DESCRIPTIVE STATISTICS

Dr Alina Gleska

Institute of Mathematics PUT

12 maja 2019

Let us remind You that - in case of the linear regression with only one predictor $Y = bX + a$ the coefficients b and a we calculate as:

$$b = \frac{cov(X,Y)}{S_X^2} = \frac{\sum_{i=1}^{n} xy - n\overline{xy}}{\sum_{i=1}^{n} x^2 - n(\overline{x})^2} = \frac{\sum_{i=1}^{n} xy - (\sum_{i=1}^{n} x \sum_{i=1}^{n} y)/n}{\sum_{i=1}^{n} x^2 - (\sum_{i=1}^{n} x)^2/n},$$

$$a = \overline{y} - b\overline{x} = (\sum_{i=1}^{n} y - b\sum_{i=1}^{n} x)/n.$$

General linear model $Y = b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$

If we have an intercept in our general model we assume that the value for the corresponding variable X always equals 1.

This time we are going to use matrices. Notation:

- b - the vertical vector of parameters, e.g. $b = [b_k]$, $k = 1, 2, \ldots, K$;
- X - the matrix of the data (given data);
- y - the vertical vector of the results.

Obtained by the classic OLS method, the vector of parameters of the linear model is the solution of the so-called normal equations

$$(X'X)b = X'y.$$

In the explicit form, the vector b is expressed by the formula:

$$b = (X'X)^{-1}X'y.$$

This formula is unique if the matrix $X'X$ is invertible. This is possible only when the rank of the matrix X is equal to the number of parameters. In particular, necessarily

- the number of non-repeated observations must be greater than the number of parameters (i.e. also $n > K$);
- no column of matrix X can be linearly dependent on others (in particular, two columns cannot be identical).

How to interpret the matrix $X'X$ and the vector $X'y$?

In case of the model $Y = bX + a = b_1 X_1 + b_2 X_2$, where $b_2 = a$, $X_2 = 1$ we have:

$$X'X = \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} 1^2 \end{bmatrix} \quad X'y = \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}.$$

For the model $Y = b_1 X_1 + b_2 X_2 + b_3 1$ we obtain:

$$X'X = \begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1} x_{i2} & \sum_{i=1}^{n} x_{i1} \\ \sum_{i=1}^{n} x_{i1} x_{i2} & \sum_{i=1}^{n} x_{i2}^2 & \sum_{i=1}^{n} x_{i2} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} 1^2 \end{bmatrix} \quad X'y = \begin{bmatrix} \sum_{i=1}^{n} x_{i1} y_i \\ \sum_{i=1}^{n} x_{i2} y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}.$$

EKONOMETRIA I BADANIA OPERACYJNE
(Econometry and operational research)
Zagadnienia podstawowe
Redakcja naukowa Bogusław GUZIK
Wydawnictwo Akademii Ekonomicznej w Poznaniu

The verification of the linear model obtained by the classical OLS

The minimal set of assumptions:

- the econometric model can not raise substantive objections,
- the model should be very well fitted to empirical data,
- all model explanatory variables must be significant.

The overall model verification activities can be divided into:

- substantive verification (postulate 1),
- statistical verification (postulates 2 and 3).

MERYTORIC VERIFICATION consists in checking whether the econometric model is consistent with economic knowledge about the phenomenon under investigation, economic theory, and finally - with common sense. During substantive (merytoric) verification, we examine e.g.

- are the model parameters characters meaningful;
- whether the scale of parameters is acceptable;
- whether the model can be sensibly extrapolated;
- whether from the model for the Y variable there are sensible models for variables related to the tested variable.

Model matching (fitting). Coefficient of determination

Coefficient of determination R^2 indicates what part of the overall observed variability of the dependent variable Y was explained by the econometric model.

Coefficient of indetermination φ^2 indicates what part of the overall observed variability of the dependent variable Y was not explained by the econometric model.

The good model is the one for which $R^2 > 90\%$, and $\varphi^2 < 10\%$.

