Colloquium Biometricum 41
2011, 27–38

VALUATION OF INFLUENCE OF CONCOMITANT VARIABLES ON ESTIMATION OF POLYNOMIAL GROWTH CURVES

Andrzej Bochniak, Mirosława Wesołowska–Janczarek

Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Akademicka 13, 20–950 Lublin, Poland
e–mail: andrzej.bochniak@up.lublin.pl; miroslawa.wesolowska@up.lublin.pl

Dedicated to the memory of Professor Wiktor Oktaba

Summary

Frequently in practice the estimation of polynomials which describe course of changes for a studied feature in a given time interval can be significantly disturbed by some concomitant variables with values changing in time. The situation where values of concomitant variables in successive time points are the same for all experimental units is considered. The influence of these variables on estimation of polynomials by two methods: Potthoff–Roy's and iterative is examined. The investigation is carried out on the data obtained by computer simulation.

Key words and phrases: growth curve methods, concomitant variables, linear mathematical models, parameter estimation methods, computer simulation

Classification AMS 2010: 62J12, 68U20

1. Introduction

The course of change of a studied feature in time for different groups of units can be described using the known growth curve method given by Potthoff and Roy (1964). Often different variables, called concomitant variables, whose values changes in time can have influence on studied feature values. For example: if a studied feature is the growth of plants in time, then concomitant variables can be the rainfall value and the air temperature in suitable time points. In the case considered here all values of those variables are the same for all plants.

Different models which take concomitant variables into account were presented by Wesołowska–Janczarek (2009). One of these models, which assumes the same influence of concomitant variables for all experimental units, was provided in the work by Wesołowska–Janczarek and Fus (1996). It also presents an iterative method of parameter estimation in that model.

It is interesting whether influence of concomitant variables ought to be ever taken into consideration while estimation the course of feature changes or the influence of concomitant variable can be omitted. Some discussion about this problem was presented by Bochniak and Wesołowska–Janczarek (2010) in the paper where the influence of degree of variation of concomitant variables values in time on the conformity of estimated function to true one was studied.

On the base of simulated data, where different influence of concomitant variables was assumed, comparison of results of Potthoff–Roy's and iterative methods was carried out.

2. Considered models and suitable estimation methods

The first model to be presented is Potthoff and Roy's (1964) one that does not contain concomitant variables. It is a multivariable model presented in the following form:

$$\mathbf{Y} = \mathbf{ABT} + \mathbf{E}, \tag{2.1}$$

where **Y** is $n \times p$ – matrix of observations of feature on n experimental units in p time points, **A** is $n \times a$ – known matrix which divides experimental units on a group, **B** – is $a \times q$ – matrix of unknown coefficients in searched polynomial growth curves of q–1 degree, **T** is $q \times p$ matrix that include the successive powers of time points from 0 to q–1 (it is Vandermonde's matrix) that defines internal structure of observations and **E** is a $n \times p$ matrix of random errors. If all units are homogeneous then $\mathbf{A} = \mathbf{J}_n$, where $\mathbf{J}_n$ is a vector of n ones, but if observations are subject to two way classification then matrix **A** is a non full rank. To continue our considerations in this paper, matrix **A** is taken as in a one way classification without the column of ones.

This model is considered under assumption that rows of matrix **Y** are uncorrelated, but columns are correlated with common covariance matrix $\underset{pp}{\mathbf{\Sigma}}$. Additionally, a matrix of observations is a multivariate normally distributed. This assumption can be presented as $\mathbf{Y} \sim \mathbf{N}_{n,p}(\mathbf{ABT}, \mathbf{I}_n \otimes \underset{pp}{\mathbf{\Sigma}})$, where $\mathbf{I}_n$ is a unit matrix of n dimension (see Gupta and Nagar, 1999, p. 55).

Estimators of parameters in this model that are coefficients in polynomials in matrix **B** and covariance matrix $\mathbf{\Sigma}$ obtained by maximum likelihood method (Kshirsagar, 1988) are given in following form:

$$\hat{\mathbf{B}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}'(\mathbf{T}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}')^{-1} \tag{2.2}$$

and

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\mathbf{Y}'\left[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right]\mathbf{Y}. \tag{2.3}$$

One of the growth curve models with concomitant variables, when the values of *s* concomitant variables are the same for all experimental units, but each of the variables values are different in observed *p* time points was given by Wesołowska–Janczarek and Fus (1996) in the following form:

$$\mathbf{Y} = \mathbf{ABT} + \mathbf{J}_n\boldsymbol{\gamma}'\mathbf{X} + \mathbf{E}, \tag{2.4}$$

where the matrices **Y**, **A**, **B**, **T** and **E** are the same as in the model (2.1), **X** is $s \times p$ matrix of values of these *s* variables in successive *p* time points, $\boldsymbol{\gamma}$ is a vector of *s* regression coefficients at concomitant variables, $\mathbf{J}_n$ is a vector of *n* ones and **E** is a $n \times p$ matrix of random errors.

