Гетероскедастичность

Общие понятия

Постоянство дисперсии ошибок $\left(\sigma_u^2 = \text{constant}\right)$, независимо от наблюдения, носит название *гомоскедастичности.*

Смысл предположения гомоскедастичности состоит в том, что вариация каждой ошибки u_i около ее математического ожидания не зависит от значения независимой переменной x_k

$$\sigma_u^2 \neq f\left(x_{1k}, x_{2k}, \ldots, x_{nk}\right).$$

В практических исследованиях явление гомоскедастичности часто нарушается.

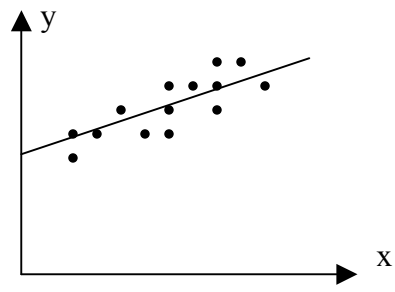
Гетероскедастичность – это нарушение классического предположения о постоянстве дисперсий ошибок, т.е.

$$\sigma_u^2 \neq \text{constant}$$

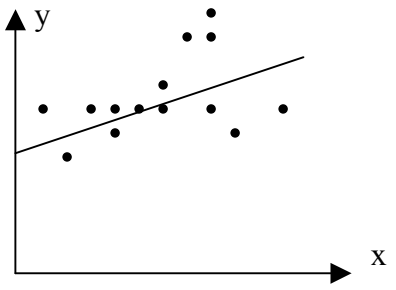
или иначе

$$\sigma_u^2 = f\left(x_{1k}, x_{2k}, \ldots, x_{nk}\right).$$

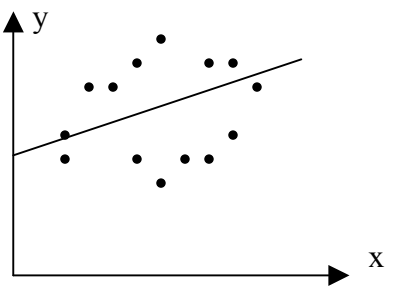
Графическая интерпретация гомо- и гетероскедастичности представлена следующим образом



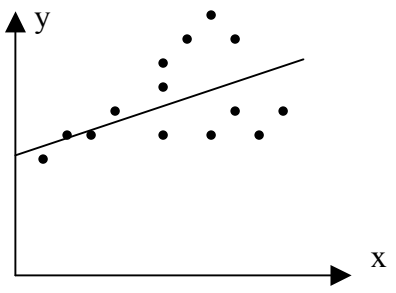
Гомоскедастичность



Гетероскедастичность



Гетероскедастичность

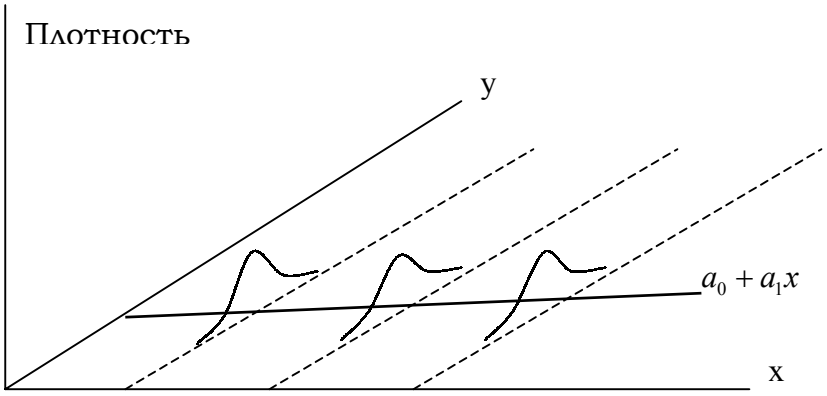


Гетероскедастичность

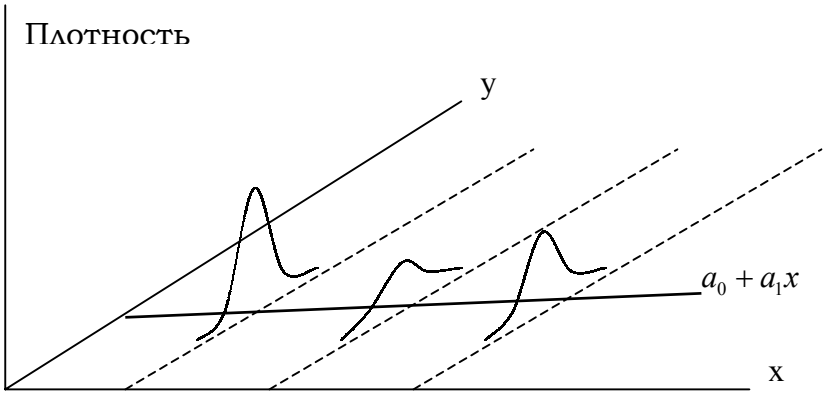
Приведенные иллюстрации в большей степени показывают природу гетероскедастичности, связанную с разбросом точек исходных данных.

Гетероскедастичность возникает чаще всего тогда, когда выборка берется в пространственном разрезе, когда имеют место большие различия между наименьшими и наибольшими значениями наблюдений, т.е. когда дисперсия значений наблюдений достаточно высока. Другой причиной гетероскедастичности является существенное изменение качества исходных данных внутри выборки.

В другом виде гомо- и гетероскедастичность иллюстрируется следующим образом:



Гомоскедастичность



Гетероскедастичность

Различают чистую гетероскедастичность, когда нарушение предположения о постоянстве дисперсий остатков возникает в корректно специфицированном уравнении регрессии, и смешанную (нечистую) гетероскедастичность, возникающую при неверной спецификации модели, в случае невключения в нее существенно влияющих переменных.

Величина ошибки u, как известно, аккумулирует в себе неточности измерений факторов, включенных в модель, влияние факторов, не включенных в модель, различия в природе наблюдений. В качестве примеров может быть использована производственная функция Кобба-Дугласа, которая учитывает только два фактора производства – труд и капитал, и не учитывает множества других; модель, описывающая зависимость уровня накопления (или потребления) от уровня доходов для различных групп населения и т.д.

Наличие гетероскедастичности не влияет на смещенность и обоснованность оценок модели, однако она затрагивает их эффективность. В связи с этим оценка дисперсии ошибок $\hat{\sigma}_u^2$ не может быть использована для проверки значимости параметров модели и расчета их доверительных интервалов.

Проверка наличия или отсутствия гетероскедастичности, как правило, не осуществляется, однако могут быть выдвинуты гипотезы относительно правдоподобности альтернативных допущений относительно пропорциональной зависимости ошибки и значений независимых переменных X.

