

# **Лекция по эконометрике № 1**

## **2 модуль**

### **Функциональные преобразования переменных**

**Демидова  
Ольга Анатольевна  
[https://www.hse.ru/staff/demidova\\_olga](https://www.hse.ru/staff/demidova_olga)  
E-mail:demidova@hse.ru  
28.10.2019**

# План лекции № 1, 2 модуль

- Влияние изменения масштаба измерения переменных на оценки коэффициентов регрессии и их дисперсий.
- Регрессия в центрированных и нормированных переменных.
- Линейная в логарифмах регрессия, как модель с постоянной эластичностью.
- Модель с постоянными темпами роста (полулогарифмическая модель).
- Интерпретация оценок коэффициентов различных функциональных форм.
- Выбор между моделями. Тесты Бокса-Кокса, Бера и МакАлера, МакКиннона, Уайта и Дэвидсона.

### 3 формы системы нормальных уравнений

1)  $Y = X\beta + \varepsilon,$

$X'X\hat{\beta} = X'Y - 1^{\text{ая}} \text{ форма} .$

2) Уравнения в отклонения  $x$  :

$$y = Y - \bar{Y}i,$$

$$x_j = X_j - \bar{X}_j i, \quad j = 1, \dots, k$$

$$y = x\alpha + \varepsilon,$$

Важный результат .  $\hat{\alpha} = \hat{\beta}_{-0}$

### 3 формы системы нормальных уравнений

Доказательство.  $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k$   
 $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_1 + \dots + \hat{\beta}_k \bar{X}_k$   
 $\hat{y} = \hat{\beta}_0 x_0 + \dots + \hat{\beta}_k x_k \Rightarrow \hat{\alpha} = \hat{\beta}_{-0}.$

$$x'x\hat{\beta}_{-0} = x'y \Leftrightarrow$$

$$Var(X)\hat{\beta}_{-1} = Cov(X, Y) - 2^{ая} \text{ форма.}$$

### 3 формы системы нормальных уравнений

$$3) \tilde{x}_j = \frac{x_j}{\hat{\sigma}_j}, j = 1, \dots, k$$

$$\hat{\sigma}_j = \frac{1}{n-1} \sum_{i=1}^n (X_{ji} - \bar{X}_j)^2, \tilde{y} = \frac{y}{\hat{\sigma}_y},$$

$$\tilde{x} \tilde{\hat{\beta}} = \tilde{x} \tilde{y} \Leftrightarrow$$

$$Cor[X] \tilde{\hat{\beta}} = Cor[X, Y] - 3^{б\text{я}} \text{ форма} .$$

## Линейная модель и ее интерпретация

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K + \varepsilon,$$

$\beta_j$  значим=>

Если  $X_j$  увеличится на 1 единицу, то  $Y$  увеличится на  $\hat{\beta}_j$  единиц.

## Линейная в логарифмах модель и ее интерпретация

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \dots + \beta_j \ln X_j + \beta_k \ln X_k + \varepsilon$$

$$\ln \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \ln X_1 + \dots + \hat{\beta}_j \ln X_j + \hat{\beta}_k \ln X_k$$

$$\frac{\dot{Y}}{Y} = \hat{\beta}_j \frac{\dot{X}_j}{X_j}$$

$$\hat{\beta}_j = \frac{\dot{Y} / Y}{\dot{X}_j / X_j} - \text{эластичность.}$$

## Полулогарифмическая модель и ее интерпретация

$$\ln Y = \beta_0 + \beta_1 X_1 + \dots + \beta_j X_j + \beta_k X_k + \varepsilon$$

$$\hat{\ln Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_j X_j + \hat{\beta}_k X_k$$

$$\frac{\dot{Y}}{Y} = \hat{\beta}_j \dot{X}_j, \quad \hat{\beta}_j = \frac{\dot{Y} / Y}{\dot{X}_j}$$

Если  $X_j$  увеличится на 1 единицу, то  $Y$  увеличится на  $\hat{\beta}_j \cdot 100\%$ .



## Выбор между моделями

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon \quad (1)$$

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \dots + \beta_k \ln X_k + \varepsilon \quad (2)$$

$$\ln Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon \quad (3)$$

*(2) и (3) можно сравнить по  $R_{adj}^2$ .*

## Тест Бокса-Кокса

$$Y^{(\theta)} = \frac{Y^\theta - 1}{\theta}, \theta \neq 0$$

$$\lim_{\theta \rightarrow 0} \frac{Y^\theta - 1}{\theta} = \lim_{\theta \rightarrow 0} \frac{Y^\theta \cdot \ln Y}{1} = \ln Y$$

$$Y^{(\theta)} = \begin{cases} \frac{Y^\theta - 1}{\theta}, & \text{если } \theta \neq 0, \\ \ln Y, & \text{если } \theta = 0 \end{cases}$$

## Тест Бокса-Кокса

$$Y^{(\theta)} = \beta_0 + \beta_1 X_1^{(\lambda)} + \dots + \beta_k X_k^{(\lambda)} + \varepsilon$$

$$RSS(\beta_0, \beta_1, \dots, \beta_k, \lambda, \theta) \rightarrow \min$$

$$\hat{\beta}_{0ML}, \hat{\beta}_{1ML}, \dots, \hat{\beta}_{kML}, \hat{\lambda}_{ML}, \hat{\theta}_{ML}$$

## Тест Бокса-Кокса

$$Y^{(\theta)} = \beta_0 + \beta_1 X_1^{(\lambda)} + \dots + \beta_k X_k^{(\lambda)} + \varepsilon$$

