TJMM
3 (2011), No. 2, 119-124

THE DIRECT AND THE INVERSE PROBLEM FOR SIMPLE REGRESSION MODELS

FELICIA ZĂVOIANU AND CONSTANTIN ZĂVOIANU

Abstract. This article contains a personal point of view regarding the implementation of simple regression models. We consider that the approach of considering in parallel both the direct and inverse problems helps in constructing a global, in depth, image regarding the usefulness of simple regression problems.

1. The formulation of the problem

Regarding the implementation of simple regression problems we mention that by using various means specific to the studied phenomenon, one can act on the variables X and Y in order to achieve a certain objective. For any given model, one can formulate and solve two essential problems:

1) – ***the direct problem*** that is aimed at obtaining predictions regarding the evolution of the studied phenomenon in order to help create various prognosis studies; the prediction estimates, using the model, a set of values $\hat{y}_{0i}$ for certain x_{0i} that have not been taken into consideration when the model was adjusted;

2) – ***the inverse problem*** that is aimed at determining, using the model, those values x_{0i} such that the target variable will take the preset values $\tilde{y}_{0i}$ established by the researcher.

Simple regression models can be divided in two classes: *additive models* and *multiplicative models*. In the present study, from the class of additive models we shall consider the linear model, the logarithmic model and the hyperbolic model and from the class of multiplicative models we shall consider the power model, the exponential model and two logistic models.

If $y_1, y_2, \dots y_n$ is a set of observations regarding the random variable Y, that correspond to the n experimental values of the factor X, noted with $x_1, x_2, \dots x_n$, i.e.:

X	x_1	x_2	x_3	$\dots$	x_i	$\dots$	x_n
Y	y_1	y_2	y_3	$\dots$	y_i	$\dots$	y_n

then, a simple regression models has the form $Y = f(\alpha_0, \alpha_1, X, u)$, where $\alpha_0, \alpha_1, u \in \mathbb{R}$, u being a residual variable that follows a normal distribution with $M(u) = 0$ and $D^2(u) = \sigma^2$, $\sigma \neq 0$, i.e. $u \sim N(0, \sigma)$.

The type of every model (additive or multiplicative) depends on the form of the function f included in the analysis, such as it appears in the following definitions:

2010 *Mathematics Subject Classification.* 62J02, 62J05.

Key words and phrases. simple regression models, direct problem, inverse problem, multiplication factor.

$$f(\alpha_0,\alpha_1,X,u)=\begin{cases}\alpha_0+\alpha_1X+u & , \ \textit{linear model}\\ \alpha_0+\alpha_1\log_a X+u,\ a>0,\ a\neq 1 & , \ \textit{logarithmic model}\\ \alpha_0+\alpha_1\frac{1}{X}+u & , \ \textit{hyperbolic model}\\ \alpha_0X^{\alpha_1}10^u & , \ \textit{power model}\\ e^{\alpha_0+\alpha_1X+u} & , \ \textit{exponential model}\\ \frac{b}{1+e^{\alpha_0+\alpha_1X+u}}\ ,\ b>0 & , \ \textit{logistic model}\\ \frac{1+e^{\alpha_0+\alpha_1X+u}}{c},\ c\neq 0 & , \ \textit{logistic model}\end{cases}$$

In case the two variables X and Y are strongly correlated, one can determine, starting from the values of these variables[1], the best estimations of the statistical parameters α_0 and α_1, noted by a_0 and a_1, and then the function $\hat{f}(a_0,a_1,X)$, such that the regression equation $\hat{Y}=\hat{f}(a_0,a_1,X)$ can be efficiently used for analysis and prognosis.

2. Solving the direct problem of simple regression models

In the general case, the direct problem of a simple regression model is formulated as: *what is the mean change of the target variable Y if the factorial variable X is changed by p units, where $p\in\mathbb{R}$?*

In the following section we shall analyze this problem for every model included in the analysis.

