PERIODIC REGRESSION

K. Čobanović, Z. Lozanov-Crvenković and E. Nikolić-Đorić
University of Novi Sad, Serbia and Montenegro
katcob@polj.ns.ac.yu

Periodic regression is seldom included in syllabus of statistical courses. However, the data following periodic or cyclic behavior are often encountered, especially in agriculture. Therefore, we think that this type of regression should be taught to students of agriculture, even in basic courses of statistics. In the paper we propose the way of teaching periodic regression through examples usually encountered in practice. The analysis of data will be based on the graphical interpretation, which would provide the visual display of the investigated problems as well.

INTRODUCTION

Many biological and agricultural series are characterized by seasonal variations. Periodic phenomena are primarily close not only to biological data, but to non-biological data as well. Periodic or cyclic phenomena are characteristic of many different types of data, which are synchronized with daily, lunar, or annual changes (Bliss, 1970). Many kinds of agricultural data tend to fluctuate up and down at regular time intervals showing periodic character (Little and Hills, 1978). The periodic or cyclic character of many phenomena in natural environment is expressed in time and in space. Many cyclical and seasonal variations of biological and other data occur depending on daily, monthly, yearly or other cycle's changes. In biology and agriculture, too, periodic and cyclic variations could be identified in a number of investigated problems (Bliss, 1970; Little and Hills, 1975; Marko and Nikolić, 1982).

The biological cycles could be divided into two categories: those which do not strictly depend on physical environment and those which strictly depend on physical environment (Bliss, 1970). In the first category the cycles are determined physiologically. In the second category the cycles depend on many potential factors. The cycle period in a natural environment, concerning animals or plants, may be of different length.

The authors have had in mind that "Periodic type is a curve that relates some variable to time and is repeated at fixed time intervals. It is known in mathematical texts as a Fourier curve and is useful for any kind of data that tends to fluctuate up and down at regular intervals. Very few statistics texts discuss fitting data of this kind, but we have found it so useful for many kinds of agricultural data, that we will give a brief outline of the general method" (Little and Hills, 1978). Using the advantages of computer techniques today, it is very easy to apply periodic regression, which is of the type of polynomial curve. In such situation, in teaching statistics, it would be simple to apply regression methods on complicated trigonometric functions, expressing wave forms, having in mind that the method of fitting a Fourier curve is very similar to the method for fitting a polynomial regression (Little and Hills, 1978). Many statistical courses include basic regression methods, which could be a solid base for introduction to the analysis of the cyclical and seasonal variations of time series data in the teaching process. Thus, students could easily apply their already acquired knowledge in a specific case of periodic regression.

Some authors suggest other methods in the analysis of cyclical and seasonal variations of explored data, like the use of dummy variable (Hebden, 1981; Čobanović and Lučić, 1992). But, the authors in this paper have found that the simplified trigonometric functions for data series with periodicity could be useful in the process of teaching Statistics. At the Faculty of Agriculture, University of Novi Sad, at the study group of Agriculture Economy, the data with periodic character which need to be analyzed are often encountered. Therefore, in this paper we wish to emphasize the need of using some specific methods of analyzing the cyclical and seasonal data usually encountered in agricultural practice. One of these methods is the use of periodic regression based on Fourier curve. We propose the way of teaching periodic regression through examples with real data from agriculture. Data series in examples are related to daily, monthly and yearly variations. The daily variations present data of the value of UV Index (ultraviolet) radiation during the period of three years; the monthly variations represent the values of price index, concerning the prices of agricultural products during the period of five years; the yearly variations represent the values of wheat yield (t/ha) during the period of 57 years.

METHODS

In the analysis of the periodicity the most important is the definition of three main parameters: the length of the cycle or fundamental period; its amplitude or the range from the minimal to the maximal response and the phase angle or angular point in time during the cycle when the response is maximal (Bliss, 1970). These parameters can be easily estimated using any statistical software. Also, F-test and t-test for testing the goodness of fit can be performed. Here we use *STATISTICA* 7.0 (University license, University of Novi Sad).

A time series $Y_t\,(t=1,...,N)$ observed at equal intervals of time and may be expressed as: $Y_t=\hat{Y}_t+\varepsilon_t$, where $\hat{Y}_t$ is unobserved fixed value at time t and $\{\varepsilon_t\}$ is a sequence of random errors identically and independently distributed with expectation 0 and variance σ^2. To determine whether the variability of the time series has periodic components, the series is approximated by finite Fourier series of the form, if the number of data is even: $N=2n$,

$$\hat{Y}_t=A_0+2\sum_{m=1}^{n-1}(A_m\cos 2\pi m f_1 t+B_m\sin 2\pi m f_1 t)+A_n\cos 2\pi n f_1 t,\ \text{or}$$

$$\hat{Y}_t=A_0+2\sum_{m=1}^{n-1}(A_m\cos 2\pi m f_1 t+B_m\sin 2\pi m f_1 t),$$

if the number of data is odd: $N=2n-1$.