$H_0 : \lambda = \theta = 1$     **Линейная модель**

$H_0 : \lambda = \theta = 0$     **Линейная в логарифмах  
модель**

## Тест Бокса-Кокса. Пример

$$W^{(\theta)} = \beta_0 + \beta_1 H^{(\lambda)} + \varepsilon$$

Log likelihood = -2659.5656      Number of obs = 540  
LR chi2(2) = 230.68  
Prob > chi2 = 0.000

| WEIGHT02 | Coef.     | Std. Err. | z     | P> z  | [95% Conf. Interval] |          |
|----------|-----------|-----------|-------|-------|----------------------|----------|
| /lambda  | 1.055498  | 1.892654  | 0.56  | 0.577 | -2.654035            | 4.76503  |
| /theta   | -.0263371 | .1471576  | -0.18 | 0.858 | -.3147607            | .2620865 |

## Тест Бокса-Кокса. Пример

$$W^{(\theta)} = \beta_0 + \beta_1 H^{(\lambda)} + \varepsilon$$

| H0:               | Test           | Restricted |             |
|-------------------|----------------|------------|-------------|
|                   | log likelihood | chi2       | Prob > chi2 |
| theta=lambda = -1 | -2680.8693     | 42.61      | 0.000       |
| theta=lambda = 0  | -2659.7618     | 0.39       | 0.531       |
| theta=lambda = 1  | -2685.5201     | 51.91      | 0.000       |

## **Выбор между линейной и полулогарифмической моделями**

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

$$\log Y = \beta_0 + \beta_1 X + \varepsilon \quad (2)$$

**$H_0$ : качество подгонки моделей (1) и (2) одинаковое**

**$H_1$ : модель с меньшей RSS лучше**

**Выбор между линейной и полулогарифмической моделью можно осуществить с помощью теста Пола Зарембки (частный случай теста Бокса – Кокса).**

## Тест П.Зарембки

- 1) Вычисляется среднее геометрическое значение  $Y$ :

$$\sqrt[n]{Y_1 \cdot Y_2 \cdot \dots \cdot Y_n}$$

- 2) Вводится вспомогательная переменная

$$Y^* = Y / \text{geometric mean of } Y$$

- 3) Оцениваются параметры вспомогательных регрессий

$$Y^* = \beta_0' + \beta_1' X + \varepsilon \quad (3)$$

$$\log Y^* = \beta_0' + \beta_1' X + \varepsilon \quad (4)$$



## Тест П.Зарембки

**4) Вычисляется значение тестовой статистики**

$$\chi^2 = \frac{n}{2} \left| \ln \frac{RSS_3}{RSS_4} \right|$$

**где  $RSS_3$ ,  $RSS_4$  – суммы квадратов остатков в оцененных регрессиях (3) и (4).**

**5) Если при выбранном уровне значимости  $\alpha$**

**то гипотеза  $H_0$  отвергается, между моделями (1) и (2) есть значимое различие. Лучше та модель, при оценивании которой меньше  $RSS$ .**

## BM тест для выбора между линейной и полулогарифмической моделями

### Bera and McAleer test

$$H_0 : \ln Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$$

$$H_1 : Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$$

$$\text{Шаг 1: } \ln \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k,$$
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k$$

## ВМ тест для выбора между линейной и полулогарифмической моделями

### Bera and McAleer test

*Шаг 2: Оцениваются вспомогательные  
регрессии:*

$$\exp(\ln \hat{Y}) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + v_1,$$

$$\ln \hat{Y} = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + v_2$$

*Сохраняются  $\hat{v}_1, \hat{v}_2$ .*

## **ВМ тест для выбора между линейной и полулогарифмической моделями**

*Шаг 3: Оцениваются вспомогательные регрессии:*

$$\ln Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \theta_1 \hat{v}_1 + \varepsilon_1,$$

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \theta_2 \hat{v}_2 + \varepsilon_2$$

*Обычные  $t$  – tests.*

*Если  $\theta_1 = 0$  не отвергается, а  $\theta_2 = 0$  отвергается  
выбирается полулогарифмическая модель.*

*Если  $\theta_2 = 0$  не отвергается, а  $\theta_1 = 0$  отвергается,  
выбирается линейная модель.*

*Возникает проблема если обе гипотезы  
отвергаются или не отвергаются.*

## PE тест для выбора между линейной и полулогарифмической моделями

### MacKinnon-White-Davidson PE test

$$H_0 : \ln Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$$

$$H_1 : Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$$

$$\text{Шаг 1: } \ln \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k,$$

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k,$$

## РЕ тест для выбора между линейной и полулогарифмической моделями

### MacKinnon-White-Davidson PE test

*Шаг 2: Оцениваются вспомогательные  
регрессии:*

$$\ln Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \theta_1 [\hat{Y} - \exp(\ln \hat{Y})] + \varepsilon_1,$$

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \theta_2 [\ln Y - \ln \hat{Y}] + \varepsilon_2,$$

## РЕ тест для выбора между линейной и полулогарифмической моделями

### MacKinnon-White-Davidson PE test

*Шаг 2: Оцениваются вспомогательные  
регрессии:*

$$\ln Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \theta_1 [\hat{Y} - \exp(\ln \hat{Y})] + \varepsilon_1,$$

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \theta_2 [\ln Y - \ln \hat{Y}] + \varepsilon_2,$$

## РЕ тест для выбора между линейной и полулогарифмической моделями

### MacKinnon-White-Davidson PE test

*Обычные  $t$  – tests.*

*Если  $\theta_1 = 0$  не отвергается,  
а  $\theta_2 = 0$  отвергается, выбирается  
полулогарифмическая модель.*

*Если  $\theta_2 = 0$  не отвергается,  
а  $\theta_1 = 0$  отвергается,  
выбирается линейная модель.*

*Возникает проблема, если обе гипотезы  
отвергаются или не отвергаются.*