Методы выявления гетероскедастичности

Так как гетероскедастичность принимает различные формы и ее точное проявление в модели почти никогда неизвестно, то и для выявления ее используются различные тесты и методы. Как результат – нет единого универсального метода для оценки данного явления, тем более что ни один из них не доказывает на 100% наличие гетероскедастичности. В литературе по эконометрии приводятся, по меньшей мере, десять различных методов для проверки наличия гетероскедастичности. К их числу относятся: анализ содержания проблемы, графический анализ, тест ранговой корреляции Спирмена, μ-критерий, параметрический и непараметрический тесты Гольдфельда-Квондта, тест

Глейсера, тест Парка, тест Бреуша-Пэйгана, тест Уайта и др. Рассмотрим некоторые из них.

Анализ существа проблемы

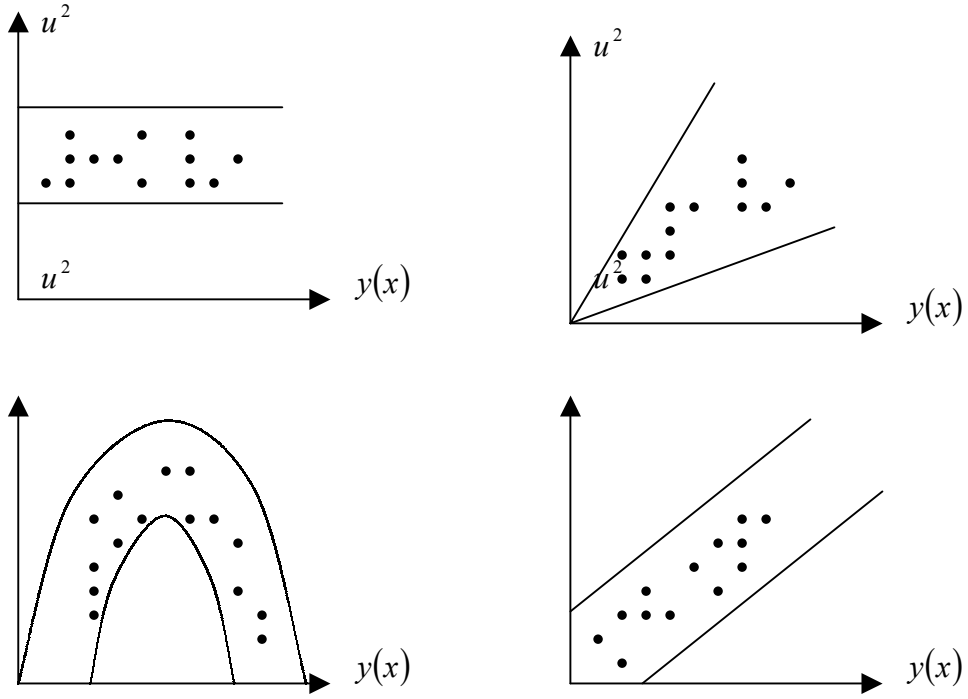
Часто наличие гетероскедастичности может быть выявлено интуитивно на стадии анализа исходной проблемы. Так, анализ бюджета любой организации или конкретной семьи может показать, что дисперсия ошибок растет пропорционально росту поступлений в бюджет (дохода). Анализ деятельности различных по размерам предприятий и фирм также может показать явную зависимость между изменением дисперсии ошибок и показателей деятельности хозяйственных объектов.

Графический анализ

Как уже было проиллюстрировано, одним из наглядных методов выявления гетероскедастичности является графический анализ. Он позволяет в отсутствие достаточного эмпирического материала сделать соответствующие выводы. Кроме того, графическая интерпретация проблемы позволяет исключить субъективность мнений исследователя.

Для проведения графического анализа необходимо рассчитать значения квадратов отклонений u_i^2 и затем определить, имеют ли они какую либо систематичность. Важно определить, действительно ли оцененное среднее значение y систематически связано с квадратом остатков. Форма связи между u_i^2 и y может быть различной, что очень важно выявить для трансформации исходных данных с целью построения модели с гомоскедастичностью ошибок.

На рисунках приведены случаи отсутствия гетероскедастичности (левый верхний) и ее наличие (остальные) с учетом гиперболической и линейной зависимостей (нижний левый и правый рисунки).



μ -критерий

Используется для проверки гетероскедастичности при большом числе наблюдений исходных данных. Алгоритм метода распадается на пять шагов.

1-й шаг. Все наблюдения зависимой переменной разбиваются на p групп $\left(r = \overline{1, p}\right)$ в соответствии с уровнем изменения величины y.

2-й шаг. Для каждой группы рассчитываются суммы квадратов отклонений

$$S_r = \sum_{i=1}^{n_r} \left(y_{ir} - \bar{y}_r\right)^2$$

3-й шаг. Находится общая сумма квадратов отклонений по всем группам

$$S = \sum_{r=1}^{p} S_r$$

4-й шаг. Вычисляется параметр w по формуле

$$w = \frac{\prod_{r=1}^{p} \left(\frac{S_r}{n_r}\right)^{n_r/2}}{\left(\frac{S}{n}\right)^{n/2}},$$

где n – общее число наблюдений;

n_r – число наблюдений r-1 группы.

5-й шаг. Рассчитывается значение критерия μ по формуле

$$\mu = -2\ln w,$$

который приближенно соответствует критерию χ^2 при числе степеней свободы $\nu = p - 1$, когда дисперсия всех наблюдений однородна.

Если $\mu \geq \chi^2$ при заданном уровне значимости α, то имеет место гетероскедастичность.

На основе данных из примера, рассмотренного для случая мультиколлинеарности, исследуем наличие гетероскедастичности с помощью μ-критерия.

Разобьем наблюдения на пять групп по пять наблюдений в каждой группе:

Группа 1	**Группа 2**	**Группа 3**	**Группа 4**	**Группа 5**
1,82	3,71	3,24	2,83	1,49
2,19	3,07	2,12	3,03	2,69
4,23	1,75	3,95	3,08	2,29
1,66	4,78	2,28	3,49	1,92
1,84	3,35	0,47	2,00	4,22