There are different definitions of these coefficients. By default, it is assumed that:

- the measure of the overall observed variability of the variable Y is the sum of the squares deviations of the value of this variable from its average, i.e. Sum of Squares Total SST:

$$SST = \sum (y - \bar{y})^2; \tag{1}$$

- the measure of the observed variability of Y, not explained by the model is the sum of squared deviations of the variable Y from the model, i.e. the sum of squares of residuals (Sum of Squares for Errors SSE):

$$SSE = \sum (y - \hat{y})^2; \tag{2}$$

Coefficient of indetermination φ^2 is defined by the formula:

$$\varphi^2 = \frac{SSE}{SST}. \tag{3}$$

Coefficient of determination R^2 is defined as

$$R^2 = 1 - \varphi^2, \tag{4}$$

provided $\varphi^2 \leq 1$. Let us observe that

$$R^2 = \frac{SSR}{SST},$$

where $SSR = SST - SSE$ is so-called Sum of Squares for Regression SSR.

Calculating the total sum of squares SST and the sum of squares for errors SSE directly from the definition can be troublesome. SST is easier to calculate as:

$$SST = \sum y^2 - 2\sum y\bar{y} + \sum \bar{y}^2 = \sum y^2 - \frac{(\sum y)^2}{n}, \tag{5}$$

and SSE - in the case of a linear model determined by the classic OLS - from the formula

$$SSE = \sum y^2 - b'X'y. \tag{6}$$

For a model with one explanatory variable $Y = bX + a$ it will be

$$SSE = \sum y^2 - (b\sum xy + a\sum y). \tag{7}$$

By definition (3) it follows that the coefficient of indetermination φ^2 compares the variability of the explained variable Y around the model (which is represented by SSE) to the variable Y around the average (which is represented by SST). Thus the coefficient of indetermination compares two descriptions of a variable: a description through a model to a description by means of an average.

The quotient $\frac{1}{\varphi^2}$ can be used as an indicator of how many times the description of the variable Y through the model is better than the description by the mean.

SIGNIFICANCE OF EXPLANATORY VARIABLES

The explanatory variable is significant when it affects the explained variable in a noticeable (explicit) way.

For the linear model $Y = b_1 X_1 + \cdots + b_k X_k$, the variable is significant when the parameter next to it is significantly different from zero.

The parameters of the econometric model are almost always non-zero, so you have to check their structure differently. Remember that the statistical material used to determine the model parameters is generally a small fragment of the set of all possible observation results of the explained variable and explanatory variables. In the set of possible observation results, the dependency of the variable Y on the variable X may not exist at all, and the fact that we obtained a non-zero parameter for the given statistical material, could only be accidental .

All explanatory variables of the econometric model must be significant

Significance testing can take place in various ways. The standard method involves the use of stochastic (probabilistic) assumptions about the manner in which the results of observation of the explained variable are generated. Such assumptions can be found in econometric textbooks.
In the case of linear models, among many systems of stochastic assumptions, the assumptions of so-called classic linear regression are used. With these assumptions - and when the model was obtained by a classic OLS - the significance of the variable is tested as follows:

1. we calculate the so-called empirical Student statistics, t_K, regarding the examined explanatory variable;
2. we establish the critical value of the Student statistics, t_{KR};
3. we compare the modulus of the empirical t-Student statistics with the critical value.

- An explanatory variable is considered to be significant if the empirical statistics associated with it are larger than the critical value, i.e. $|t_K| > t_{KR}$.
- In the opposite case, e.g. for $|t_K| \leq t_{KR}$, an explanatory variable is considered to be non-significant (negligible).
- The critical value t_{KR} is read from the tables as the value of the Student's t-distribution with the assumed significance level α (usually $\alpha = 5\%$) and regarding the tested model the degrees of freedom Q:

$$Q = n - K, \tag{8}$$

where n is the number of observations, and K is the number of estimated parameters.

- IT programs usually give the values of empirical Student's statistics. Sometimes, however, they are not displayed but give so-called estimated mean errors d_k (related to individual model parameters). This information is enough to calculate Student statistics, because

$$\text{empirical t-Student statistics} = \frac{\text{model parameter}}{\text{estimated mean error}},$$

$$t_k = \frac{b_k}{d_k} \quad (k = 1, 2, \ldots, K). \tag{9}$$

If, on the other hand, we have to calculate the empirical t-Student statistics, we proceed as follows:

- first we count the estimation of the standard deviation of random components s:

$$s = \sqrt{\frac{SSE}{Q}}; \tag{10}$$

- then we count the estimated mean errors d_k:

$$d_k = s\sqrt{c_k}, \tag{11}$$

  where c_k is the $k-th$ element of the main diagonal of the inverse matrix $(X'X)^{-1}$;
- finally, based on (9), we count the t-Student empirical statistics.

