

Семинар 8. Решение.

1) По месячным данным с 01.2001 по 06.2003 были оценены три регрессии:

$$Y = 18 - 0.427P + 0.000007I \quad (1)$$

(0.23) (0.006) (0.00001)

$$\ln Y = 4.7 - 0.096P + 0.000001I \quad (2)$$

(0.15) (0.004) (0.000007)

$$\ln Y = 12 - 3.11 \ln P + 0.0317 \ln I \quad (3)$$

(0.73) (0.15) (0.029)

где Y – агрегированные расходы на медицинские услуги (в млрд. руб.), P – индекс цен на медицинские услуги, I – средний среднемесячный доход россиян (руб.), в скобках указаны стандартные отклонения.

Дайте экономическую интерпретацию полученным результатам (не забывайте про значимость коэффициентов!).

Решение:

Линейная модель

Проверим значимость коэффициентов:

Для всех коэффициентов критическое значение будет одно и то же: $t_{2,5\%} \approx 2$.

$$t_2 = \frac{-0.427}{0.006} = -71.167 \Rightarrow \text{коэффициент перед } P \text{ значим на уровне } 5\%;$$

$$t_3 = \frac{0.000007}{0.00001} = 0.7 \Rightarrow \text{коэффициент перед } I \text{ не значим на уровне } 5\%.$$

Интерпретация значимого коэффициента:

При увеличении индекса цен на медицинские услуги P на 1 единицу агрегированные расходы на медицинские услуги Y уменьшаются на 0,427 единиц.

Полулогарифмическая модель

Проверим значимость коэффициентов:

$$t_2 = \frac{-0.096}{0.004} = -24 \Rightarrow \text{коэффициент перед } P \text{ значим на уровне } 5\%;$$

$$t_3 = \frac{0.000001}{0.00007} = 0.0143 \Rightarrow \text{коэффициент перед } I \text{ не значим на уровне } 5\%.$$

Интерпретация значимого коэффициента:

При увеличении индекса цен на медицинские услуги P на 1 единицу агрегированные расходы на медицинские услуги Y уменьшаются на $0.096 \cdot 100\% = 9.6\%$.

Логарифмическая модель

Проверим значимость коэффициентов:

$$t_2 = \frac{-3.11}{0.15} = -20.733 \Rightarrow \text{коэффициент перед } P \text{ значим на уровне } 5\%;$$

$$t_3 = \frac{0.0317}{0.029} = 1.093 \Rightarrow \text{коэффициент перед } I \text{ не значим на уровне } 5\%.$$

Интерпретация значимого коэффициента:

При увеличении индекса цен на медицинские услуги P на 1% агрегированные расходы на медицинские услуги Y уменьшаются на 3,11%. Отношение процентного изменения одной величины к процентному изменению другой величины называется эластичностью. В данном случае эластичность агрегированных расходов на мед. услуги Y по индексу цен на мед. услуги P равна 3,11.

2) Зависимость расходов на непродовольственные товары Y от располагаемого дохода X имеет вид: $\ln \hat{Y} = -1.2 + 0.02X$ (все коэффициенты регрессии значимы). При увеличении дохода на 1 единицу расходы увеличатся на

- 1) 0.02 единицы 2) 2 единицы 3) 2% 4) 0.02% 5) 20%

3) Зависимость расходов на продовольственные товары Y от располагаемого дохода X имеет вид: $\ln \hat{Y} = -1.2 + 2 \ln X$ (все коэффициенты регрессии значимы). При увеличении дохода на 1% расходы увеличатся на

- 1) 0.02 единицы 2) 2 единицы 3) 2% 4) 0.02% 5) 20%

4) Доугерти, издание 3.

Выписать, используя табл. на с.123, внизу – линейную спецификацию модели,

На стр.211, 1- ая табл. – полулогарифмическую спецификацию модели

На стр.175- вспомогательные регрессии.