1. It is known that for the *linear model* $\hat{Y}=a_0+a_1X$, the regression coefficient a_1 shows the mean modification of the variable Y when variable X is modified by one unit, because $\hat{Y}^{(\pm 1)}=a_0+a_1(X\pm 1)=a_0+a_1X\pm a_1=\hat{Y}\pm a_1$ (particular case of the direct problem). If variable X is modified by p units, then

$$\hat{Y}^{(+p)}=a_0+a_1(X+p)=a_0+a_1X+a_1p=\hat{Y}+a_1p$$

i.e.

$$\hat{Y}^{(+p)}=\hat{Y}+a_1p.$$

In conclusion, if variable X is modified by p units, then variable Y will be modified in average by a_1p units.

2. In the case of the *logarithmic model* $\hat{Y}=a_0+a_1\log_a X$, if variable X is modified by p units then

$$\hat{Y}^{(+p)}=a_0+a_1\log_a(X+p)$$

and

$$\hat{Y}^{(+p)}-\hat{Y}=a_1\log_a(X+p)-a_1\log_a X$$

i.e.

$$\hat{Y}^{(+p)}=\hat{Y}+a_1\log_a\left(\frac{X+p}{X}\right).$$

In conclusion, if variable X is modified by p units, then variable Y will be modified in average by $a_1\log_a\left(\frac{X+p}{X}\right)$ units.

3. In the case of the *hyperbolic model* $\hat{Y}=a_0+a_1\frac{1}{X}$ the modification of variable X by p units leads to

$$\hat{Y}^{(+p)}=a_0+a_1\frac{1}{X+p}$$

and

$$\hat{Y}^{(+p)}-\hat{Y}=a_1\left(\frac{1}{X+p}-\frac{1}{X}\right)$$

[1]Through the linearization of the models and the least squares method.

i.e.

$$\hat{Y}^{(+p)} = \hat{Y} - a_1 \frac{p}{X(X+p)}.$$

In conclusion, if variable X is modified by p units, then variable Y will be modified in average by $-a_1 \frac{p}{X(X+p)}$ units.

4. In the case of the *power model* $\hat{Y} = a_0 X^{a_1}$ the modification of variable X by p units leads to

$$\hat{Y}^{(+p)} = a_0 (X+p)^{a_1}$$

and because

$$\frac{\hat{Y}^{(+p)}}{\hat{Y}} = \frac{a_0 (X+p)^{a_1}}{a_0 X^{a_1}}$$

we have that

$$\hat{Y}^{(+p)} = \hat{Y} \left(\frac{X+p}{X} \right)^{a_1}, p \in \mathbb{R}.$$

In conclusion, the modification of variable X by p units implies the average modification of the variable Y induced by multiplying the latter with the factor $\left(1 + \frac{p}{X}\right)^{a_1}$.

5. In the case of the *exponential model* $\hat{Y} = e^{a_0 + a_1 X}$ the modification of variable X by p units leads to

$$\hat{Y}^{(+p)} = e^{a_0 + a_1 (X+p)} = e^{a_0 + a_1 X} e^{a_1 p}$$

and because

$$\frac{\hat{Y}^{(+p)}}{\hat{Y}} = \frac{e^{a_0 + a_1 X} e^{a_1 p}}{e^{a_0 + a_1 X}} = e^{a_1 p}$$

we have that

$$\hat{Y}^{(+p)} = \hat{Y} e^{a_1 p}.$$

In conclusion, the modification of variable X by p units implies the average modification of the variable Y induced by multiplying the latter with the factor $e^{a_1 p}$.

6. In the case of the *logistic model* $\hat{Y} = \frac{b}{1 + e^{a_0 + a_1 X}}$ the modification of variable X by p units leads to

$$\hat{Y}^{(+p)} = \frac{b}{1 + e^{a_0 + a_1 (X+p)}} = \frac{b}{1 + e^{a_0 + a_1 X} e^{a_1 p}}$$

and

$$\frac{\hat{Y}^{(+p)}}{\hat{Y}} = \frac{1 + e^{a_0 + a_1 X}}{1 + e^{a_0 + a_1 X} e^{a_1 p}}$$

from where we arrive at

$$\hat{Y}^{(+p)} = \hat{Y} \frac{1 + e^{a_0 + a_1 X}}{1 + e^{a_0 + a_1 X} e^{a_1 p}}.$$

In conclusion, the modification of variable X by p units implies the average modification of the variable Y induced by multiplying the latter with the factor $\frac{1 + e^{a_0 + a_1 X}}{1 + e^{a_0 + a_1 X} e^{a_1 p}}$.