Найдем сумму квадратов отклонений индивидуальных значений каждой группы от своего среднего значения:

$$\bar{y}_1 = 2{,}384,\ \bar{y}_2 = 3{,}332,\ \bar{y}_3 = 2{,}412,\ \bar{y}_4 = 2{,}886,\ \bar{y}_5 = 2{,}522,$$

$$S_1 = \sum_{i=1}^{5}\left(y_{i1} - \bar{y}_1\right)^2 = 4{,}57708,$$

$$S_2 = \sum_{i=1}^{5}\left(y_{i2} - \bar{y}_2\right)^2 = 4{,}81128,$$

$$S_3 = \sum_{i=1}^{5}\left(y_{i3} - \bar{y}_3\right)^2 = 6{,}92508,$$

$$S_4 = \sum_{i=1}^{5}\left(y_{i4} - \bar{y}_4\right)^2 = 1{,}21132,$$

$$S_5 = \sum_{i=1}^{5}\left(y_{i5} - \bar{y}_5\right)^2 = 4{,}39268.$$

Рассчитаем общую сумму квадратов отклонений по пяти группам:

$$S = \sum_{r=1}^{5} S_r = 21{,}91744.$$

Определим параметр w:

$$w = \frac{\prod_{r=1}^{p}\left(\frac{S_r}{n_r}\right)^{n_r/2}}{\left(\frac{S}{n}\right)^{n/2}} =$$

$$= \frac{\left(\frac{4{,}57708}{5}\right)^{5/2}\cdot\left(\frac{4{,}81128}{5}\right)^{5/2}\cdot\left(\frac{6{,}92508}{5}\right)^{5/2}\cdot\left(\frac{1{,}21132}{5}\right)^{5/2}\cdot\left(\frac{4{,}39268}{5}\right)^{5/2}}{\left(\frac{21{,}91744}{25}\right)^{25/2}} = 0{,}177995.$$

Найдем критерий μ:

$$\mu = -2\ln w = -2\cdot(-1.7259989) = 3{,}541998\,.$$

Для числа степеней свободы $\nu = p - 1 = 5 - 1 = 4$ и уровня значимости $\alpha = 0{,}05$ находим табличное значение критерия $\chi^2 = 9{,}49$ и сравниваем с ним полученное значение критерия μ. Так как $\mu < \chi^2$, то делаем заключение об отсутствии гетероскедастичности для исследуемого набора данных.

Тест ранговой корреляции Спирмена

Данный простой тест может быть использован для выборок различных размеров. Проверка наличия гетероскедастичности осуществляется за три шага.

1-й шаг. На основе построенного уравнения регрессии рассчитываем величины остатков u_i.

2-й шаг. Ранжируем либо по возрастанию, либо по убыванию абсолютные значения остатков $|u_i|$ и значения независимой переменной x_j. Рассчитываем коэффициент ранговой корреляции Спирмена по формуле

$$r_S = 1 - 6\cdot\left[\frac{\sum_{i=1}^{n} d_i^2}{n\left(n^2-1\right)}\right],$$

где d_i – разность между рангами, которые приписываются двум характеристикам $(x_j$ и $u)$ i-го объекта,

n – количество ранжируемых объектов.

3-й шаг. Проверяем значимость коэффициента ранговой корреляции по критерию Стьюдента. С этой целью рассчитываем t-статистику Стьюдента по формуле

$$t = \frac{r_S\sqrt{n-2}}{\sqrt{1-r_S^2}}.$$

Сравниваем полученное значение с табличным значением критерия для числа степеней свободы $\nu = n-2$ и уровня значимости α. Если $t_{расч.} \geq t_{табл.}$, то принимается гипотеза о наличии гетероскедастичности остатков. Если $t_{расч.} < t_{табл.}$, то считается, что в модели имеет место гомоскедастичность.

Используем пример, рассмотренный в вопросе, связанном с мультиколлинеарностью, для оценки наличия гетероскедастичности остатков. Будем рассматривать независимую переменную x_1 – численность населения и величины остатков модели u.

Проранжируем по возрастанию указанные показатели:

Население, тыс. чел.	**Ранг**	**Модули остатков**	**Ранг**	d_i	d_i^2
940,8	1	0,01667	1	0	0
1071,8	2	0,63702	16	-14	196
1172,3	3	0,79428	18	-15	225
1192,5	4	0,69994	17	-13	169
1211,2	5	0,10498	5	0	0
1255,1	6	1,80037	25	-19	361
1288,6	7	0,21139	9	-2	4
1332,1	8	0,81888	19	-11	121
1333,4	9	1,65011	24	-15	225
1384,3	10	1,18054	20	-10	100
1465,6	11	0,09936	3	8	64
1467,9	12	0,56415	15	-3	9
1491,1	13	0,42480	12	1	1
1498,4	14	0,11667	6	8	64
1723,9	15	1,57790	23	-8	64
1862,2	16	0,08118	2	14	196
1880,4	17	0,37616	11	6	36
2059,2	18	0,44007	13	5	25
2566,8	19	0,16261	7	12	144

2580,2	20	0,20826	8	12	144
2743	21	0,10447	4	17	289
2750,6	22	0,53837	14	8	64
3055,1	23	0,25425	10	13	169
3811,2	24	1,36929	21	3	9
5125,4	25	1,38299	22	3	9
				$\sum$	2688

Коэффициент ранговой корреляции равен

$$r_S = 1 - 6 \cdot \left[\frac{2688}{25\left(25^2 - 1\right)} \right] = \text{-}0{,}033846154 .$$

Определим статистику Стьюдента

$$t = \frac{-0{,}0338\sqrt{25-2}}{\sqrt{1-\left(-0{,}0338\right)^2}} = -17{,}9251 .$$

Табличное значение критерия при 23 степенях свободы и уровне значимости 5% составляет 1,714. Сравнивая по абсолютной величине расчетное значение t-статистики с табличным, делаем вывод о наличии гетероскедастичности.

Тест Парка

В основе теста лежит предположение, что дисперсия остатков изменяется пропорционально значению некоторого фактора, т.е.

$$Var\left(u_i\right) = \sigma^2 Z_i^2 ,$$

где σ^2 – дисперсия гомоскедастической ошибки,

Z – фактор пропорциональности.

Прежде чем использовать тест Парка, необходимо ответить на три вопроса:

1. *Имеют ли место очевидные ошибки спецификации модели?* Если в уравнении регрессии отсутствует (еще не включена) существенно влияющая переменная или уравнение должно быть построено заново в силу ряда других причин, то использование теста Парка должно быть отложено до того момента, пока спецификация модели не будет как можно более хорошей.

2. *Насколько часто предмет исследований «страдает» гетероскедастичностью?* Выборке, взятой в пространственном разрезе, наиболее часто присуще явление гетероскедастичности в связи с часто возникающими большими разрывами в значениях зависимой переменной.
3. *Дает ли график отклонений какое-либо свидетельство наличия гетероскедастичности?* Иногда выгодно построить график изменения отклонений по отношению к какому-то потенциальному фактору пропорциональности Z.

Рассмотрим тест Парка, который распадается на три шага.

1-й шаг. Построение уравнения регрессии и расчет отклонений $u_i \ \left(i = \overline{1,n}\right)$

$$u_i = y_i - \widehat{a}_0 - \widehat{a}_1 x_1 - \ldots - \widehat{a}_m x_m .$$

2-й шаг. Использование значений отклонений для получения новой зависимой переменной. В качестве последней используются логарифмы квадратов отклонений. Независимая переменная – фактор пропорциональности (наиболее существенно влияющая переменная на основную переменную). Построение вторичного уравнения регрессии

$$\ln\left(u_i^2\right) = \alpha_0 + \alpha_1 \ln Z_i + \varepsilon_i .$$

В качестве фактора пропорциональности часто выбирают ту независимую переменную, которая имеет дисперсию, близкую к дисперсии ошибок.