Under the assumption of matrix variate normal distribution of **Y** denoted by $\mathbf{Y} \sim \mathbf{N}_{n,p}(\mathbf{ABT} + \mathbf{J}_n\boldsymbol{\gamma}'\mathbf{X}, \mathbf{I}_n \otimes \mathbf{\Sigma})$ and $\mathbf{\Sigma}(p \times p) > 0$ assumptions estimators of parameters in this model obtained by maximum likelihood method were given in following form:

$$\begin{aligned}
n\hat{\mathbf{\Sigma}} &= (\mathbf{Y} - \mathbf{A}\hat{\mathbf{B}}\mathbf{T} - \mathbf{J}_n\hat{\boldsymbol{\gamma}}'\mathbf{X})'(\mathbf{Y} - \mathbf{A}\hat{\mathbf{B}}\mathbf{T} - \mathbf{J}_n\hat{\boldsymbol{\gamma}}'\mathbf{X}) \\
\hat{\mathbf{B}}_{\hat{\Sigma}} &= (\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'(\mathbf{Y} - \mathbf{J}_n\hat{\boldsymbol{\gamma}}'\mathbf{X})\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}'(\mathbf{T}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}')^{-1} \\
\hat{\boldsymbol{\gamma}}'_{\hat{\Sigma}} &= [\mathbf{J}'_n\mathbf{Y} - \mathbf{J}'_n\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{Y}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}'(\mathbf{T}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}')^{-1}\mathbf{T}]\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}'\mathbf{R}_{\hat{\Sigma}} \\
\mathbf{R}_{\hat{\Sigma}} &= [n\mathbf{X}\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}' - \mathbf{J}'_n\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{J}_n\mathbf{X}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}'(\mathbf{T}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}')^{-1}\mathbf{T}\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}']^{-1}.
\end{aligned} \tag{2.5}$$

The values of these estimators can be calculated by the iterative method where in the first step the following form of a matrix will be taken $n\hat{\mathbf{\Sigma}} = \mathbf{Y}'[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{Y}$.

We are interested in the answer to the question when the concomitant variables must be considered in the model and when they can be omitted. This is examined by comparison the fitting of estimated polynomials given by two mentioned method with assumed ones. The investigations presented here were undertaken using the data obtained by a computer simulation. Details about the method of simulation is presented in next part of this paper.

3. Computer simulation

The computer simulation was conducted using own procedures and some built–in functions programmed in *Matlab*. The values of observations obtained in the real experiment were taken as the basis for the computer simulation. In the experiment described by Wesołowska–Janczarek and Fus (1996) the yielding of 16 raspberry varieties was examined where values of meteorological elements such as air temperature, sunshine and rainfall were taken as concomitant variables.

The values of parameters, such as covariance matrix $\mathbf{\Sigma}$, matrix $\mathbf{B}$ of polynomials coefficient, averages and standard deviations of concomitant variables given by matrix $\mathbf{X}$ which appear in the models given by equations (2.1) and (2.4), were estimated from original data.

In the second step on the basis of estimators calculated in the first step new polynomials (exactly coefficients matrix $\mathbf{B}$) representing different varieties were generated assuming similar course of yielding. Next, for such polynomials new values (10 000 times) were generated for required parameters: positive defined matrix $\mathbf{\Sigma}$, values of matrix $\mathbf{X}$ containing information about concomitant variables, vector $\boldsymbol{\gamma}$ of regression coefficients and finally observation matrix $\mathbf{Y}$ by rows from distribution $\mathbf{N}_{n,p}(\mathbf{ABT} + \mathbf{J}_n\boldsymbol{\gamma}'\mathbf{X};\ \mathbf{I}_n \otimes \mathbf{\Sigma})$. Estimators of polynomials were next calculated and compared with assumed ones.

Such simulation was repeated for different variability in concomitant variables values (matrix $\mathbf{X}$) and in observation matrix $\mathbf{Y}$ (changed by matrix $\mathbf{\Sigma}$). The influence of concomitant variables was changed during the computer simulation aiming to establish cases in which iterative method gives better estimators than Potthoff–Roy′s method.

Arrangement of time points and matrix $\mathbf{A}$ dividing experimental units into groups were kept unchanged in comparison to the conducted experiment. The values of **X** or **Y** were generated assuming similarity to the values from experimental data.

4. Results

The results obtained in all conducted simulations are similar to those presented in the examples in further part of paper. All the analysis are shown using the same 16 polynomials assumed in simulations. The shapes of polynomials are shown on fig. 1. As Bochniak and Wesołowska–Janczarek (2010) suggested the shape of polynomial has also influence on estimation, so the most (variety 15th) and the least bent (variety 9th) polynomials are specially marked with wider lines.