Here $R_m=\sqrt{A_m^2+B_m^2}$ is the amplitude, and $\phi_m=arctg(B_m/A_m)$ is the phase of the i-th component. The function $\hat{Y}_t$ is a linear combination of sinus and cosinus functions with frequencies proportional to fundamental frequencies $f_1=1/N$, so it is linear multiple regression with sinus and cosinus functions as regressors. Since $\frac{1}{N}\sum_{t=1}^{N}Y_t^2=R_0^2+2\sum_{m=1}^{n-1}R_m^2+R_n^2$, a contribution of i-th harmonical component to the mean of the total sum of squares of time series is equal to R_i^2. By decomposing the mean of the total sum of squares it is possible to single out harmonical components which describe the series well.

With additional assumption that errors are normally distributed, the estimated periodic regression model may be tested by F-test and particular estimates of parameters with t-test.

RESULTS

Example 1: The first examined time series contained daily measurements of maximal UV index, made in period 25.04.2003. –29.06.2005. at the Department of Physics, University of Novi Sad in the Vojvodina Province, Serbia and Montenegro. Line plot of time series suggested cyclic behavior with the constant periodicity of approximately one year ($f_1=1/365$). The periodic regression may be considered as a special case of multiple regression with time series (Y_t, $t=1,...,615$) as dependent and two regressors $X_{1t}=\sin(2\pi f_1 t)$ and $X_{21t}=\cos(2\pi f_1 t)$ as independent variables. Ordinary least squares estimates of unknown regression coefficients, standard errors of coefficients, t-values, F-value, coefficients of multiple correlation, determination, adjusted coefficient of determination and standard error of the regression were calculated by means of program *STATISTICA* 7 and presented in Table 1. Special attention should be given to low values of UV Index during the summer day, due to cloudy days. It should be pointed out to students that better fit could be obtained by including the dummy variable.

Table 1: Results of periodic regression of UV data

N=615	Regression Summary for Dependent Variable: Y R= .90421869 R²= .81761143 Adjusted R²= .81701539 F(2,612)=1371.7 p<0.0000 Std.Error of estimate: 1.2218					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(612)	p-level
Intercept			3,920087	0,050101	78,24385	0,0000
sin(2*pi*t*f1)	0,789134	0,017342	3,272730	0,071922	45,50395	0,0000
cos(2*pi*t*f1)	0,372651	0,017342	1,490214	0,069350	21,48823	0,0000

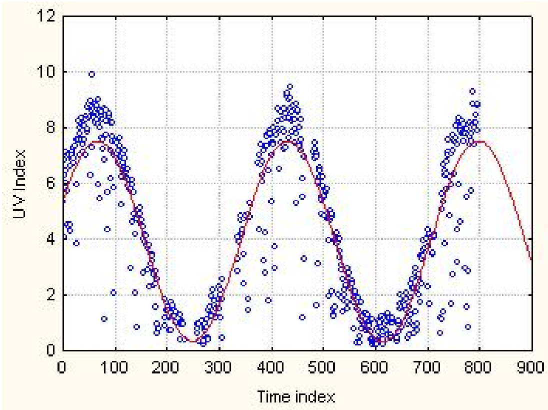

Figure 1: Periodic regression of UV data

Example 2: As an example of monthly data, the time series of agricultural production indices of prices in period 1999-2004, with the base period (2004=100) for Serbia, was considered. Line diagram of time series suggested seasonal oscillations of indices around quadratic trend. So the adequate model for trend and seasonal movements is the linear combination of polynomial of second degree and periodic components. It may be shown that fitting of model will be improved by the introduction of trigonometric components with period of 6 months. See Table 2 and Fig. 2.

Example 3: Time series of yearly wheat data (t/ha) in the period 1947-2002 is characterized by cyclic variations around third degree polynomial trend. The dominant frequency was obtained by means of periodogram applied on detrended time series. The dominant frequency in this case is 0.2857. The OLS coefficients of regression model that includes variables selected by application of step-wise regression are given in Table 3.