3-й шаг. Проверка значимости коэффициента регрессии при Z с помощью критерия Стьюдента. Если коэффициент существенно отличается от нуля, то это является очевидным признаком наличия гетероскедастичности в отклонениях по отношению к независимой переменной Z; в противном случае, гетероскедастичность, относящаяся к данному конкретному Z, не подкрепляется с очевидностью в данных отклонениях. Тем не менее, невозможно полностью доказать, что ошибка уравнения регрессии обладает свойством гомоскедастичности.

В качестве **примера** для оценки гетероскедастичности с помощью теста Парка используем следующую задачу.

С целью выбора места строительства нового круглосуточно работающего кафе необходимо определить зависимость его дохода (y) *от трех факторов: количества прямых конкурентов, находящихся в радиусе двух ки-*

лометров от кафе (x_1)*, количества людей, живущих в радиусе трех километров от кафе* (x_2) *и среднегодового дохода семьи* (x_3)*, живущей в данной окрестности. Исходные данные для построения модели приведены в таблице*

Наблюдение	Доход кафе, грн.	Число конкурентов, ед.	Численность проживающих, чел.	Доход семьи, грн.
1	107919	3	65044	13240
2	118866	5	101376	22554
3	98579	7	124989	16916
4	122015	2	55249	20967
5	152827	3	73775	19576
6	91259	5	48484	15039
7	123550	8	138809	21857
8	160931	2	50244	26435
9	98496	6	104300	24024
10	108052	2	37852	14987
11	144788	3	66921	30902
12	164571	4	166332	31573
13	105564	3	61951	19001
14	102568	5	100441	20058
15	103324	2	39462	16194
16	127030	5	139900	21384
17	166755	6	171740	18800
18	125343	6	149894	15289
19	121886	3	57386	16702
20	134594	6	185105	19093
21	152937	3	114520	26502
22	109622	3	52933	18760
23	149884	5	203500	33242
24	98388	4	39334	14988
25	140791	3	95120	18505
26	101260	3	49200	16839
27	139517	4	113566	28915
28	115236	9	194125	19033
29	136749	7	233844	19200
30	105067	7	83416	22833
31	136872	6	183953	14409
32	117146	3	60457	20307
33	163538	2	65065	20111

1-й шаг. Построим уравнение множественной регрессии

$$\hat{y} = 102188 - 9074x_1 + 0{,}355x_2 + 1{,}288x_3 .$$

Показатели значимости модели и ее параметров следующие: $F = 15{,}65$, $t_1 = -4{,}42$, $t_2 = 4{,}88$, $t_3 = 2{,}37$, $\overline{R}^2 = 0{,}579$.

Рассчитаем остатки:

Наблюдаемое значение y	Предсказанное значение y	Остатки
107919	115087,8	-7168,786
118866	121821,6	-2955,611
98579	104785,9	-6206,934
122015	130640,6	-8625,642
152827	126345,3	26481,71
91259	93383,01	-2124,01
123550	106977,3	16572,75

160931	135908,4	25022,58
98496	115677,7	-17181,75
108052	116768,1	-8716,113
144788	138502,5	6285,495
164571	165550,4	-979,4206
105564	121411,1	-15847,08
102568	118275,1	-15707,1
103324	118893,8	-15569,77
127030	133977,9	-6947,881
166755	132868	33887,02
125343	120597,7	4745,348
121886	116830,9	5055,142
134594	137985,5	-3391,515
152937	149717,1	3219,912
109622	117902,3	-8280,264
149884	171808,1	-21924,14
98388	99146,41	-758,4093
140791	132536,2	8254,777
101260	114104	-12844
139517	143412,4	-3895,421
115236	113884,4	1351,572
136749	146335,2	-9586,22
105067	97662,51	7404,49
136872	131543,9	5328,148
117146	122563,4	-5417,357
163538	133019,5	30518,48

2-й шаг. Преобразуем остатки – рассчитаем логарифмы их квадратов. В качестве фактора пропорциональности Z возьмем размер рынка, обслуживаемого кафе, т.е. население $Z = x_2$. Найдем логарифмы фактора пропорциональности. Исходные данные для построения второй регрессионной модели представлены в таблице

u_i	u_i^2	$\ln\left(u_i^2\right)$	$\ln x_{2i}$
-7168,786	51391485,68	17,75498	11,08282
-2955,611	8735636,01	15,98292	11,52659
-6206,934	38526023,5	17,46684	11,73598
-8625,642	74401693,06	18,12499	10,91961
26481,71	701280919	20,36842	11,20878
-2124,01	4511420,475	15,32212	10,78899
16572,75	274655904,8	19,43103	11,84085
25022,58	626129613,6	20,25507	10,82465
-17181,75	295212383,8	19,50321	11,55503
-8716,113	75970631,17	18,14586	10,54144
6285,495	39507443,21	17,492	11,11127
-979,4206	959264,7604	13,77392	12,02174
-15847,08	251129881,2	19,34148	11,0341
-15707,1	246712934,4	19,32374	11,51733

-15569,77	242417729,1	19,30617	10,58309
-6947,881	48273043,93	17,69238	11,84868
33887,02	1148330143	20,86157	12,05374
4745,348	22518328,46	16,92984	11,91768
5055,142	25554458,54	17,05632	10,95756
-3391,515	11502375,9	16,25806	12,12868
3219,912	10367831,76	16,15422	11,6485
-8280,264	68562769,46	18,04326	10,87678
-21924,14	480668027,2	19,99069	12,22342
-758,4093	575184,6448	13,26245	10,57984
8254,777	68141348,59	18,03709	11,46289
-12844	164968331,8	18,92126	10,80365
-3895,421	15174303,91	16,53511	11,64014
1351,572	1826745,66	14,41805	12,17626
-9586,22	91895622,88	18,33616	12,36241
7404,49	54826477,02	17,81968	11,3316
5328,148	28389161,01	17,16152	12,12244
-5417,357	29347757,64	17,19473	11,00969
30518,48	931377408	20,65218	11,08314

Уравнение регрессии, построенное на основе данной информации, имеет вид

$$\ln\left(u^2\right) = 21{,}05 - 0{,}286 \ln x_2 .$$

Параметры качества уравнения следующие: $F = 0{,}209$, $t = -0{,}457$, $\overline{R}^2 = 0{,}0067$.

3-й шаг. Проверка значимости коэффициента регрессии в построенном уравнении при 5% уровне значимости и числе степеней свободы, равном 31 $\left(t_{табл.} = 2{,}042\right)$, показал, что нулевая гипотеза о гомоскедастичности остатков принимается.