7. In the case of the *logistic model* $\hat{Y} = \frac{1 + e^{a_0 + a_1 X}}{c},\ c \neq 0$ the modification of variable X by p units leads to

$$\hat{Y}^{(+p)} = \frac{1 + e^{a_0 + a_1 (X+p)}}{c},\ c \neq 0$$

and

$$\frac{\hat{Y}^{(+p)}}{\hat{Y}} = \frac{1 + e^{a_0 + a_1 X} e^{a_1 p}}{1 + e^{a_0 + a_1 X}}$$

and finally

$$\hat{Y}^{(+p)} = \hat{Y} \frac{1 + e^{a_0 + a_1 X} e^{a_1 p}}{1 + e^{a_0 + a_1 X}}.$$

In conclusion, the modification of variable X by p units implies the average modification of the variable Y induced by multiplying the latter with the factor $\frac{1+e^{a_0+a_1X}e^{a_1p}}{1+e^{a_0+a_1X}}$.

3. Solving the inverse problem of simple regression models

In the general case, the inverse problem of a simple regression model is: *how much must factorial variable X be modified by such that the target variable Y will take the preset value $\tilde{Y}$?* If we mark by ε the value with which the factorial variable $X$is modified, the problem is to determine that value of ε such that $\tilde{Y} = \hat{f}(a_0, a_1, X+\varepsilon)$ for every considered model.

1. For the *linear model* $\hat{Y} = a_0 + a_1X$, one must determine the value ε from the equation $\tilde{Y} = a_0 + a_1(X+\varepsilon)$. Because $\tilde{Y} = \hat{Y} + a_1\varepsilon$ we obtain the solution $\varepsilon = \frac{\tilde{Y}-\hat{Y}}{a_1}$.

2. In the case of the *logarithmic model* $\hat{Y} = a_0+a_1\log_a X$, if variable Y takes the preset value $\tilde{Y}$ then we must determine the value of ε from the equation $\tilde{Y} = a_0+a_1\log_a(X+\varepsilon)$. Because $\tilde{Y} - \hat{Y} = \log_a\left(\frac{X+\varepsilon}{X}\right)^{a_1}$, i.e. $\tilde{Y} - \hat{Y} = \log_a\left(1+\frac{\varepsilon}{X}\right)^{a_1}$, we obtain the equation $\left(1+\frac{\varepsilon}{X}\right)^{a_1} = a^{\tilde{Y}-\hat{Y}}$ which is equivalent to the equation $1+\frac{\varepsilon}{X} = a^{\frac{\tilde{Y}-\hat{Y}}{a_1}}$ that has the solution $\varepsilon = X\left(a^{\frac{\tilde{Y}-\hat{Y}}{a_1}}-1\right)$. Another solution for this case is obtained from the equation $\tilde{Y} = a_0 + a_1\log_a(X+\varepsilon)$ that is equivalent to the equation $\tilde{Y} - a_0 = \log_a(X+\varepsilon)^{a_1}$ from which we deduce that $(X+\varepsilon)^{a_1} = a^{\tilde{Y}-a_0}$, i.e. $X+\varepsilon = a^{\frac{\tilde{Y}-a_0}{a_1}}$, and $\varepsilon = -X + a^{\frac{\tilde{Y}-a_0}{a_1}}$.

3. In the case of the *hyperbolic model* $\hat{Y} = a_0 + a_1\frac{1}{X}$, the inverse problem consists in determining ε from the equation $\tilde{Y} = a_0 + a_1\frac{1}{X+\varepsilon}$ that is equivalent to the equation $\left(\tilde{Y}-a_0\right)(X+\varepsilon) = a_1$ and has the solution $\varepsilon = \frac{a_1}{\tilde{Y}-a_0} - X$.

