспериментальных расчетов, для случая одного фактора
рекомендовано при $n=30$ принимать $C=8$.

3 этап.

Разделение совокупности из $(n-C)$ наблюдений на две группы
(соответственно с малыми и большими значениями фактора x) и
определение по каждой из групп уравнений регрессии.

4 этап.

Определение остаточной суммы квадратов для первой (S_1) и второй (S_2) групп и нахождение их отношения: $R=S_1:S_2$, где $S_1 > S_2$.

При выполнении нулевой гипотезы о гомоскедастичности отношение R будет удовлетворять **F-критерию** с $(n-C-2p):2$ степенями свободы для каждой остаточной суммы квадратов.

Если $F_{\text{факт}} > F_{\text{теор}}$, то основная гипотеза отклоняется, и в основной модели регрессии присутствует гетероскедастичность, зависящая от факторной переменной x .

Если $F_{\text{факт}} < F_{\text{теор}}$, то основная гипотеза принимается, и гетероскедастичность в основной модели регрессии не зависит от факторной переменной x .

Чем больше величина R превышает табличное значение **F-критерия**, тем более нарушена предпосылка о равенстве дисперсий

Возможны варианты: если e_i зависит от y_x , то:

1. остатки e_i не случайны.
2. остатки e_i , не имеют постоянной дисперсии.
3. остатки e_i носят систематический характер в данном случае отрицательные значения e_i , соответствуют низким значениям y_x , а положительные — высоким значениям.

Коэффициент корреляции между e_i и e_j , где e_i — остатки текущих наблюдений, e_j — остатки предыдущих наблюдений, может быть определен по обычной формуле линейного коэффициента корреляции

$$r_{e_i e_j} = \frac{\text{cov}(e_i, e_j)}{\sigma_{e_i} \cdot \sigma_{e_j}}$$

Если этот коэффициент окажется существенно отличным от нуля, то остатки **автокоррелированы** и функция плотности вероятности $F(e)$ зависит от j -й точки наблюдения и от распределения значений остатков в других точках наблюдения.

Обобщенный МНК для корректировки гетероскедастичности.

В общем виде для уравнения

$$y_i = a + bx_i + e_i$$

$$\text{при } \sigma_{ei}^2 = \sigma^2 \cdot K_i$$

где K_i – коэффициент пропорциональности.

Модель примет вид:

$$y_i = \alpha + \beta x_i + \sqrt{K_i} e_i$$

В ней остаточные величины гетероскедастичны. Предполагая в них отсутствие автокорреляции, можно перейти к уравнению с гомоскедастичными остатками, поделив все переменные, зафиксированные в ходе i -го наблюдения на

$$\sqrt{K_i}$$

Тогда дисперсия остатков будет величиной постоянной. От регрессии y по x перейдем к регрессии на новых переменных: $y/\sqrt{K}$ и $x/\sqrt{K}$. Уравнение регрессии примет вид:

$$y_i / \sqrt{K_i} = \alpha / \sqrt{K_i} + \beta \cdot x_i / \sqrt{K_i} + e_i$$

По отношению к обычной регрессии уравнение с новыми, преобразованными переменными представляет собой взвешенную регрессию, в которой переменные y и x взяты с весами $1/\sqrt{K}$. Коэффициент регрессии b можно определить как:

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

При использовании обобщенного МНК с целью корректировки гетероскедастичности коэффициент регрессии b представляет собой взвешенную величину по отношению к обычному МНК с весами $1/K$. Аналогичный подход возможен не только для уравнения парной, но и для множественной регрессии. Модель примет вид:

$$y_i = a + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} + K_i \cdot e_i$$

Модель с преобразованными переменными составит:

$$y_i / K_i = a / K_i + b_1 \cdot x_{1i} / K_i + b_2 \cdot x_{2i} / K_i + e_i$$

Это уравнение не содержит свободного члена, и, применяя, обычный МНК получим:

$$y_i / K_i = A + b_1 \cdot x_{1i} / K_i + b_2 \cdot x_{2i} / K_i + e_i$$

Тест Глейзера обнаружения гетероскедастичности остатков модели регрессии

Тест Глейзера основывается на регрессии абсолютных значений остатков $|\varepsilon|$, т.е. рассматривается функция

$$|\varepsilon_i| = a + bx_i^c + u_i,$$

Регрессия $|\varepsilon_i|$ от x_i строится при разных значениях параметра c , и далее отбирается та функция, для которой коэффициент регрессии b оказывается наиболее значимым, т.е. имеет место наибольшее значение t -критерия Стьюдента или F -критерия Фишера и R^2 .

Устранение гетероскедастичности остатков модели регрессии

Автокорреляция остатков может быть вызвана следующими причинами:

- 1) Ошибками измерения при первоначальном сборе данных по результативному признаку;
- 2) Неправильно выбранная формулировка исходной модель.