Параметрический тест Гольдфельда-Квондта

Как и в предыдущих случаях, тест Гольдфельда-Квондта проверяет гипотезу $H_0 : u_i$ – гомоскедастичны против гипотезы $H_1 : u_i$ – гетероскедастичны (с возрастающей дисперсией). Тест используется для значительных по объему выборок. При этом ставится условие, чтобы число наблюдений было, по крайней мере, в два раза больше числа переменных. Для проведения теста необходимо выполнить пять шагов.

1-й шаг. Ранжируем наблюдения в порядке возрастания значений независимой переменной x. Если переменная не одна, то либо выбирают наиболее су-

щественную с точки зрения постановки задачи, либо попеременно используют тест для каждой из переменных.

2-й шаг. Выбираем C центральных наблюдений переменной и исключаем их из выборки. Число C обычно принимают равным от одной четвертой до одной трети общего числа наблюдений. Остаток наблюдений делится на две подвыборки, первая из которых состоит из наименьших значений переменной, вторая – из наибольших.

3-й шаг. Строим две эконометрические модели на основе каждой из подвыборок, содержащих по $(n-C)/2$ наблюдений.

4-й шаг. Рассчитываем суммы квадратов ошибок

$$S_1 = \sum u_1^2$$

и

$$S_2 = \sum u_2^2$$

для каждой из моделей, где u_1 – ошибки, соответствующие первой модели, u_2 – ошибки, соответствующие второй модели.

5-й шаг. Рассчитываем значение критерия

$$F^* = \frac{S_2}{S_1},$$

который в случае выполнения гипотезы о гомоскедастичности соответствует F-распределению с числом степеней свободы $\nu_1 = \nu_2 = [(n-C)/2] - k$ и уровнем значимости α. Рассчитанное значение критерия сравнивается с теоретическим (табличным) и в случае, когда $F^* \le F_{табл.}$, то гипотеза $H_0 : u_i$ о гомоскедастичности величин u_i принимается, т.е. гетероскедастичность отсутствует.

Можно заметить, что если имеет место гомоскедастичность, то дисперсии и для первой, и для второй модели совпадут, и значение критерия F^* будет равно единице.

В качестве **примера** снова возьмем задачу по оценке зависимости уровня доходов кафе от числа конкурентов, численности населения и среднегодового дохода семьи. В качестве основного фактора будем снова использовать численность населения.

Проранжируем выборку по возрастанию численности проживающих

Наблюдение	Доход кафе, грн.	Число конкурентов, ед.	Численность проживающих, чел.	Доход семьи, грн.
10	108052	2	37852	14987
24	98388	4	39334	14988
15	103324	2	39462	16194
6	91259	5	48484	15039
26	101260	3	49200	16839
8	160931	2	50244	26435
22	109622	3	52933	18760
4	122015	2	55249	20967
19	121886	3	57386	16702
32	117146	3	60457	20307
13	105564	3	61951	19001
1	107919	3	65044	13240
33	163538	2	65065	20111
11	144788	3	66921	30902
5	152827	3	73775	19576
30	105067	7	83416	22833
25	140791	3	95120	18505
14	102568	5	100441	20058
2	118866	5	101376	22554
9	98496	6	104300	24024
27	139517	4	113566	28915
21	152937	3	114520	26502
3	98579	7	124989	16916
7	123550	8	138809	21857
16	127030	5	139900	21384
18	125343	6	149894	15289
12	164571	4	166332	31573
17	166755	6	171740	18800
31	136872	6	183953	14409
20	134594	6	185105	19093
28	115236	9	194125	19033
23	149884	5	203500	33242
29	136749	7	233844	19200

Выберем число $C = 9$, что позволит нам исключить девять центральных наблюдений, оставив две выборки по $\frac{33-9}{2} = 12$ наблюдений

Наблюдение	Доход кафе, грн.	Число конкурентов, ед.	Численность проживающих, чел.	Доход семьи, грн.
10	108052	2	37852	14987
24	98388	4	39334	14988
15	103324	2	39462	16194
6	91259	5	48484	15039
26	101260	3	49200	16839
8	160931	2	50244	26435
22	109622	3	52933	18760
4	122015	2	55249	20967
19	121886	3	57386	16702
32	117146	3	60457	20307
13	105564	3	61951	19001
1	107919	3	65044	13240

21	152937	3	114520	26502
3	98579	7	124989	16916
7	123550	8	138809	21857
16	127030	5	139900	21384
18	125343	6	149894	15289
12	164571	4	166332	31573
17	166755	6	171740	18800
31	136872	6	183953	14409
20	134594	6	185105	19093
28	115236	9	194125	19033
23	149884	5	203500	33242
29	136749	7	233844	19200

Для полученных выборок строим уравнения регрессии и находим остатки, после чего рассчитываем их суммы квадратов.

Для первого уравнения имеем:

Доход кафе (наблюдаемые значения)	**Доход кафе (предсказанные значения)**	**Остатки**	**Квадраты остатков**
108052	105064,6477	2987,352279	8924273,641
98388	94935,84469	3452,155306	11917376,25
103324	109581,2784	-6257,278383	39153532,76
91259	91309,57658	-50,57658276	2557,990724
101260	108147,1782	-6887,178172	47433223,17
160931	147472,3824	13458,61759	181134387,5
109622	115510,9168	-5888,916781	34679340,85
122015	128852,6893	-6837,689349	46753995,64
121886	108887,7396	12998,26036	168954772,4
117146	122117,5448	-4971,544784	24716257,54
105564	117715,0525	-12151,05248	147648076,4
107919	97771,14901	10147,85099	102978879,8
		Сумма квадратов остатков $S_1 = \sum u_1^2$	**814296674**

Для второго уравнения:

Доход кафе (наблюдаемые значения)	**Доход кафе (предсказанные значения)**	**Остатки**	**Квадраты остатков**
152937	151368,3231	1568,676943	2460747,352
98579	115153,0984	-16574,09845	274700739,3
123550	112708,731	10841,26895	117533112,5
127030	137871,758	-10841,758	117543716,6
125343	128580,4067	-3237,406666	10480801,92

164571	158019,5706	6551,429435	42921227,64
166755	135659,4428	31095,55723	966933679,6
136872	136198,2259	673,7741099	453971,5512
134594	138992,1854	-4398,185371	19344034,55
115236	115944,4171	-708,4170899	501854,7733
149884	159361,3923	-9477,392304	89820964,87
136749	142242,4488	-5493,448795	30177979,66
	Сумма квадратов остатков $S_2 = \sum u_2^2$		**1672872830**

Рассчитываем значение критерия F^*:

$$F^* = \frac{S_2}{S_1} = \frac{1672872830}{814296674} = 2{,}05437758\,.$$

Сравниваем полученное значение с табличным значением F-распределения для числа степеней свободы

$$\nu_1 = \nu_2 = \left[(n - C)/2\right] - k = \left[(33 - 9)/2\right] - 4 = 8$$

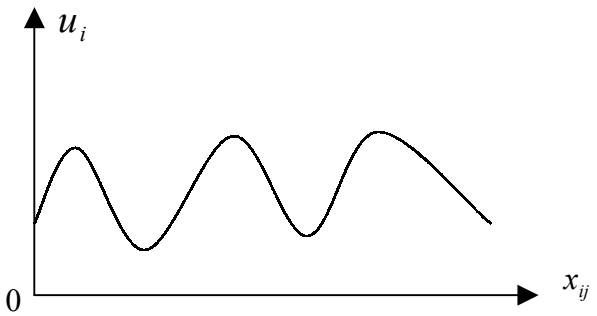
и уровня значимости $\alpha = 0{,}05$ $\left(F_{табл.} = 3{,}44\right)$.