4. In the case of the *power model* $\hat{Y} = a_0X^{a_1}$ if variable Y must take the preset value $\tilde{Y}$ then one must determine the value of ε from the equation $\tilde{Y} = a_0(X+\varepsilon)^{a_1}$ that is equivalent to the equation $\frac{\tilde{Y}}{a_0} = (X+\varepsilon)^{a_1}$ that has the solution $\varepsilon = \left(\frac{\tilde{Y}}{a_0}\right)^{\frac{1}{a_1}} - X$.

5. In the case of the exponential model $\hat{Y} = e^{a_0+a_1X}$, the inverse problem consists in determining ε from the equation $\tilde{Y} = e^{a_0+a_1(X+\varepsilon)}$. But $\tilde{Y} = e^{a_0+a_1(X+\varepsilon)} = e^{a_0+a_1X}e^{a_1\varepsilon} = \hat{Y}e^{a_1\varepsilon}$, i.e. $e^{a_1\varepsilon} = \frac{\tilde{Y}}{\hat{Y}}$, from which we get $\varepsilon = \frac{\ln\tilde{Y}-\ln\hat{Y}}{a_1}$.

6. For the *logistic model* $\hat{Y} = \frac{b}{1+e^{a_0+a_1X}}$ if variable Y must take the preset value $\tilde{Y}$ then the value of ε can be determined from the equation $\tilde{Y} = \frac{b}{1+e^{a_0+a_1(X+\varepsilon)}}$, that is equivalent to $\frac{b}{\tilde{Y}} - 1 = e^{a_0+a_1X+a_1\varepsilon}$ from where we deduce that $\ln\left(\frac{b}{\tilde{Y}}-1\right) = a_0 + a_1X + a_1\varepsilon$ and because $\ln\left(\frac{b}{\hat{Y}}-1\right) = a_0 + a_1X$ we obtain that $\ln\left(\frac{b}{\tilde{Y}}-1\right) - \ln\left(\frac{b}{\hat{Y}}-1\right) = a_1\varepsilon$, and $\varepsilon = \frac{\ln\left(\frac{b}{\tilde{Y}}-1\right)-\ln\left(\frac{b}{\hat{Y}}-1\right)}{a_1}$.

7. For the *logistic model* $\hat{Y} = \frac{1+e^{a_0+a_1X}}{c}$, $c \neq 0$ if variable Y must take the preset value $\tilde{Y}$ then the value of ε can be determined from the equation $\hat{Y} = \frac{1+e^{a_0+a_1(X+\varepsilon)}}{c}$ that is equivalent to $c\tilde{Y}-1 = e^{a_0+a_1X+a_1\varepsilon}$ from where we deduce that $\ln\left(c\tilde{Y}-1\right) = a_0+a_1X+a_1\varepsilon$ and because $\ln\left(c\hat{Y}-1\right) = a_0+a_1X$ we obtain that $\ln\left(c\tilde{Y}-1\right)-\ln\left(c\hat{Y}-1\right) = a_1\varepsilon$, and $\varepsilon = \frac{\ln\left(c\tilde{Y}-1\right)-\ln\left(c\hat{Y}-1\right)}{a_1}$.

4. The implementation of simple regression models

The implementation of a simple regression model requires the creation of scenarios in order to adopt informed decisions. The accomplishment of this objective is possible

only through solving the two problems: the direct and the inverse problem attached to the model. For example, if regarding n territorial units $U_1, U_2, \cdots, U_n$, the target variable Y represents the *rate of economic dependency*, the factorial variable X represents *the workforce employment rate* and the optimal regression model is the power model $\hat{Y} = a_0 X^{a_1}$ then:

1) – ***by solving the direct problem*** we deduce that the rate of economic dependency from the territorial unit U_i, $i = \overline{1,n}$ is modified on average $\left(1 + \frac{p_i}{x_i}\right)^{a_1}$ times, when the workforce employment rate x_i grows by p_i; if $\frac{\hat{y}_i^{(+p_i)}}{\hat{y}_i} = \left(1 + \frac{p_i}{x_i}\right)^{a_1} < 1$, then $\hat{y}_i^{(+p_i)} < \hat{y}_i$ which highlights the fact that a growth of the workforce employment rate, generally implies, a decrease of the rate of economic dependency; we also need to remember that $x_i + p_i$ must not exceed the limits of the interval $[x_{\min}, x_{\max}]$ where $x_{\min} = \min\limits_{i=\overline{1,n}} \{x_i\}$, and $x_{\max} = \max\limits_{i=\overline{1,n}} \{x_i\}$;