Interpretation of the estimated mean error

The stochastic assumptions of linear regression, among others, admit that

- there exist the real model $\phi = \beta_1 X_1 + \cdots + \beta_k X_k$ of the variable Y, with real parameters $\beta_1, \ldots, \beta_k$. It is so-called hypothetical model of the variable Y;
- the value y of the variable Y in the given observation is just one of the many possible realizations of the variable Y. The possibility of other values of the Y variable results from the so-called a random component, representing e.g. the impact of random circumstances.

In this situation, the statistical material used to determine the parameters of the model is one of many possible statistical materials, and the parameters $b_1, \ldots, b_k$ of the econometric model we have set are the one of many possible ratings of parameters $\beta_1, \ldots, \beta_k$ of the hypothetical model. In particular, the b_k parameter is one of many possible evaluations of the hypothetical β_k parameter.

The estimated mean error, d_k, is an assessment of the discrepancy between the possible β_k parameter estimates around this parameter.

Of course, the smaller the mean error, the better situation. Then, possible evaluations of the parameter β_k are less deviating from this parameter. And because one of these possible ratings is our rating b_k so we think that the estimated mean error d_k is smaller, the presumable precision of estimating β_k by b_k is greater.

Interpretation of the estimation of the standard deviation of random components

The estimation of the standard deviation of the random components s is an assessment of the discrepancy between the possible values of the variable explained around the hypothetical model.

Example.

Based on 20 observations, the following econometric model was obtained for the demand for domestic cars (PSK) against income (DO), the domestic cars price index (CSK) and the imported car price index (CSI):

$$\underset{}{PSK} = \underset{(4,9)}{1,2DO} - \underset{(3,2)}{0,23CSK} + \underset{(1,9)}{0,12CSI} + \underset{(7,3)}{3,4}$$

$$R^2 = 0,987.$$

The modules of the t-Student empirical statistics were given under the parameters of the econometric model. At the 5% significance level, check the significance of the explanatory variables.

Since $n = 20$, $K = 4$, the degrees of freedom are $Q = 20 - 4 = 16$. We find the critical value from the table of t-Student distribution for 5% significance level and 16 degrees of freedom. There is $t_{KR} = 2.1199$.

Model verification

Wartości krytyczne rozkładu t-Studenta

$X \sim t_\nu$ - X zmienna losowa o rozkładzie t-Studenta z liczbą stopni swobody ν,
α - poziom istotności,
$t_{\nu,\alpha}$ - wartość krytyczna - liczba taka, że $P(|X| > t_{\nu,\alpha}) = \alpha$