Так как $F^* < F_{табл.}$, то делаем вывод о наличии гомоскедастичности ошибок модели, что созвучно с тестом Парка.

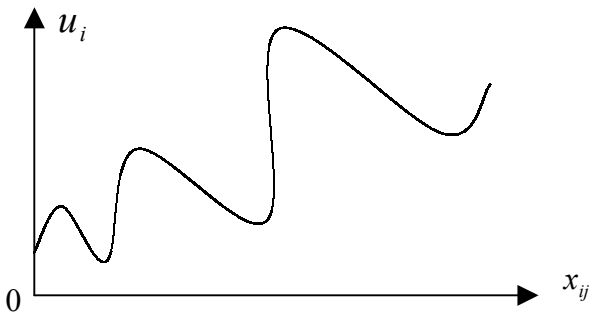
Непараметрический тест Гольдфельда-Квондта

В основе данного теста лежит оценка числа вершин величины остатков, получаемых после упорядочения наблюдений переменной x_j. Оценка осуществляется визуально путем анализа графика изменения остатков u_i при изменении значений переменной x_j.

Случай гомоскедастичности может быть описан следующим графиком изменения остатков, имеющих постоянную дисперсию:



Гетероскедастичность проявляет себя таким образом, что дисперсия остатков меняется:



Данный тест отличается своей простотой, однако он не так надежен, как остальные.

Тест Глейсера

Данный тест основан на построении и оценке уравнений регрессии, в которых в качестве зависимой переменной выступают абсолютные значения остатков $|u_i|$, а в качестве независимой – фактор пропорциональности (величина или независимая переменная, которая соответствует изменению дисперсии σ_u^2).

1-й шаг. Рассчитываются параметры уравнения регрессии и находятся величины отклонений u_i.

2-й шаг. Строятся уравнения регрессии, увязывающие модули остатков и фактор пропорциональности. При этом могут использоваться различные виды зависимостей:

$$|u| = \alpha_0 + \alpha_1 x + \varepsilon ,$$

$$|u| = \alpha_0 + \alpha_1 x^{-1} + \varepsilon ,$$

$$|u| = \alpha_0 + \alpha_1 x^{\frac{1}{2}} + \varepsilon ,$$

$$|u| = \sqrt{\alpha_0 + \alpha_1 x} + \varepsilon ,$$

$$|u| = \sqrt{\alpha_0 + \alpha_1 x^2} + \varepsilon$$

и другие.

Выбирается та модель, которая наиболее точно описывает связь между рассматриваемыми величинами, учитывая коэффициент корреляции и среднеквадратические ошибки параметров α_0 и α_1. Проверяется статистическая значимость α_0 и α_1. На этой основе делается вывод о наличии или отсутствии гетероскедастичности остатков u_i.

Если оба параметра α_0 и α_1 являются значимыми (т.е. $\alpha_0 \neq 0$ и $\alpha_1 \neq 0$), то имеет место смешанная гетероскедастичность. Если $\alpha_0 = 0$, а $\alpha_1 \neq 0$, то – чистая.

В отличие от ранее рассмотренных тестов тест Глейсера позволяет выявить не только наличие гетероскедастичности и определить ее вид.

Тест Бреуша-Пэйгана

Данный тест осуществляет попытку определить гетероскедастичность путем оценки общей значимости вторичного уравнения регрессии, построенного на основе квадратов отклонений (зависимой переменной) и сразу нескольких факторов пропорциональности (переменных, определяющих изменение дисперсии остатков). Таким образом, для оценки гетероскедастичности остатков многофакторной модели нет надобности поочередно рассматривать все или многие независимые переменные, как факторы пропорциональности, рассчитывая уравнения отдельно для каждого случая.

1-й шаг. Находим отклонения на основе построенного уравнения регрессии.

2-й шаг. Используем квадраты отклонений в качестве зависимой переменной. В качестве независимых переменных используются те из основной модели, которые влияют на вариацию отклонений

$$u^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_j + \varepsilon ,$$

где p – число факторов, определяющих вариацию отклонений (в качестве таковых могут использоваться и те, которые не попали в основное уравнение регрессии).

3-й шаг. Проверяем статистическую значимость полученного уравнения, формулируя гипотезы:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0,$$
$$H_1 : \alpha_1 \neq 0, \alpha_2 \neq 0, \ldots, \alpha_p \neq 0.$$

С этой целью рассчитывается статистика

$$L = \frac{SSR}{2\left[\sum_{i=1}^{n} \frac{u_i^2}{n}\right]^2},$$

где SSR – сумма квадратов, объясняющая регрессию, из вторичного уравнения регрессии.

Для больших по размеру выборок L имеет χ^2-распределение с числом степеней свободы, равным p, и уровнем значимости α.

Если

$$L > \chi^2_{табл.},$$

то нулевая гипотеза отвергается и уравнение считается значимым, а, следовательно, делается вывод о наличии гетероскедастичности.

Преимуществом теста Бреуша-Пэйгана является то, что он использует более одного фактора пропорциональности. Но, с другой стороны, проблема выбора этих факторов все же остается.

Тест Уайта

Данный тест, также как и предыдущий, использует в качестве зависимой переменной во вторичном уравнении регрессии квадраты отклонений. В качестве независимых переменных (факторов пропорциональности) выступают все исходные независимые переменные, их квадраты и попарные произведения. Тест Уайта не предполагает выдвижения никаких гипотез относительно формы гетероскедастичности.

1-й шаг. Находим отклонения наблюдаемых значений зависимой переменной от расчетных.

2-й шаг. Используем квадраты отклонений в качестве независимой переменной и все независимые переменные исходной модели, их квадраты и попарные произведения в качестве независимых. Строим вторичное уравнение регрессии:

$$u^2 = \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_m x_m + \\ + \beta_1 x_1^2 + \ldots + \beta_m x_m^2 + \\ + \gamma_1 x_1 x_2 + \ldots + \gamma_m x_{m-1} x_m + \varepsilon.$$

3-й шаг. Проверяем общую значимость уравнения с помощью критерия χ^2. Для этого рассчитываем статистику nR^2, где R^2 – нескорректированный коэффициент детерминации, которая имеет χ^2-распределение с числом степеней свободы, равным числу угловых коэффициентов модели, и уровнем значимости α.