С помощью RSS-ов вспомогательных регрессий провести тест Бокса-Кокса и ответить на вопрос, какая из моделей, линейная или полулогарифмическая лучше?

| .reg EARNINGS S EXP | | | | | | |
|---------------------|------------|-----------|------------|---------------------|----------------------|-----------|
| Source | SS | df | MS | Number of obs = 540 | | |
| Model | 22513.6473 | 2 | 11256.8237 | F(2,537) | = 67.54 | |
| Residual | 89496.5838 | 537 | 166.660305 | Prob > F | = 0.0000 | |
| Total | 112010.231 | 539 | 207.811189 | R-squared | = 0.2010 | |
| | | | | Adj R-squared | = 0.1980 | |
| | | | | Root MSE | = 12.91 | |
| EARNINGS | Coef. | Std. Err. | t | P > t | [95% Conf. Interval] | |
| S | 2.678125 | .2336497 | 11.46 | 0.000 | 2.219146 | 3.137105 |
| EXP | .5624326 | .1285136 | 4.38 | 0.000 | .3099816 | .8148837 |
| _cons | -26.48501 | 4.27251 | -6.20 | 0.000 | -34.87789 | -18.09213 |

| .reg LG EARN S EXP | | | | | | |
|--------------------|------------|-----------|------------|---------------------|----------------------|----------|
| Source | SS | df | MS | Number of obs = 540 | | |
| Model | 50.9842581 | 2 | 25.492129 | F(2,537) | = 100.86 | |
| Residual | 135.723365 | 537 | .252743734 | Prob > F | = 0.0000 | |
| Total | 186.707643 | 539 | .34639637 | R-squared | = 0.2731 | |
| | | | | Adj R-squared | = 0.2704 | |
| | | | | Root MSE | = .50274 | |
| LG EARN | Coef. | Std. Err. | t | P > t | [95% Conf. Interval] | |
| S | .1235911 | .0090989 | 13.58 | 0.000 | .1057173 | .141465 |
| EXP | .0350826 | .0050046 | 7.01 | 0.000 | .0252515 | .0449137 |
| _cons | .5093196 | .1663823 | 3.06 | 0.002 | .1824796 | .8361596 |

| gen EARNSTAR = EARNINGS/16.3135 | | | | | | |
|---------------------------------|------------|-----------|------------|---------------------|----------------------|-----------|
| gen LG EARNST = ln(EARNSTAR) | | | | | | |
| .reg EARNSTAR S EXP | | | | | | |
| Source | SS | df | MS | Number of obs = 540 | | |
| Model | 84.5963381 | 2 | 42.298169 | F(2, 537) | = 67.54 | |
| Residual | 336.288615 | 537 | .626235783 | Prob > F | = 0.0000 | |
| Total | 420.884953 | 539 | .780862622 | R-squared | = 0.2010 | |
| | | | | Adj R-squared | = 0.1980 | |
| | | | | Root MSE | = .79135 | |
| EARNSTAR | Coef. | Std. Err. | t | P > t | [95% Conf. Interval] | |
| S | .1641662 | .0143225 | 11.46 | 0.000 | .1360312 | .1923011 |
| EXP | .0344765 | .0078777 | 4.38 | 0.000 | .0190015 | .0499515 |
| _cons | -1.623503 | .2619003 | -6.20 | 0.000 | -2.137977 | -1.109028 |
| .reg LG EARNST S EXP | | | | | | |
| Source | SS | df | MS | Number of obs = 540 | | |
| Model | 50.9842589 | 2 | 25.4921295 | F(2, 537) | = 100.86 | |
| Residual | 135.72339 | 537 | .252743742 | Prob > F | = 0.0000 | |
| Total | 186.707649 | 539 | .346396379 | R-squared | = 0.2731 | |
| | | | | Adj R-squared | = 0.2704 | |
| | | | | Root MSE | = .50274 | |
| LG EARNST | Coef. | Std. Err. | t | P > t | [95% Conf. Interval] | |
| S | .1235911 | .0090989 | 13.58 | 0.000 | .1057173 | .141465 |
| EXP | .0350826 | .0050046 | 7.01 | 0.000 | .0252515 | .0449137 |
| _cons | -2.282673 | .1663823 | -13.72 | 0.000 | -2.609513 | -1.955833 |

Решение:

Линейная и полулогарифмическая спецификации имеют следующий вид

$$EARNINGS = \beta_0 + \beta_1 S + \beta_2 EXP + u \quad (1)$$

$$\ln EARNINGS = \beta_0 + \beta_1 S + \beta_2 EXP + u \quad (2)$$

Тест Бокса-Кокса предполагает следующие гипотезы:

Но: между моделями (1) и (2) нет статистической разницы

Н1: между моделями (1) и (2) есть статистическая разница, модель с меньшей RSS лучше.