ν \ a	0,400	0,300	0,200	0,100	0,050	0,025	0,025	0,010	0,005	0,001
1	1,3764	1,9626	3,0777	6,3137	12,7062	25,4519	25,4519	63,6559	127,3211	636,5776
2	1,0607	1,3862	1,8856	2,9200	4,3027	6,2054	6,2054	9,9250	14,0892	31,5998
3	0,9785	1,2498	1,6377	2,3534	3,1824	4,1765	4,1765	5,8408	7,4532	12,9244
4	0,9410	1,1896	1,5332	2,1318	2,7765	3,4954	3,4954	4,6041	5,5975	8,6101
5	0,9195	1,1558	1,4759	2,0150	2,5706	3,1634	3,1634	4,0321	4,7733	6,8685
6	0,9057	1,1342	1,4398	1,9432	2,4469	2,9687	2,9687	3,7074	4,3168	5,9587
7	0,8960	1,1192	1,4149	1,8946	2,3646	2,8412	2,8412	3,4995	4,0294	5,4081
8	0,8889	1,1081	1,3968	1,8595	2,3060	2,7515	2,7515	3,3554	3,8325	5,0414
9	0,8834	1,0997	1,3830	1,8331	2,2622	2,6850	2,6850	3,2498	3,6896	4,7809
10	0,8791	1,0931	1,3722	1,8125	2,2281	2,6338	2,6338	3,1693	3,5814	4,5868
11	0,8755	1,0877	1,3634	1,7959	2,2010	2,5931	2,5931	3,1058	3,4966	4,4369
12	0,8726	1,0832	1,3562	1,7823	2,1788	2,5600	2,5600	3,0545	3,4284	4,3178
13	0,8702	1,0795	1,3502	1,7709	2,1604	2,5326	2,5326	3,0123	3,3725	4,2209
14	0,8681	1,0763	1,3450	1,7613	2,1448	2,5096	2,5096	2,9768	3,3257	4,1403
15	0,8662	1,0735	1,3406	1,7531	2,1315	2,4899	2,4899	2,9467	3,2860	4,0728
16	0,8647	1,0711	1,3368	1,7459	2,1199	2,4729	2,4729	2,9208	3,2520	4,0149
17	0,8633	1,0690	1,3334	1,7396	2,1098	2,4581	2,4581	2,8982	3,2224	3,9651
18	0,8620	1,0672	1,3304	1,7341	2,1009	2,4450	2,4450	2,8784	3,1966	3,9217
19	0,8610	1,0655	1,3277	1,7291	2,0930	2,4334	2,4334	2,8609	3,1737	3,8833
20	0,8600	1,0640	1,3253	1,7247	2,0860	2,4231	2,4231	2,8453	3,1534	3,8496
21	0,8591	1,0627	1,3232	1,7207	2,0796	2,4138	2,4138	2,8314	3,1352	3,8193
22	0,8583	1,0614	1,3212	1,7171	2,0739	2,4055	2,4055	2,8188	3,1188	3,7922
23	0,8575	1,0603	1,3195	1,7139	2,0687	2,3979	2,3979	2,8073	3,1040	3,7676
24	0,8569	1,0593	1,3178	1,7109	2,0639	2,3910	2,3910	2,7970	3,0905	3,7454
25	0,8562	1,0584	1,3163	1,7081	2,0595	2,3846	2,3846	2,7874	3,0782	3,7251
26	0,8557	1,0575	1,3150	1,7056	2,0555	2,3788	2,3788	2,7787	3,0669	3,7067
27	0,8551	1,0567	1,3137	1,7033	2,0518	2,3734	2,3734	2,7707	3,0565	3,6895
28	0,8546	1,0560	1,3125	1,7011	2,0484	2,3685	2,3685	2,7633	3,0470	3,6739
29	0,8542	1,0553	1,3114	1,6991	2,0452	2,3638	2,3638	2,7564	3,0380	3,6595
30	0,8538	1,0547	1,3104	1,6973	2,0423	2,3596	2,3596	2,7500	3,0298	3,6460
31	0,8534	1,0541	1,3095	1,6955	2,0395	2,3556	2,3556	2,7440	3,0221	3,6335
32	0,8530	1,0535	1,3086	1,6939	2,0369	2,3518	2,3518	2,7385	3,0149	3,6218
33	0,8526	1,0530	1,3077	1,6924	2,0345	2,3483	2,3483	2,7333	3,0082	3,6109
34	0,8523	1,0525	1,3070	1,6909	2,0322	2,3451	2,3451	2,7284	3,0020	3,6007
35	0,8520	1,0520	1,3062	1,6896	2,0301	2,3420	2,3420	2,7238	2,9961	3,5911
40	0,8507	1,0500	1,3031	1,6839	2,0211	2,3289	2,3289	2,7045	2,9712	3,5510
45	0,8497	1,0485	1,3007	1,6794	2,0141	2,3189	2,3189	2,6896	2,9521	3,5203
50	0,8489	1,0473	1,2987	1,6759	2,0086	2,3109	2,3109	2,6778	2,9370	3,4960
55	0,8482	1,0463	1,2971	1,6730	2,0040	2,3044	2,3044	2,6682	2,9247	3,4765
60	0,8477	1,0455	1,2958	1,6706	2,0003	2,2990	2,2990	2,6603	2,9146	3,4602
65	0,8472	1,0448	1,2947	1,6686	1,9971	2,2945	2,2945	2,6536	2,9060	3,4466
70	0,8468	1,0442	1,2938	1,6669	1,9944	2,2906	2,2906	2,6479	2,8987	3,4350
75	0,8464	1,0436	1,2929	1,6654	1,9921	2,2873	2,2873	2,6430	2,8924	3,4249
80	0,8461	1,0432	1,2922	1,6641	1,9901	2,2844	2,2844	2,6387	2,8870	3,4164
85	0,8459	1,0428	1,2916	1,6630	1,9883	2,2818	2,2818	2,6349	2,8822	3,4086
90	0,8456	1,0424	1,2910	1,6620	1,9867	2,2795	2,2795	2,6316	2,8779	3,4019
95	0,8454	1,0421	1,2905	1,6611	1,9852	2,2775	2,2775	2,6286	2,8741	3,3958
100	0,8452	1,0418	1,2901	1,6602	1,9840	2,2757	2,2757	2,6259	2,8707	3,3905

Let us make the analysis knowing $t_{KR} = 2.12$.