Если

$$nR^2 > \chi^2_{табл.},$$

то гипотеза об отсутствии гетероскедастичности остатков отвергается.

Проблемой использования теста Уайта является то, что при построении вторичного уравнения регрессии число вновь образованных независимых переменных (факторов пропорциональности) может оказаться больше числа наблюдений и, следовательно, данное уравнение не может быть оценено.

Обобщенный метод наименьших квадратов (метод Эйткена)

Для оценки параметров обобщенной эконометрической модели, т.е. модели, которой свойственна гетероскедастичность остатков, используется обобщенный метод наименьших квадратов.

Если модель описывается уравнением

$$Y = X\hat{A} + u$$

и имеет дисперсию остатков, которая описывается выражением

$$Var(u) = \sigma_u^2 Z,$$

где фактор пропорциональности Z представлен в виде симметричной, положительно определенной матрицы, то для нахождения оценок параметров модели $\hat{A}$, необходимо произвести корректировку исходной информации с учетом указанной матрицы.

Матрица Z может быть представлена следующим образом:

$$Z = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\lambda_3} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{pmatrix}.$$

Структура матрицы определяется исходя из того, что наличие гетероскедастичности связано с изменением дисперсии остатков, в то время как между ними отсутствует ковариация. Диагональные элементы определяют пропорции изменения дисперсий в зависимости от наблюдения объясняющей переменной x.

Значения величин λ_i определяются в зависимости от того, в какой форме фактор пропорциональности представлен в выражении для дисперсии остатков. В частности, если $Z = x_j$ и

$$Var(u) = \sigma_u^2 x_j,$$

то

$$\lambda_i = \frac{1}{x_{ij}},$$

если $Z = x_j^2$ и

$$Var(u) = \sigma_u^2 x_j^2,$$

то

$$\lambda_i = \frac{1}{x_{ij}^2}$$

и т.д.

Так как матрица Z является симметричной и положительно определенной, то ее можно представить в виде произведения

$$Z = P^{/} P,$$

где

$$P=\begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sqrt{\lambda_3}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}, \quad P^{-1}=\begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & \cdots & 0 \\ 0 & 0 & \sqrt{\lambda_3} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix},$$

причем матрица P – невырожденная.

Используя соотношение $Z = P^{/}P$ и вытекающие из него

$$P^{-1}Z\left(P^{-1}\right)^{/} = E$$

и

$$\left(P^{-1}\right)^{/}P^{-1} = Z^{-1},$$

преобразуем исходное уравнение регрессии.

Умножим его на матрицу P^{-1}:

$$P^{-1}Y = P^{-1}X\widehat{A} + P^{-1}u$$

и введем новые обозначения

$$Y^{*} = P^{-1}Y,$$

$$X^{*} = P^{-1}X,$$

$$u^{*} = P^{-1}u.$$

В результате преобразованная модель примет вид:

$$Y^{*} = X^{*}\widehat{A} + u^{*}.$$

Используя выражение $P^{-1}Z\left(P^{-1}\right)^{/} = E$, можно показать, что

$$Var\left(u^{*}\right) = \sigma^{2}_{u^{*}}E,$$

то есть полученная модель удовлетворяет условию гомоскедастичности и для нахождения ее параметров может быть применен метод наименьших квадратов (1МНК).

Итак, имеем

$$\widehat{A} = \left[\left(X^{*}\right)^{/}X^{*}\right]^{-1}\left(X^{*}\right)^{/}Y^{*} = \left(X^{/}Z^{-1}X\right)^{-1}X^{/}Z^{-1}Y.$$

Данные оценки являются несмещенными и имеют наименьшую дисперсию. Дисперсионно-ковариационная матрица в этом случае рассчитывается следующим образом:

$$Var\left(\widehat{A}\right)=\widehat{\sigma}_u^2\left[\left(X^*\right)' X^*\right]^{-1}=\widehat{\sigma}_u^2\left(X' Z^{-1} X\right)^{-1}.$$

Несмещенная оценка дисперсии остатков находится из выражения:

$$\widehat{\sigma}_u^2=\frac{\left(Y^*-X^*\widehat{A}\right)'\left(Y^*-X^*\widehat{A}\right)}{n-k}=\frac{(Y-X\widehat{A})' Z^{-1}\left(Y-X\widehat{A}\right)}{n-k}=$$

$$=\frac{u' Z^{-1} u}{n-k}.$$

Оценка параметров $\widehat{A}$ является оценкой обобщенного метода наименьших квадратов (метода Эйткена).

Разложим общую сумму квадратов на сумму квадратов регрессии и ошибок

$$Y' Z^{-1} Y=\widehat{A}' X' Z^{-1} Y+u' Z^{-1} u\,.$$

Отсюда можно рассчитать общую дисперсию

$$\sigma_{общ.}^2=\frac{Y' Z^{-1} Y}{n-1},$$

дисперсию, объясняемую регрессией

$$\sigma_{рег.}^2=\frac{\widehat{A}' X' Z^{-1} Y}{k-1},$$

и дисперсию ошибок

$$y_{ош.}^2=\frac{u' Z^{-1} u}{n-k}.$$

Возможно и другое преобразование для расчета оценок уравнения регрессии.

Если имеет место система соотношений

$$Y=X\widehat{A}+u\,,$$

$$M(u)=0\,,$$

$$Var(u)=V\,,$$

где $V=\sigma_u^2 Z$ – известная симметричная положительно определенная матрица, то вектор $\widehat{A}$ рассчитывается как

$$\hat{A} = \left(X' V^{-1} X\right)^{-1} X' V^{-1} Y ,$$

а дисперсионно-ковариационная матрица определяется из выражения

$$Var\left(\hat{A}\right) = \hat{\sigma}_u^2 \left(X' V^{-1} X\right)^{-1} .$$

Рассмотрим **пример**, для которого сначала проверим наличие гетероскедастичности остатков методом Гольдфельда–Квондта, а затем сопоставим расчеты, произведенные методом 1МНК с расчетами, осуществленными обобщенным методом наименьших квадратов.

Итак, имеем данные о затратах на питание (y) и общих затратах (x).

