

Эконометрика
Лекция 26.11.18
**Различные функциональные формы модели,
выбор между ними**

Демидова О.А.
E-mail:demidova@hse.ru

Интерпретация коэффициентов линейной модели

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

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

β_j значим $\Rightarrow$

Если X_j увеличится на 1 единицу, то β_j Y увеличится на единиц (при зафиксированных остальных факторах, *Ceteris paribus*).

Интерпретация коэффициентов литейной модели. Пример

. sum price lotsize prefarea bedrooms bathrms driveway airco

Variable	Obs	Mean	Std. Dev.	Min	Max
price	546	68121.6	26702.67	25000	190000
lotsize	546	5150.266	2168.159	1650	16200
prefarea	546	.2344322	.4240319	0	1
bedrooms	546	2.965201	.7373879	1	6
bathrms	546	1.285714	.5021579	1	4
driveway	546	.8589744	.3483672	0	1
airco	546	.3168498	.465675	0	1

Интерпретация коэффициентов литейной модели. Пример как не надо оценивать

```
. reg price lotsize prefarea bedrooms bathrms driveway airco
```

Source	SS	df	MS	Number of obs =	546
Model	2.3627e+11	6	3.9379e+10	F(6, 539) =	139.33
Residual	1.5233e+11	539	282618294	Prob > F =	0.0000
Total	3.8860e+11	545	713032635	R-squared =	0.6080
				Adj R-squared =	0.6036
				Root MSE =	16811

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lotsize	3.852852	.3650144	10.56	0.000	3.135827	4.569878
prefarea	10439.47	1770.057	5.90	0.000	6962.411	13916.52
bedrooms	4999.33	1063.874	4.70	0.000	2909.481	7089.178
bathrms	18033.47	1573.191	11.46	0.000	14943.13	21123.81
driveway	10190.74	2188.033	4.66	0.000	5892.628	14488.86
airco	15241.56	1614.189	9.44	0.000	12070.68	18412.43
_cons	-5761.746	3667.776	-1.57	0.117	-12966.63	1443.141

Интерпретация коэффициентов литейной модели. Пример как надо оценивать

```
. reg price lotsize prefarea i.bedrooms i.bathrms driveway airco
```

Source	SS	df	MS	Number of obs =	546
Model	2.3851e+11	12	1.9876e+10	F(12, 533) =	70.58
Residual	1.5009e+11	533	281599675	Prob > F =	0.0000
Total	3.8860e+11	545	713032635	R-squared =	0.6138
				Adj R-squared =	0.6051
				Root MSE =	16781

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lotsize	3.838742	.3664153	10.48	0.000	3.118946	4.558537
prefarea	10000.93	1789.829	5.59	0.000	6484.951	13516.92
bedrooms						
2	1396.901	11977	0.12	0.907	-22131.02	24924.82
3	9579.447	11949.71	0.80	0.423	-13894.86	33053.75
4	12313.94	12063.24	1.02	0.308	-11383.39	36011.27
5	14082.43	13094.65	1.08	0.283	-11641.02	39805.88
6	14837.49	16839.34	0.88	0.379	-18242.12	47917.11
bathrms						
2	17741.71	1832.683	9.68	0.000	14141.54	21341.88
3	34326.39	5558.11	6.18	0.000	23407.9	45244.88
4	83973.21	16984.45	4.94	0.000	50608.53	117337.9
driveway	10140.71	2202.863	4.60	0.000	5813.351	14468.07
airco	14739.1	1637.48	9.00	0.000	11522.39	17955.81
_cons	19437.91	11962.38	1.62	0.105	-4061.282	42937.11

Линейная в логарифмах модель

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

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

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

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

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

Линейная в логарифмах модель. Пример

```
. reg lnprice lnlotsize prefarea i.bedrooms i.bathrms driveway airco
```

Source	SS	df	MS	Number of obs = 546		
Model	47.4358442	12	3.95298701	F(12, 533) = 75.31		
Residual	27.977326	533	.052490293	Prob > F = 0.0000		
Total	75.4131702	545	.138372789	R-squared = 0.6290		
				Adj R-squared = 0.6207		
				Root MSE = .22911		

lnprice	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnlotsize	.324457	.02773	11.70	0.000	.2699835	.3789305
prefarea	.1414937	.0243289	5.82	0.000	.0937014	.1892861
bedrooms						
2	.1034527	.1635184	0.63	0.527	-.217767	.4246723
3	.2446626	.1631393	1.50	0.134	-.0758123	.5651376
4	.2645225	.1646884	1.61	0.109	-.0589955	.5880405
5	.3087426	.1786531	1.73	0.085	-.042208	.6596932
6	.3332718	.2299128	1.45	0.148	-.1183746	.7849181
bathrms						
2	.2311596	.0250621	9.22	0.000	.181927	.2803921
3	.3897792	.0757544	5.15	0.000	.2409653	.5385931
4	.7036469	.2317105	3.04	0.003	.248469	1.158825
driveway	.1514601	.0304716	4.97	0.000	.091601	.2113192
airco	.1900504	.0225145	8.44	0.000	.1458223	.2342786
_cons	7.810132	.2762548	28.27	0.000	7.26745	8.352814