- Empirical t-Student statistics for income (equal to 4.9) is greater than the critical value; income is therefore the significant explanatory variable.
- In the case of prices of domestic cars, the t-Student empirical statistics is 3.2 and is greater than the critical value. Thus, the domestic car prices has a significant impact on the demand for domestic cars.
- For imported car prices, the t-Student empirical statistics (1.9) is less than the critical value of 2.12. Therefore, we consider that for these data, prices of imported cars are irrelevant.

Final conclusion: the model - despite a very good fit - can not be considered as final, because it contains a non-significant explanatory variable (CSI). Further actions would be to exclude this variable and re-estimating the model, this time as the dependence of demand on income and prices of domestic cars. Due to the change in the list of explanatory variables, the ”new” parameters for income and prices of domestic cars will generally not be the same as before.

Among the nonlinear models, the following models are distinguished:

- linear with respect to parameters;
- linearizable;
- nonlinear in the strict sense (others).

At the moment we will deal with the first of them.

The model $\hat{Y} = f(X, b)$ is linear with respect to parameters, if it can be presented as linear function of the unique transformations of explanatory variables X; the coefficients of these transformations are known with numerical accuracy.

So the model is linear with respect to parameters if it can be expressed in the form

$$\hat{Y} = \sum_k b_k Z_k, \tag{12}$$

where

$$Z_k = h_k(X), \tag{13}$$

and all transformations h_k are unique, and their coefficients are known. The variables Z_k, which are the transformations of original explanatory variables X,we will call auxiliary explanatory variables, and the model (12) - auxiliary linear model.

The determination and verification of the linear model with respect to parameters comes down to the determination and the verification of the auxiliary linear model (12), in which:

- the explained variable is the original variable Y,
- the explanatory variables are the auxiliary explanatory variables Z_k.

The parameter vector obtained from the classic *OLS* is therefore expressed by the formula:

$$b = (Z'Z)^{-1}Z'y. \tag{14}$$

PARABOLA

The course of the parabola depends on the sign of the coefficient standing next to the square of the explanatory variable b_2. The parabola is suitable for description of runs:

- with one minimum, first decreasing slowly and then growing faster and faster ($b_2 > 0$);
- with one maximum, first growing slower and then decreasing faster and faster ($b_2 < 0$).

The model is described by the equation:

$$Y = b_0 + b_1 X + b_2 X^2 \quad (b_2 \neq 0). \tag{15}$$

The successive powers of variable X, e.g X and X^2, are treated as auxiliary explanatory variables and in relation to the auxiliary model we use the methods of estimation and verification of linear models.

Instead of the original model

$$Y = b_0 + b_1 X + b_2 X^2$$

we estimate and verify the auxiliary model

$$Y = b_0 + b_1 Z_1 + b_2 Z_2, \tag{16}$$

where

$$Z_1 = X, \quad Z_2 = X^2. \tag{17}$$

POLYNOMIAL OF K-TH ORDER

The model is described by the equation:

$$Y = b_0 + b_1 X + b_2 X^2 + \cdots + b_k X_k \quad (b_k \neq 0). \tag{18}$$

The successive powers of variable X are treated as auxiliary explanatory variables and for the auxiliary model we use the methods of estimation and verification of linear models. Instead of the original model

$$Y = b_0 + b_1 X + b_2 X^2 + \cdots + b_k X^k$$

we estimate and verify the auxiliary model

$$Y = b_0 + b_1 Z_1 + b_2 Z_2 + \cdots b_k Z_k, \tag{19}$$

where

$$Z_k = X^k, \quad k = 1, \ldots, K. \tag{20}$$

HYPERBOLE

The model is described by the equation:

$$Y = b_0 + \frac{b_1}{X} \quad (X > 0;\ \ b_1 \neq 0). \tag{21}$$

Properties of the hyperbole:

- hyperbole has a horizontal asymptote $Y = b_0$;
- the course of the hyperbole depends on the sign of the coefficient b_1 - if it is positive, the hyperbole is decreasing, and if it is negative, the hyperbole is increasing;
- for $X = 0$ the hyperbole is not defined.

The hyperbole is suitable for description of such phenomena which for $X > 0$:

- are growing slower and slower and finally they lead to a saturation level b_0 (in this case $b_1 < 0$);
- are decreasing slower and slower to the bottom level b_0 (in this case $b_1 > 0$).

