

Гетероскедастичность

1) Доугерти, 3-е издание, 7.1, с.236

В таблице приведены данные о государственных расходах G , инвестициях I , валовом внутреннем продукте Y и численности населения P для 30 стран в 1997 г. Исследователь рассматривает вопрос о том, происходит ли «вытеснение» инвестиций государственными расходами, оценивает регрессию:

$$\hat{I} = 18,10 - 1,07 G + 0,36 Y; R^2 = 0,99.$$

(7,79) (0,14) (0,02)

Исследователь упорядочивает наблюдения по увеличению Y и оценивает регрессии снова для 11 стран с наименьшим уровнем дохода и для 11 стран с наибольшим уровнем дохода. Величины RSS для этих регрессий равны 321 и 28101 соответственно. Выполните тест Голдфелда-Квандта на гетероскедастичность.

Решение:

Модель, которую оценивает исследователь, имеет следующий вид:

$$I_i = \beta_1 + \beta_2 G_i + \beta_3 Y_i + u_i; \quad i=1,2,\dots,30, \quad (0)$$

Тест Голдфелда-Квандта проверяет следующие гипотезы:

H_0 : гомоскедастичность ($\sigma_{u_i}^2 = \sigma_u^2$),

H_1 : гетероскедастичность ($\sigma_{u_i}^2 \sim Y_i$, $i=1,2,\dots,30$, т.е. стандартное отклонение распределения вероятностей случайного члена в наблюдении i пропорционально значению Y_i).

$$F = \frac{RSS_2 / (n_2 - k)}{RSS_1 / (n_1 - k)} = \frac{RSS_2}{RSS_1} = \frac{28101}{321} = 87,5$$

$$F_{0,05}^{cr}(8,8) = 3,44$$

$F > F^{cr} \Rightarrow H_0$ отвергается. В модели присутствует гетероскедастичность, стандартное отклонение распределения вероятностей случайного члена в наблюдении i пропорционально значению Y_i .

2) Доугерти, 3-е издание, 7.7, с.245

Ответить на вопросы и дать интерпретацию полученным результатам.

Исследователь из предыдущей задачи оценивает следующие регрессии в качестве альтернативных спецификаций модели:

$$\frac{\hat{I}}{P} = -0,03 \frac{1}{P} - 0,69 \frac{G}{P} + 0,34 \frac{Y}{P}; \quad R^2 = 0,97; \quad (1)$$

(0,28) (0,16) (0,03)

$$\frac{\hat{I}}{Y} = 0,39 + 0,03 \frac{1}{Y} - 0,93 \frac{G}{Y}; \quad R^2 = 0,78; \quad (2)$$

(0,04) (0,42) (0,22)

$$\log \hat{I} = -2,44 - 0,63 \log G + 1,60 \log Y; \quad R^2 = 0,98. \quad (3)$$

(0,26) (0,12) (0,12)

Он упорядочивает выборку по Y/P , G/Y и $\log Y$ соответственно и в каждом случае оценивает регрессии для 11 самых больших и 11 самых малых значений переменной упорядочения. Суммы квадратов отклонений показаны в таблице:

	11 самых малых значений переменной	11 самых больших значений переменной
(1)	1,43	12,63
(2)	0,0223	0,0155
(3)	0,573	0,155

Выполните тест Голдфелда-Квандта для каждой спецификации модели и интерпретируйте достоинства каждой спецификации. Подтверждается ли то, что инвестиции являются убывающей функцией от государственных расходов?

Решение:

Напомним, что модель, которую оценивает исследователь, имеет следующий вид:

$$I_i = \beta_1 + \beta_2 G_i + \beta_3 Y_i + u_i; \quad i=1,2,\dots,30 \quad (0).$$

Рассмотрим модель (1) зависимости инвестиций от государственных расходов и ВВП на душу населения. Коэффициенты в этой модели значимы за исключением коэффициента перед $\frac{1}{P}$.

Выполним тест Голдфелда-Квандта для этой модели:

$$F = \frac{RSS_2 / (n_2 - k)}{RSS_1 / (n_1 - k)} = \frac{RSS_2}{RSS_1} = \frac{12,63}{1,43} = 8,83$$

$$F_{0,05}^{cr}(8,8) = 3,44$$

$F > F^{cr} \Rightarrow H_0$ отвергается. Проблема гетероскедастичности осталась.