Для проведения теста Бокса-Кокса были оценены следующие вспомогательные регрессии:

$$EARNSTAR = \beta_0' + \beta_1' S + \beta_2' EXP + u \quad (3), \text{ RSS}=336,288615$$

$$\ln EARNSTAR = \beta_0' + \beta_1' S + \beta_2' EXP + u \quad (4), \text{ RSS}=135,72339$$

где $EARNSTAR = EARNINGS / \sqrt[n]{Y_1 \cdot Y_2 \cdot \dots \cdot Y_n}$.

$$\chi^2 = \frac{n}{2} \left| \ln \frac{RSS_3}{RSS_4} \right| = \frac{540}{2} \left| \ln \frac{336,288615}{135,72339} \right| = 244,985 > \chi_{cr}^2(1) \text{ при любом разумном уровне}$$

значимости. Следовательно, полулогарифмическая модель лучше.

5) Исследователь считает, что уровень активности в теневой экономике Y зависит либо положительно от налогового бремени X , либо отрицательно от уровня государственных расходов на предотвращение теневой экономической деятельности Z . Переменная Y может также зависеть от обеих переменных X и Z . Получены международные данные двух перекрестных выборок по Y , X и Z (в млн долл. США): для группы из 30 индустриально развитых и для группы из 30 развивающихся стран. Исследователь оценивает регрессионные зависимости: 1) $\log Y$ от $\log Z$; 2) $\log Y$ только от $\log X$; 3) $\log Y$ только от $\log Z$ одновременно для каждой выборки, получая следующие результаты (в скобках приведены стандартные ошибки):

| | Индустриально развитые страны | | | Развивающиеся страны | | |
|-----------|-------------------------------|---------|---------|----------------------|---------|---------|
| | (1) | (2) | (3) | (1) | (2) | (3) |
| $\log X$ | 0,699 | 0,201 | - | 0,806 | 0,727 | - |
| | (0,154) | (0,112) | | (0,137) | (0,090) | |
| $\log Z$ | -0,646 | - | -0,053 | -0,091 | - | 0,427 |
| | (0,162) | | (0,124) | (0,117) | | (0,116) |
| Константа | -1,137 | -1,065 | 1,230 | -1,122 | -1,024 | 2,824 |
| | (0,863) | (1,069) | (0,896) | (0,873) | (0,858) | (0,835) |
| R^2 | 0,44 | 0,10 | 0,01 | 0,71 | 0,70 | 0,33 |

Переменная X положительно коррелирована с Z в обеих выборках. Выполнив соответствующие статистические тесты, напишите краткий обзор, дав рекомендации исследователю относительно интерпретации полученных результатов.

Выбрать, какая из моделей лучше

А) Для индустриально развитых стран, Б) Для развивающихся стран. Объяснить изменения в оценках коэффициентов и их стандартных отклонений в других моделях.

Решение:

Запишем оцененные модели для индустриально развитых стран:

$$\ln \hat{Y} = -1,137 + 0,699 \ln X - 0,646 \ln Z \quad R^2 = 0,44 \quad (1)$$

(0,863) (0,154) (0,162)

$$\ln \hat{Y} = -1,065 + 0,201 \ln X \quad R^2 = 0,10 \quad (2)$$

(1,069) (0,112)

$$\ln \hat{Y} = 1,230 - 0,053 \ln Z \quad R^2 = 0,01 \quad (3)$$

(0,896) (0,124)

Предположим, лучшая модель – модель (1), тогда в модели (2) коэффициент при переменной $\ln X$ является смещенным. Знак смещения совпадает со знаком произведения коэффициента при невключенном факторе и ковариации включенного и невключенного факторов. $\hat{\beta}_{\ln Z} < 0, \text{cor}(X, Z) > 0 \Rightarrow$ смещение отрицательно, что мы и наблюдаем в модели (2): $0,201 < 0,699$.