General case

The hyperbole of many explanatory variables:

$$Y = b_0 + \frac{b_1}{X_1} + \frac{b_2}{X_2} + \cdots + \frac{b_k}{X_k} \quad (\text{all } X_k > 0). \tag{22}$$

The reciprocals of every successive variables we treat as successive auxiliary explanatory variables and then we estimate and verify the proper auxiliary linear model. Instead of the model (22) we consider the auxiliary linear model:

$$Y = b_0 + b_1 Z_1 + b_2 Z_2 + \cdots + b_k Z_k, \tag{23}$$

where

$$Z_1 = \frac{1}{X_1}, \; Z_2 = \frac{1}{X_2}, \; \ldots, \; Z_k = \frac{1}{X_k}. \tag{24}$$

LOGARITHMIC FUNCTION

The logarithmic function with one explanatory variable X:

$$Y = b_0 + b_1 \log(X), \quad (X > 0, b_1 \neq 0). \tag{25}$$

The course of the logarithmic function depends on sign of the coefficient b_1. For positive X the logarithmic function can described two runs:

- increasing (without any bounds) slower and slower (for the coefficient $b_1 > 0$);
- decreasing (without any bounds) slower and slower (for the coefficient $b_1 < 0$).

General case

The general case of the logarithmic function:

$$Y = b_0 + b_1 \log(X_1) + \cdots + b_k \log(X_k), \quad \text{all } X_k > 0. \tag{26}$$

The logarithms of every successive variables we treat as successive auxiliary explanatory variables and then we estimate and verify the proper auxiliary linear model.
In case of the model (26), the auxiliary linear model is

$$Y = b_0 + b_1 Z_1 + b_2 Z_2 + \cdots + b_k Z_k, \tag{27}$$

where

$$Z_1 = \log(X_1), \; Z_2 = \log(X_2), \; \ldots, \; Z_k = \log(X_k). \tag{28}$$

LINEARIZABLE MODELS

We call the model linearizable, if there exists a unique transformation of both sides of the equation such that finally we get the linear model or the model linear with respect to parameters.

The linear model obtained after the linearization we will call the auxiliary linear model, and its variables - auxiliary variables. We denote the auxiliary explained variable by V if needed; and the auxiliary explanatory variables by $Z_1, Z_2, \ldots, Z_k$.

EXPONENTIAL MODEL

The exponential model with one explanatory variable:

$$Y = Ap^{bX} \quad (A > 0), \tag{29}$$

where p - positive constant; A, b - parameters. The most commonly used is the exponential function at the base of the natural logarithm ($p = e$) or at the base of the decimal logarithm ($p = 10$). The parameter A defines the theoretical level of the explained variable Y in the period preceding the first given period (e.g. for $X = 0$). The course of the exponential function depends on the sign of the coefficient (the exponent) b. For $A > 0$ the function:

- is growing faster and faster (unboundly), if the exponent $b > 0$;
- is decreasing slower and slower to zero, if the component $b < 0$.

General case

The exponential model with many variables:

$$Y = Ap^{b_1X_1+b_2X_2+\cdots+b_KX_K} \quad (A, p > 0). \tag{30}$$

The main property of the exponential function - fixed growth rates

The growth rate shows what is the expected relative growth of Y, when the given explanatory variable X is growing by one unit,and the other explanatory variables stay unchanged. Only exponential functions have such properties.

In case of an exponential function, the growth rate (r_i) with respect to the explanatory variable X_i is defined by the coefficient standing next this variable, e.g.

$$r_i = b_i \cdot \ln(p) \quad (1 \leq i \leq K). \tag{31}$$

In particular, for base $p = e$ we have

$$r_i = b_i. \tag{32}$$

The power model

$$Y = AX^{\beta} \tag{33}$$

can be logarithmized to the linear model

$$ln(Y) = ln(A) + \beta ln(X), \tag{34}$$

which can be rewritten as

$$Y^* = A^* + \beta X^*.$$

Using the OLS method we obtain:

$$A = \exp(A^*), \quad \beta = \beta^*.$$

The parameter A defines the theoretical level of the explained variable Y at the unit value of X, and the parameter β defines percentage change in the value of the dependent variable caused by the change in the value of the independent variable by 1% (this property is called the elasticity of a dependent variable with respect to an independent variable).