Наблюдение	Затраты на питание	Общие затраты
1	3,3	14
2	3,2	15
3	3	15
4	3,2	17
5	3,1	17
6	3,3	17
7	3,4	18
8	3,5	19
9	3,2	20
10	4,1	21
11	3,5	22
12	3,8	39
13	4	55
14	3,7	72
15	4,9	80
16	4,1	85
17	4,95	90

Выделим $S = 5$ центральных наблюдений и образуем две подвыборки, для которых рассчитаем уравнения регрессии и суммы квадратов отклонений

Наблюдение	Затраты на питание	Общие затраты	Наблюдение	Затраты на питание	Общие затраты
1	3,3	14			
2	3,2	15			
3	3	15	12	3,8	39
4	3,2	17	13	4	55
5	3,1	17	14	3,7	72
6	3,3	17	15	4,9	80
			16	4,1	85
			17	4,95	90

Расчеты для первой подвыборки:

Регрессионная статистика	
Множественный R	0,021452116
R-квадрат	0,000460193
Нормированный R-квадрат	-0,249424758
Стандартная ошибка	0,130673148
Наблюдения	6

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	3,14465E-05	3,14465E-05	0,001841621	0,967826762
Остаток	4	0,068301887	0,017075472		
Итого	5	0,068333333			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	3,213207547	0,698180772	4,60225729	0,010013849	1,274742946	5,151672149
Переменная X 1	-0,001886792	0,043966718	-0,042914108	0,967826762	-0,123958224	0,120184639

ВЫВОД ОСТАТКА			
Наблюдение	*Предсказанное Y*	*Остатки*	*Квадраты остатков*
1	3,186792453	0,113207547	0,012815949
2	3,18490566	0,01509434	0,000227839
3	3,18490566	-0,18490566	0,034190103
4	3,181132075	0,018867925	0,000355999
5	3,181132075	-0,081132075	0,006582414
6	3,181132075	0,118867925	0,014129583
		$S_1 =$	**0,068301887**

Расчеты для второй подвыборки:

Регрессионная статистика	
Множественный R	0,649971203
R-квадрат	0,422462564
Нормированный R-квадрат	0,278078205
Стандартная ошибка	0,465700912
Наблюдения	6

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	**Значимость F**
Регрессия	1	0,634573977	0,634573977	2,925957958	0,162337446
Остаток	4	0,867509357	0,216877339		
Итого	5	1,502083333			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	2,964326747	0,770567731	3,846938598	0,018350474	0,824883311	5,103770184
Переменная X 1	0,018204369	0,01064245	1,710543176	0,162337446	-0,011343869	0,047752607

ВЫВОД ОСТАТКА			
Наблюдение	*Предсказанное Y*	*Остатки*	*Квадраты остатков*
1	3,674297154	0,125702846	0,015801206
2	3,965567064	0,034432936	0,001185627
3	4,275041344	-0,575041344	0,330672547
4	4,420676299	0,479323701	0,22975121
5	4,511698146	-0,411698146	0,169495363
6	4,602719993	0,347280007	0,120603403
		$S_2 =$	**0,867509357**

Отношение

$$F^* = \frac{S_2}{S_1} = \frac{0{,}867509357}{0{,}068301887} = 12{,}7011$$

свидетельствует о наличии гетероскедастичности остатков, так как $F_{табл.} = 6{,}39$.

Если применить метод 1МНК, то получим следующие результаты расчетов

Регрессионная статистика	
Множественный R	0,84293208
R-квадрат	0,710534491
Нормированный R-квадрат	0,691236791
Стандартная ошибка	0,325421219
Наблюдения	17

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	3,899162512	3,899162512	36,8196454	2,15247E-05
Остаток	15	1,588484547	0,10589897		
Итого	16	5,487647059			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	3,025773574	0,131205637	23,06130774	3,9702E-13	2,746115207	3,305431942
Переменная X 1	0,017551703	0,002892541	6,06791936	2,1525E-05	0,011386395	0,023717012

ВЫВОД ОСТАТКА		
Наблюдение	*Предсказанное Y*	*Остатки*
1	3,271497421	0,028502579
2	3,289049124	-0,089049124
3	3,289049124	-0,289049124
4	3,324152531	-0,124152531
5	3,324152531	-0,224152531
6	3,324152531	-0,024152531
7	3,341704234	0,058295766
8	3,359255937	0,140744063
9	3,37680764	-0,17680764
10	3,394359344	0,705640656

11	3,411911047	0,088088953
12	3,710290003	0,089709997
13	3,991117256	0,008882744
14	4,289496212	-0,589496212
15	4,429909839	0,470090161
16	4,517668355	-0,417668355
17	4,605426872	0,344573128

Полученные оценки уравнения регрессии $\widehat{a}_0 = 3{,}025775374$ и $\widehat{a}_1 = 0{,}017551703$ нельзя считать эффективными из-за наличия гетероскедастичности остатков. В связи с этим применим метод Эйткена для расчета эффективных оценок. Выдвинем гипотезу, в соответствии с которой $\lambda_{ij} = \frac{1}{x_{ij}}$.

Тогда матрица Z^{-1} будет выглядеть следующим образом:

0,07143	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0,06667	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0,06667	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0,05882	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0,05882	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0,05882	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0,05556	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0,05263	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0,05	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,04762	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0,04545	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0,02564	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0,01818	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0,01389	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0125	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,01176	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,01111

Используем оператор

$$\widehat{A} = \left(X^{/} Z^{-1} X\right)^{-1} X^{/} Z^{-1} Y .$$

Найдем произведение $X^{/} Z^{-1}$

$$X^{/} Z^{-1} = \begin{pmatrix} 0{,}0714 & 0{,}0667 & 0{,}0667 & 0{,}0588 & 0{,}0588 & 0{,}0588 & 0{,}0556 & 0{,}0526 & 0{,}05 & 0{,}0476 & 0{,}0455 & 0{,}0256 & 0{,}0182 & 0{,}0139 & 0{,}0125 & 0{,}0118 & 0{,}0111 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Рассчитаем произведение $X^{/} Z^{-1} X$

$$X^{/} Z^{-1} X = \begin{pmatrix} 0{,}72558 & 17 \\ 17 & 616 \end{pmatrix}.$$

Обратим полученное выражение

$$\left(X^{/} Z^{-1} X\right)^{-1} = \begin{pmatrix} 3{,}89978 & -0{,}1076 \\ -0{,}1076 & 0{,}00459 \end{pmatrix}.$$

Найдем произведение $X^{/} Z^{-1} Y$

$$X^{/}Z^{-1}Y = \begin{pmatrix} 2{,}48722 \\ 62{,}25 \end{pmatrix}.$$

Теперь можно рассчитать вектор оценок $\widehat{A}$:

$$\widehat{A} = \begin{pmatrix} 3{,}00002 \\ 0{,}01826 \end{pmatrix}.$$

Полученное уравнение регрессии имеет вид:

$$\widehat{Y} = 3{,}00002 + 0{,}01826X \ .$$

По сравнению с полученным ранее уравнением

$$\widehat{Y} = 3{,}025773574 + 0{,}017551703X \ .$$

оценки нового уравнения считаются более эффективными.