2) – ***by solving the inverse problem*** we determine by how much the workforce employment rate must rise in the territorial unit i, $i = \overline{1,n}$ such that the rate of economic dependency drops with an order of size k_i, $k_i > 0$? The solution of the problem is $\varepsilon_i = \left(\frac{\tilde{y}_i}{a_0}\right)^{\frac{1}{a_1}}$, where $\tilde{y}_i = y_i - k_i$, on condition that $\tilde{y}_i$ is not smaller than the lower limit of the confidence interval for the mean of $\hat{y}_i$.

5. Conclusions

1. If $\tilde{Y} = \hat{Y}^{(+p)}$, then for every analyzed model the solution of the inverse problem is $\varepsilon = p$, and this result confirms the fact that the two problems, direct and inverse, have been formulated and solved correctly. For example, for the logarithmic model $\hat{Y} = a_0 + a_1 \log_a X$ the solution of the direct problem is $\hat{Y}^{(+p)} = \hat{Y} + a_1 \log_a \left(\frac{X+p}{X}\right)$, and a solution to the inverse problem is $\varepsilon = X\left(a^{\frac{\tilde{Y}-\hat{Y}}{a_1}} - 1\right)$ and if we assume that $\tilde{Y} = \hat{Y}^{(+p)}$, then the solution to the inverse problem is

$$\varepsilon = X\left(a^{\frac{\hat{Y}^{(+p)}-\hat{Y}}{a_1}} - 1\right) = X\left(a^{\frac{\hat{Y}+a_1\log_a\left(\frac{X+p}{X}\right)-\hat{Y}}{a_1}} - 1\right) =$$

$$= X\left(a^{\log_a\left(\frac{X+p}{X}\right)} - 1\right) = X\left(\frac{X+p}{X} - 1\right) = p.$$

2. The types of simple regression models analyzed in this article are the most commonly used ones, but we also need to state the fact that other types of regression models also exist and implementing these latter models through solving the direct and inverse problems associated with them is extremely useful and constitutes an essential analysis tool.

References

[1] Harrell, F. E. Jr., *Regression Modeling Strategies: With Applications to Linear Models*, Logistic Regression and Survival Analysis, Springer, 2006.

[2] Isaic-Maniu, Al., Mitruț, C., Voineagu, V., *Statistică*, Editura Universitară, București, 2003.

[3] Kleinbaum, David G., Klein, Mitchel, *Logistic Regression*, Springer, 2010.

[4] Pecican, E. Șt. *Econometrie*, Editura All, București, 1994.

[5] Panaretos, J., `http://www.stat-athens.aueb.gr/~jpan/diatrives/Tsiptsis/chapter3.pdf`.

[6] Powers, Daniel A., Xie, Yu, *Statistical Methods for Categorical Data Analysis*, Emeral Group Publishing Limited, 2008.

[7] Rencher, A. C., *Linear Models in Statistics*, Wiley-Interscience, 1999.

[8] von Storch, Hans, Zwiers F. W., *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, UK, 2002.
[9] Xin Yan, Xiao Gang Su., *Linear Regression Analysis – Theory and Computing*, World Scientific Publishing Co. Pte. Ltd. 2009.
[10] Zăvoianu, Felicia, *Models for evaluating the impact of economic restructuring at a territorial level*, Doctoral Thesis, A.S.E. București, 2007.

University of Petroșani
Department of Mathematics and Computer Science
Universității 20, 332006 Petroșani, România
E-mail address: fzavoianu@yahoo.com

University of Petroșani
Department of Mathematics and Computer Science
Universității 20, 332006 Petroșani, România
E-mail address: constantin.zavoianu@yahoo.com