Полулогарифмическая модель

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

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

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

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

Полулогарифмическая модель. Пример

```
. reg lnpricel lotsize prefarea i.bedrooms i.bathrms driveway airco
```

Source	SS	df	MS	Number of obs =	546
Model	46.452897	12	3.87107475	F(12, 533) =	71.25
Residual	28.9602732	533	.054334471	Prob > F =	0.0000
Total	75.4131702	545	.138372789	R-squared =	0.6160
				Adj R-squared =	0.6073
				Root MSE =	.2331

lnpricel	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lotsize	.0000544	5.09e-06	10.68	0.000	.0000444	.0000644
prefarea	.1367604	.0248618	5.50	0.000	.0879212	.1855995
bedrooms						
2	.0935436	.166368	0.56	0.574	-.2332739	.420361
3	.2333898	.1659889	1.41	0.160	-.0926829	.5594625
4	.2595624	.1675659	1.55	0.122	-.0696082	.588733
5	.2773368	.1818928	1.52	0.128	-.0799779	.6346514
6	.324802	.2339089	1.39	0.166	-.1346944	.7842984
bathrms						
2	.2385983	.0254571	9.37	0.000	.1885898	.2886069
3	.3790502	.0772056	4.91	0.000	.2273856	.5307148
4	.6849466	.2359246	2.90	0.004	.2214905	1.148403
driveway	.1745734	.0305991	5.71	0.000	.1144637	.2346831
airco	.2035277	.0227456	8.95	0.000	.1588457	.2482098
_cons	10.26266	.1661649	61.76	0.000	9.93624	10.58908

Модели с дамми переменными

$$\ln Y = \beta_0 + \dots + \beta_D D + \dots + \varepsilon$$

$$\ln \hat{Y} = \hat{\beta}_0 + \dots + \hat{\beta}_D D + \dots$$

β_D значим $\Rightarrow$

Y для категории $D=1$ >

Y для категории $D=0$ на $\hat{\beta}_D \cdot 100\%$.

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

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

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

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

Трансформация переменных

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

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

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

Тест Бокса-Кокса

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

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

Box-Cox test

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

$$H_0 : \lambda = \theta = 1 \quad \text{linear regression}$$

$$H_0 : \lambda = \theta = 0 \quad \text{double-logarithmic model}$$

Box-Cox test. 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

Box-Cox test. Пример

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

H0:	Test log likelihood	Restricted 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

Bera and McAleer test

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

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

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

Bera and McAleer test

Шаг 2: Оцениваем вспомогательные регрессии:

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

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

Сохраняем $\hat{v}_1, \hat{v}_2$.

Шаг 3: Оцениваем 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.

BM Test: Полупологарифмическая модель vs линейная модель

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

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

Если $\theta_1 = 0$ не отвергается, $\theta_2 = 0$ отвергается, то выбираем полупологарифмическую модель.

Если $\theta_2 = 0$ не отвергается, $\theta_1 = 0$ отвергается, то выбираем линейную модель.

Если одновременно θ_1, θ_2 значимы или незначимы, то нельзя сделать вывод.

MacKinnon-White-Davidson PE test

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

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

Шаг 2: Оценка вспомогательных регрессий:

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

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

Обычные t -tests.

PE Test: Полулогарифмическая модель vs линейная модель

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

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

Если $\theta_1 = 0$ не отвергается, $\theta_2 = 0$ отвергается, то выбираем полулогарифмическую модель.

Если $\theta_2 = 0$ не отвергается, $\theta_1 = 0$ отвергается, то выбираем линейную модель.

Если одновременно θ_1, θ_2 значимы или незначимы, то нельзя сделать вывод.

PE Test: Линейная в логарифмах модель vs линейная модель

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

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k,$$

Шаг 2 : Оценка вспомогательных регрессий :

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

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

Если $\theta_1 = 0$ не отвергается, $\theta_2 = 0$ отвергается, то выбираем линейную в логарифмах модель.

Если $\theta_2 = 0$ не отвергается, $\theta_1 = 0$ отвергается, то выбираем линейную модель.

Если одновременно θ_1, θ_2 значимы или незначимы, то нельзя сделать вывод.



NATIONAL RESEARCH
UNIVERSITY

Thank you for your attention!

20, Myasnitskaya str., Moscow, Russia, 101000

Tel.: +7 (495) 628-8829, Fax: +7 (495) 628-7931

www.hse.ru