Параметрический тест Гольдфельда-Квондта

- Тест Гольдфельда-Квондта проверяет гипотезу

$H_0 : u_i$ – остатки гомоскедастичны

против гипотезы

$H_1 : u_i$ – гетероскедастичны (с возрастающей дисперсией)

- 1-й шаг

  Ранжируем наблюдения в порядке возрастания значений независимой переменной x_j.

- <u>2-й шаг</u>

  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 отклонений

$$S_1 = \sum u_1^2$$

$$S_2 = \sum u_2^2$$

- <u>5-й шаг</u>

Рассчитываем значение критерия

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

(соответствует *F*-распределению с числом степеней свободы

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

и уровнем значимости α)

Если $F^* \le F_{табл.}$, то гипотеза $H_0 : u_i$ о гомоскедастичности величин u_i принимается

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

$$C = 9$$

Исключим 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

Строим уравнения регрессии, находим отклонения,

рассчитываем их суммы квадратов

Доход кафе (наблюдаемые значения)	Доход кафе (предсказанные значения)	Отклонения	Квадраты отклонений
108052	105064,6	2987,3	8924273,6
98388	94935,8	3452,1	11917376,2
103324	109581,2	-6257,2	39153532,7
91259	91309,5	-50,5	2557,9
101260	108147,1	-6887,1	47433223,1
160931	147472,3	13458,6	181134387,5
109622	115510,9	-5888,9	34679340,8
122015	128852,6	-6837,6	46753995,6
121886	108887,7	12998,2	168954772,4
117146	122117,5	-4971,5	24716257,5
105564	117715,0	-12151,0	147648076,4
107919	97771,1	10147,8	102978879,8
		Сумма квадратов отклонений	$S_1 = \sum u_1^2$ **814296674**

Доход кафе (наблюдаемые значения)	**Доход кафе (предсказанные значения)**	**Отклонения**	**Квадраты отклонений**
152937	151368,32	1568,67	2460747,3
98579	115153,09	-16574,09	274700739,3
123550	112708,73	10841,26	117533112,5
127030	137871,75	-10841,75	117543716,6
125343	128580,40	-3237,40	10480801,9
164571	158019,57	6551,42	42921227,6
166755	135659,44	31095,55	966933679,6
136872	136198,22	673,77	453971,5
134594	138992,18	-4398,18	19344034,5
115236	115944,41	-708,41	501854,7
149884	159361,39	-9477,39	89820964,8
136749	142242,44	-5493,44	30177979,6
		Сумма квадратов отклонений	$S_2 = \sum u_2^2$ **1672872830**