Рассмотрим теперь модель (2). Тест Голдфелда-Квандта:

$$F = \frac{RSS_2 / (n_2 - k)}{RSS_1 / (n_1 - k)} = \frac{RSS_2}{RSS_1} = \frac{0,0155}{0,0223} = 0,695.$$

$$F_{0,05}^{cr}(8,8) = 3,44$$

$F < F^{cr} \Rightarrow H_0$ не отвергается. Теперь гетероскедастичности в модели нет.

Сравним результаты оценивания регрессии для модели (0) и модели (2):

Из модели (0) $\hat{\beta}_2 = 1,7; \quad \hat{\beta}_3 = 0,36;$
 $s.e.(\hat{\beta}_2) = 0,14; \quad s.e.(\hat{\beta}_3) = 0,02.$

Из модели (2) $\hat{\beta}_2 = -0,93; \quad \hat{\beta}_3 = 0,39;$
 $s.e.(\hat{\beta}_2) = 0,22; \quad s.e.(\hat{\beta}_3) = 0,04.$

Оценки из модели (2) являются несмещенными оценками β_2 и β_3 в модели (0) (гетероскедастичность не приводит к смещению оценок), поэтому существенных изменений в значениях коэффициентов не наблюдается. Оценки модели значимы, однако коэффициент детерминации стал значительно ниже.

В модели (3) гетероскедастичности нет: $F = \frac{RSS_2 / (n_2 - k)}{RSS_1 / (n_1 - k)} = \frac{RSS_2}{RSS_1} = \frac{0,155}{0,573} = 0,27.$

$F_{0,05}^{cr}(8,8) = 3,44, F < F^{cr} \Rightarrow H_0$ не отвергается.

Все коэффициенты значимы. В логарифмической модели коэффициенты перед переменными трактуются как эластичности (отношение процентного изменения одной переменной по отношению к процентному изменению другой переменной). Инвестиции растут пропорционально государственным расходам и валовому внутреннему продукту. Эта модель имеет более высокий коэффициент детерминации, чем модель (2), поэтому для нас она более предпочтительна.

- 3) Предположим, что для модели парной регрессии $\sigma_{ui}^2 = \sigma^2 X_i^4$.
Как избавиться от проблемы гетероскедастичности ошибок?

Решение:

Дисперсия ошибок пропорциональна X_i^4 , а следовательно стандартная ошибка пропорциональна X_i^2 ($\sigma_{ui}^2 \sim X_i^4 \Rightarrow \sigma_{ui} \sim X_i^2$). Решением проблемы будет замена переменных:

$$Y^* = \begin{pmatrix} \frac{Y_1}{X_1^2} \\ \vdots \\ \frac{Y_n}{X_n^2} \end{pmatrix}; \quad 1^* = \begin{pmatrix} \frac{1}{X_1^2} \\ \vdots \\ \frac{1}{X_n^2} \end{pmatrix}; \quad X^* = \begin{pmatrix} \frac{1}{X_1} \\ \vdots \\ \frac{1}{X_n} \end{pmatrix},$$

т.е. оценить модель

$$Y_i^* = \beta_1^* 1^* + \beta_2^* X^* + u_i^*$$

или

$$\frac{Y}{X^2} = \tilde{\beta}_1 \frac{1}{X^2} + \tilde{\beta}_2 \frac{1}{X} + \tilde{u}_i.$$

- 4) По данным с 1946 г. по 1975 г. Hanushek и Jackson оценили коэффициенты уравнений регрессии (в скобках указаны оценки стандартных отклонений)

$$\hat{C}_t = 26.19 + 0.6248 GNP_t - 0.4398 D_t$$

(2.73) (0.006) (0.0736)

$$\left(\frac{\hat{C}}{GNP} \right)_t = 25.92 \frac{1}{GNP_t} + 0.6246 - 0.4315 \left(\frac{D}{GNP} \right)_t$$

(2.22) (0.00597) (0.0597)

где C – агрегированные частные потребительские расходы,
GNP (Gross National Product) – валовой национальный продукт,
 D – национальные расходы на оборону.

С какой целью оценили второе уравнение? Какое при этом было сделано предположение о дисперсии ошибок? Можно ли сравнивать R^2 в двух регрессиях? Ответ обоснуйте. Дайте интерпретацию полученным результатам.