Запишем оцененные модели для развивающихся стран:

$$\ln \hat{Y} = -1,122 - 0,806 \ln X - 0,091 \ln Z \quad R^2 = 0,71 \quad (1)$$

(0,873) (0,137) (0,117)

$$\ln \hat{Y} = -1,024 + 0,727 \ln X \quad R^2 = 0,70 \quad (2)$$

(0,858) (0,090)

$$\ln \hat{Y} = 2,824 + 0,427 \ln Z \quad R^2 = 0,33 \quad (3)$$

(0,835) (0,116)

Проанализировав коэффициенты и их стандартные ошибки модели (1), видим, что коэффициент перед $\ln Z$ незначим на уровне значимости 5%: $t = \frac{0,091}{0,117} = 0,778 < 2$. В модели (2) коэффициент перед этой же переменной равен 0,727, а $0,727 \sim 0,8 \Rightarrow$ модель (2) лучшая.

- 6) По 150 наблюдениям оценили зависимость почасовой заработной платы от пола (переменная MALE равно 1 для мужчин и 0 для женщин), длительности обучения S и возраста AGE.

$$\hat{Y} = 3.6 + 3.5 MALE + 3.24 S + 0.44 AGE, \text{ RSS} = 7632$$

(3.09) (1.21) (0.53) (0.057)

Используя результаты двух вспомогательных регрессий, приведенных ниже, проведите RESET – тест и ответьте, правильная ли спецификация модели выбрана.

$$\hat{Y} = 12.37 - 0.29 MALE - 0.49 S - 0.08 AGE + 0.0064 \hat{Y}^2, \text{ RSS} = 7154$$

(4.09) (1.7) (1.3) (0.17) (0.002)

$$\hat{Y} = -18.1 + 9.2 MALE + 7.93 S + 1.1 AGE - 0.012 \hat{Y}^2 + 1.75 \cdot 10^{-10} \hat{Y}^3, \text{ RSS} = 6069$$

(4.42) (2.44) (2.05) (0.28) (0.004) (3.45 \cdot 10^{-11})

Решение:

Оцениваемая модель $Y = \beta_0 + \beta_1 MALE + \beta_2 S + \beta_3 AGE + u (*)$

RESET-тест Рамсея предполагает следующие гипотезы

H0 : спецификация модели (*) является правильной

H1: спецификация модели (*) является неправильной

Для проведения этого теста строятся вспомогательные регрессии:

$$Y = \beta_0 + \beta_1 MALE + \beta_2 S + \beta_3 AGE + \alpha_0 \hat{Y}^2 + u (1),$$

$$Y = \beta_0 + \beta_1 MALE + \beta_2 S + \beta_3 AGE + \alpha_0 \hat{Y}^2 + \alpha_1 \hat{Y}^3 + u (2),$$

где \hat{Y} - столбец оцененных значений регрессии (*).

Используя вспомогательную регрессию (1), проверим следующую гипотезу:

Но: $\alpha_0 = 0$

H1: $\alpha_0 \neq 0$

Рассчитаем значение F-статистики

$$F = \frac{(RSS_R - RSS_{UR})/1}{RSS_{UR}/(150 - 5)} = \frac{(7632 - 7154)/1}{7154/145} \approx 9,7 > F_{(1;145)}$$

=> Гипотеза Но отвергается, спецификация модели (*) является неправильной.

Используя вспомогательную регрессию (2), проверим следующую гипотезу:

Но: $\alpha_0 = \alpha_1 = 0$

H1: $\alpha_0 \neq 0$ или $\alpha_1 \neq 0$

Рассчитаем значение F-статистики

$$F = \frac{(RSS_R - RSS_{UR})/2}{RSS_{UR}/(150 - 6)} = \frac{(7632 - 6069)/2}{6069/144} \approx 18,56 > F_{(2;144)}$$

=> Гипотеза Но отвергается, спецификация модели (*) является неправильной.