$$F^* = \frac{S_2}{S_1} = \frac{1672872830}{814296674} = 2{,}05$$

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

$$\alpha = 0{,}05$$

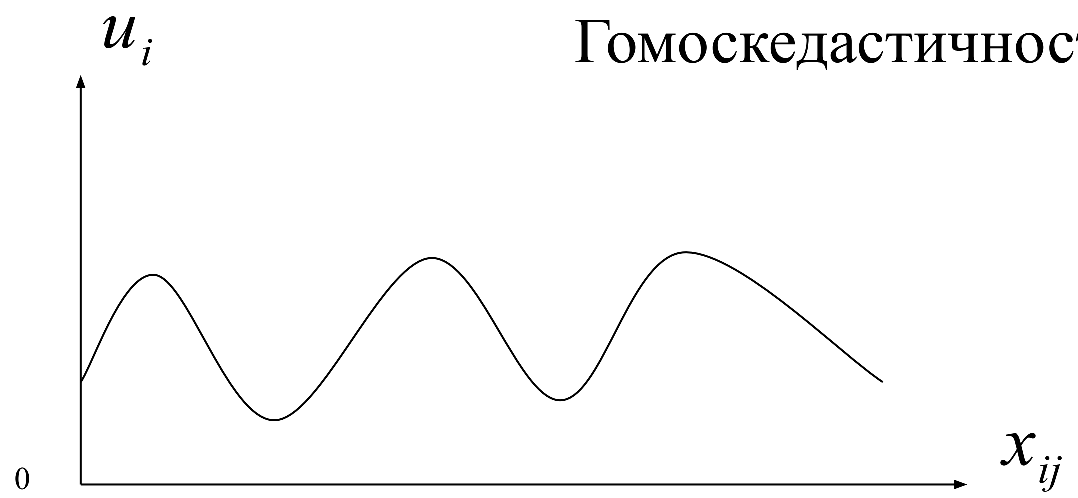
$$(F_{табл.} = 3{,}44)$$

Вывод: имеет место гомоскедастичность отклонений модели

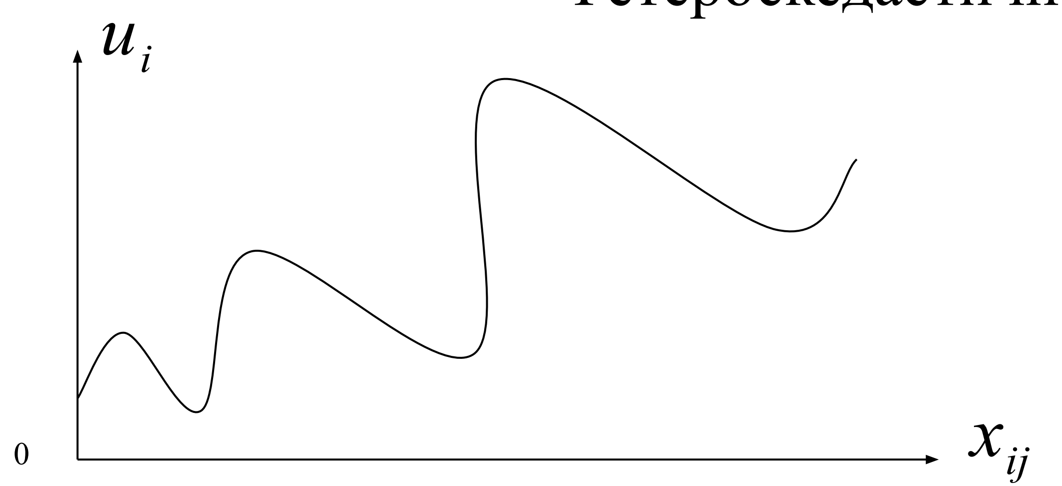
Непараметрический тест Гольдфельда-Квондта

- В основе 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.

Гомоскедастичность



Гетероскедастичность



Тест Глейсера

- 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

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

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

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

Выбирается та модель, которая наиболее точно описывает связь между рассматриваемыми величинами

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

Тест Бреуша-Пэйгана

- Осуществляет попытку определить гетероскедастичность путем оценки общей значимости вторичного уравнения регрессии, построенного на основе квадратов отклонений зависимой переменной и сразу нескольких факторов пропорциональности

- 1-й шаг

  Находим отклонения на основе построенного уравнения регрессии

2-й шаг

Строим вторичное уравнение регрессии.

Зависимая переменная – квадраты отклонений. Независимые переменные – те из основной модели, которые влияют на вариацию отклонений

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

p – число факторов, определяющих вариацию отклонений

- 3-й шаг

Проверяем статистическую значимость уравнения, формулируя гипотезы

$$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0,$$

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

Рассчитываем статистику

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

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

Если

$$L > \chi^2_{табл.} \quad ,$$

то нулевая гипотеза отвергается и

уравнение считается значимым.

Т.о. делается вывод о наличии

гетероскедастичности

Тест Уайта

- В качестве независимых переменных (факторов пропорциональности) выступают все исходные независимые переменные, их квадраты и попарные произведения

- <u>1-й шаг</u>

  Находим отклонения наблюдаемых значений зависимой переменной от расчетных

- 2-й шаг

Строим вторичное уравнение регрессии

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

- 3-й шаг

Проверяем общую значимость уравнения с помощью критерия χ^2

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

Обобщенный метод наименьших квадратов (метод Эйткена)

Пусть модель описывается уравнением

$$Y = XA + u$$

и имеет дисперсию остатков, которая описывается выражением

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

Фактор пропорциональности Z представлен в виде симметричной, положительно определенной матрицы

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

Диагональные элементы определяют пропорции изменения дисперсий в зависимости от наблюдения объясняющей переменной x

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

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

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

$$P = P^{/} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sqrt{\lambda_3}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix} \qquad P^{-1} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & \cdots & 0 \\ 0 & 0 & \sqrt{\lambda_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix}$$

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

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

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

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

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

$$Y^{*} = X^{*}\overset{\sqcup}{A} + u^{*}$$

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

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

Вектор оценок модели рассчитывается так:

$$\tilde{A} = \left[\left(X^{*} \right)' X^{*} \right]^{-1} \left(X^{*} \right)' Y^{*} = \left(X' Z^{-1} X \right)^{-1} X' Z^{-1} Y$$

Дисперсионно-ковариационная матрица

$$Var\left(\tilde{A} \right) = \tilde{\sigma}_{u}^{2} \left[\left(X^{*} \right)' X^{*} \right]^{-1} = \sigma_{u}^{2} \left(X' Z^{-1} X \right)^{-1}$$