Решение:

Вероятно, с помощью теста Голдфелда-Квандта была выявлена гетероскедастичность по переменной GNP. Предполагалось, что дисперсия ошибок пропорциональна переменной GNP: $\sigma_t \sim GNP_t$.

R^2 – это доля дисперсии зависимых переменных. В этих двух моделях зависимые переменные разные, поэтому нельзя сравнивать ни R^2 , ни скорректированный R^2 . Исследователями оценены модели:

$$C_t = \beta_1 + \beta_2 GNP_t + \beta_3 D_t + u_t, \quad (1)$$

$$\left(\frac{C}{GNP} \right)_t = \beta_1 \frac{1}{GNP_t} + \beta_2 - \beta_3 \left(\frac{D}{GNP} \right)_t. \quad (2)$$

Результаты оценивания модели (1): $\hat{\beta}_1 = 26,19$; $\hat{\beta}_2 = 0,6248$; $\hat{\beta}_3 = -0,4398$;
 $s.e.(\hat{\beta}_1) = 2,73$; $s.e.(\hat{\beta}_2) = 0,006$. $s.e.(\hat{\beta}_3) = 0,0736$.

Результаты оценивания модели (2): $\hat{\beta}_1 = 25,92$; $\hat{\beta}_2 = 0,6246$; $\hat{\beta}_3 = -0,4315$;
 $s.e.(\hat{\beta}_1) = 2,22$; $s.e.(\hat{\beta}_2) = 0,00597$. $s.e.(\hat{\beta}_3) = 0,0597$.

Как видно, оценки коэффициентов и их стандартные отклонения в этих моделях близки по значению, чего и следовало ожидать, ведь гетероскедастичность не приводит к смещению оценки.

Модель (2) оценивает зависимость агрегированных частных потребительских расходов в расчете на одну единицу валового национального продукта (разумеется, в денежном выражении) от обратной величины от валового национального продукта и национальных расходов на оборону в расчете на одну единицу валового национального продукта.

5) По данным для 39 районов Балтимора в 1970 г. были оценены уравнения

$$\ln \hat{Y}_i = 10.093 - 0.239 X_i, \quad R^2 = 0.803$$

$t=54.7 \quad t=-12.28$

$$\frac{\ln \hat{Y}_i}{\sqrt{X_i}} = 9.093 \frac{1}{\sqrt{X_i}} - 0.2258 \sqrt{X_i},$$

$t=47.87 \quad t=-15.10$

где Y – плотность населения района,

X – расстояние до центрального делового квартала.

С какой целью оценили второе уравнение? Какое при этом было сделано предположение о дисперсии ошибок? Дайте интерпретацию полученным результатам.

Решается аналогично задаче 4.

Подсказка: Предположение о дисперсии ошибок: $\sigma_i \sim \sqrt{X_i} \Rightarrow \sigma_i^2 \sim X_i$.

6) По данным для 20 стран были оценены коэффициенты уравнения регрессии

$$\hat{Y}_i = 111.78 - 0.0042 X_{2i} - 0.4898 X_{3i}, \quad R^2 = 0.492,$$

$t=4.79 \quad t=-2.53 \quad t=-1.71$

где Y – младенческая смертность (количество в расчете на тысячу рожденных живыми),

X_2 – GNP в расчете на душу населения,

X_3 – процент имеющих начальное образование в определенной возрастной группе.

При проведении теста Уайта была оценена регрессия

$$e_i^2 = 4987 - 0.4718 X_{2i} - 0.8442 X_{3i} + 0.00001 X_{2i}^2 + 0.4435 X_{3i}^2 + 0.0026 X_{2i} X_{3i}, \quad R^2 = 0.649.$$

$t=4.86 \quad t=-0.59 \quad t=-2.45 \quad t=1.27 \quad t=1.62 \quad t=0.35$

Проведя тест Уайта, проверьте гипотезу об отсутствии гетероскедастичности.

Решение:

Выполним тест Уайта.

Гипотезы:

Но: гомоскедастичность,

H1: гетероскедастичность.

Статистика Уайта: $\chi^2 = n \cdot R^2 = 20 \cdot 0,649 = 12,98$, где R^2 – это коэффициент детерминации во вспомогательной регрессии.

$$\chi_{0,05,cr}^2(5) = 11,07.$$

$\chi^2 > \chi_{0,05,cr}^2 \Rightarrow$ в модели присутствует гетероскедастичность.