Table 2: Results of periodic of agricultural production indices of prices regression

N=72	Regression Summary for Dependent Variable: Y R= .97903602 R²= .95851153 Adjusted R²= .95468182 F(6,65)=250.28 p<0.0000 Std.Error of estimate: 5.8190					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(65)	p-level
Intercept			10,36715	2,123943	4,8811	0,000007
t	2,25428	0,102516	2,94431	0,133896	21,9895	0,000000
t2	-1,38139	0,102449	-0,02395	0,001776	-13,4837	0,000000
sin(2*Pi*t*1/12)	0,07824	0,025472	3,00363	0,977816	3,0718	0,003106
cos(2*Pi*t*1/12)	-0,02447	0,025287	-0,93935	0,970723	-0,9677	0,336790
sin(2*Pi*t*1/6)	-0,05781	0,025309	-2,21922	0,971600	-2,2841	0,025641
cos(2*Pi*t*1/6)	0,06112	0,025280	2,34621	0,970383	2,4178	0,018426

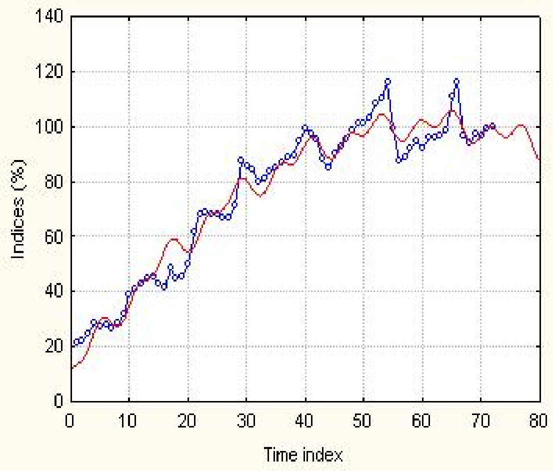

Figure 2: Periodic regression of agricultural production wheat indices of prices data

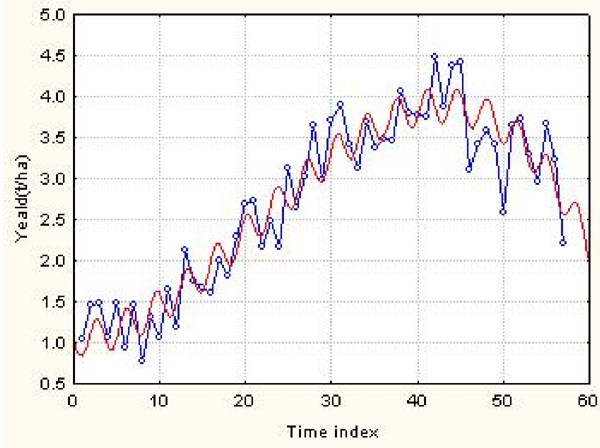

Figure 3: Polynomial-periodic regression of data

Table 3: Results of polynomial-periodic regression of wheat data

N=57	R= .94889712 R²= .90040575 Adjusted R²= .89476834 F(3,53)=159.72 p<0.0000 Std.Error of estimate: .33750					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(53)	p-level
Intercept			1.046239	0.090675	11.5384	0.000000
t2	4.42319	0.261654	0.004633	0.000274	16.9047	0.000000
t3	-3.78387	0.261708	-0.000072	0.000005	-14.4584	0.000000
sin(2*Pi*f1*t)	-0.15030	0.043417	-0.217534	0.062839	-3.4618	0.001069

CONCLUSION

This paper illustrates that the periodic regression is a simple and flexible method for the detection of cyclical and seasonal pattern in time series and should be included in statistical courses for students of agriculture, economy and biology. This kind of model can represent time series with a small number of parameters which is of great importance especially in the case when the time series is not monotonic and stationary and contains a nonlinear trend, cyclical and seasonal components with different periodicity. As periodic regression is a special case of multiple regression, it can be performed at any commonly used statistical software. This method is not adequate for time series with varying amplitudes, frequencies and phases. Time series of sun spot numbers is the example of time series with non-symmetric cycles with variable amplitudes.

REFERENCES

Bliss, C. I. (1970). *Statistics in Biology*. New York: McGraw-Hill Book Company.

Čobanović, K. and Lučić, Đ. (1992). The investigation of seasonality in milk production using dummy variable. *Zbornik radova, Savez inženjera i tehničara Jugoslavije*.

Hebden, J. (1981). *Statistics for Economists*. Oxford: Philip Allan, Publishers Limited.

Little, T. M. and Hills, F. J. (1978). *Agricultural Experimentation Design and Analysis*. New York: John Wiley and Sons, Inc.

Marko, J. and Nikolić, E. (1982). Characteristics of yield development of the main field crops in SAP of Vojvodina. *Contemporary Agriculture,* 30(2), 87-98.

STATISTICA 7.0. StatSoft, Inc. University License

Statistical Yearbook of Serbia 2003. Statistical Office of Serbia, Belgrade.