Несмещенная оценка дисперсии остатков

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

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

Разложим общую сумму квадратов на сумму квадратов регрессии и остатков

$$Y^{/}Z^{-1}Y = \hat{A}^{/}X^{/}Z^{-1}Y + u^{/}Z^{-1}u$$

Рассчитаем общую дисперсию

$$\sigma^2_{общ.} = \frac{Y^{/}Z^{-1}Y}{n-1},$$

дисперсию, объясняемую регрессией

$$\sigma^2_{рег.} = \frac{\hat{A}^{/}X^{/}Z^{-1}Y}{k-1}$$

и дисперсию остатков

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

Если имеет место система соотношений

$$Var(u) = V$$

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

$$M(u) = 0$$

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

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

а дисперсионно-ковариационная матрица определяется из выражения

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

Имеются данные о затратах на питание и общих затратах

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

$$S = 5$$

Наблюден ие	Затраты на питание	Общие затраты	Наблюден ие	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	**0,021**
R-квадрат	**0,00046**
Нормированный R-квадрат	**-0,249**
Стандартная ошибка	**0,131**
Наблюдения	**6**

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	**1**	**0,00003**	**0,00003**	**0,0018**	**0,9678**
Остаток	**4**	**0,0683**	**0,0171**		
Итого	**5**	**0,0683**			

	Коэф-ты	*Стандартная ошибка*	*t-стат.*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересеч.	**3,213**	**0,698**	**4,602**	**0,01**	**1,275**	**5,152**
Перем. X 1	**-0,00189**	**0,044**	**-0,0429**	**0,968**	**-0,124**	**0,12**

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Y*	*Остатки*	*Квадраты остатков*
1	**3,187**	**0,113**	**0,013**
2	**3,185**	**0,015**	**0,0002**
3	**3,185**	**-0,185**	**0,0342**
4	**3,181**	**0,0189**	**0,00036**
5	**3,181**	**-0,081**	**0,0066**
6	**3,181**	**0,1189**	**0,0141**
		$S_1 =$	**0,0683**

Расчеты для второй подвыборки

Регрессионная статистика	
Множественный R	**0,65**
R-квадрат	**0,42**
Нормированный R-квадрат	**0,278**
Стандартная ошибка	**0,466**
Наблюдения	**6**

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	**1**	**0,635**	**0,635**	**2,926**	**0,162**
Остаток	**4**	**0,8675**	**0,217**		
Итого	**5**	**1,502**			

	Коэф-ты	*Стандартная ошибка*	*t-стат.*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересеч.	**2,964**	**0,771**	**3,847**	**0,018**	**0,825**	**5,104**
Перем. X 1	**0,0182**	**0,011**	**1,711**	**0,162**	**-0,011**	**0,0478**

ВЫВОД ОСТАТКА			
Наблюдение	*Предсказанное Y*	*Остатки*	*Квадраты остатков*
1	**3,674**	**0,1257**	**0,0158**
2	**3,966**	**0,0344**	**0,0012**
3	**4,275**	**-0,5750**	**0,3307**
4	**4,421**	**0,4793**	**0,2298**
5	**4,512**	**-0,4117**	**0,1695**
6	**4,603**	**0,3473**	**0,1206**
		$S_2 =$	**0,8675**

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

$$F_{табл.} = 6,39$$

Имеет место гетероскедастичность

Применим метод 1МНК

Регрессионная статистика	
Множественный R	**0,843**
R-квадрат	**0,711**
Нормированный R-квадрат	**0,691**
Стандартная ошибка	**0,325**
Наблюдения	**17**

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	**1**	**3,899**	**3,899**	**36,82**	**0,00002**
Остаток	**4**	**1,588**	**0,1059**		
Итого	**5**	**5,4876**			

	Коэф-ты	*Стандартная ошибка*	*t-стат.*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересеч.	**3,026**	**0,131**	**23,06**	**3,9702E-13**	**2,746**	**3,3054**
Перем. X 1	**0,0176**	**0,0029**	**6,068**	**0,00002**	**0,0114**	**0,0237**

ВЫВОД ОСТАТКА

Наблюдение	Предсказанное Y	Остатки
1	3,27	0,028
2	3,28	-0,089
3	3,28	-0,289
4	3,32	-0,12
5	3,32	-0,22
6	3,32	-0,024
7	3,34	0,058
8	3,35	0,14
9	3,37	-0,176
10	3,39	0,705
11	3,41	0,088
12	3,71	0,089
13	3,99	0,008
14	4,28	-0,58
15	4,42	0,47
16	4,51	-0,41
17	4,60	0,34

0,07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0,07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0,07	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0,06	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0,06	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0,06	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0,06	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	0,05	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	0,05	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0,03	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0,02	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0,01	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,01	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,01	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,01

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 $X^{/}Z^{-1}Y$

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

Рассчитаем вектор оценок $\hat{A}$

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

Уравнение регрессии $\hat{Y} = 3{,}00002 + 0{,}01826X$

По сравнению с полученным ранее уравнением

$$\hat{Y} = 3{,}02577357 + 0{,}0175513X$$

оценки нового уравнения считаются более эффективными