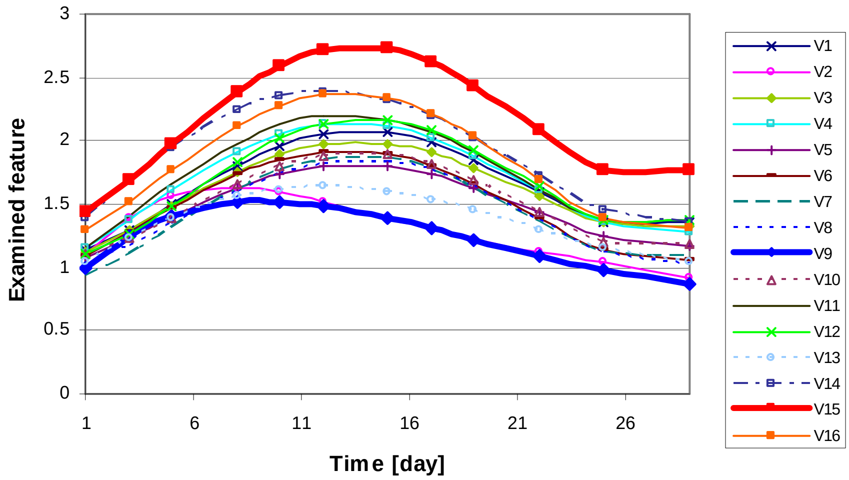


Fig. 1. Exemplary polynomials assumed in simulation

For such polynomials new data (**Y**, **X**, **Σ)** was generated 10 000 times using different random regression vectors **γ** which determine the influence of concomitant variables. The values of mentioned parameters were similar to original experimental data. Estimators given by (2.2) and (2.5) were used to calculate coefficients of matrix **B**, and obtained polynomials were compared with assumed ones. Graphical comparison of estimators for the first 100 iterations of simulation given by both methods is presented on fig.2. Dark line correspond to assumed polynomial for variety 15th and lighter lines show the first 100 estimators of the given polynomial from all 10 000 repetitions of simulation. It is visible that polynomials given by Pothoff–Roy's method is much more scattered from assumed line.

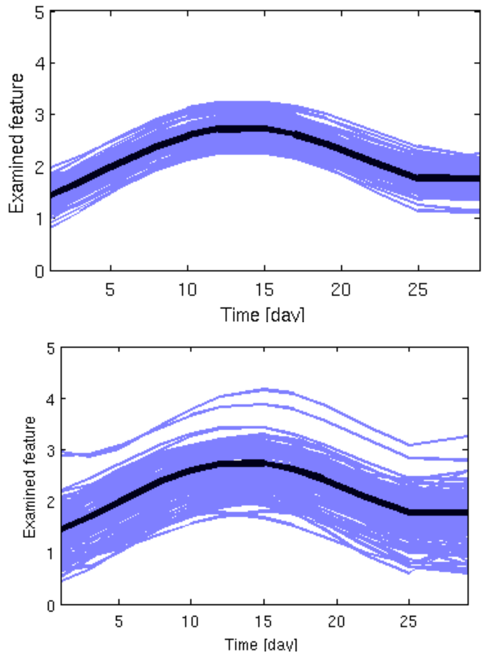


Fig. 2. Comparison of polynomials estimated by iterative (top) and Potthoff–Roy′s (bottom) methods with assumed growth curve

Fig. 3 presents an exemplary generated polynomial for some variety, curve with considered concomitant variables, generated observations for 4 replication and finally estimators of polynomial growth curve calculated by Potthoff–Roy's and iterative methods. In this case estimated polynomial by iterative method is closer to assumed one than given by Potthoff–Roy's method. Iterative method gives larger values for variety polynomial and concomitant variables cause lowering cumulative curve to fit to observations. In other cases the situation is varying.

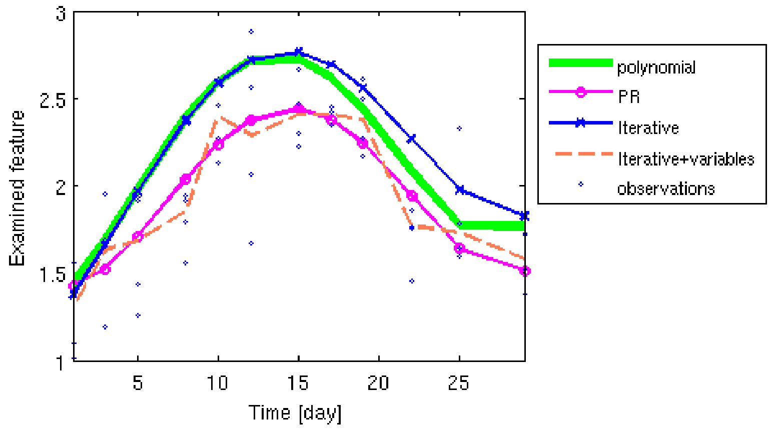


Fig. 3. Generated observations and estimations of assumed curve by iterative and Potthoff–Roy's methods for example simulation

For each of 10 000 repetitions of simulation two polynomials were estimated using Potthoff–Roy's method and iterative one. For each estimated polynomials relative error, which was averaged in all time points, was calculated using following formula:

$$err_{ij} = \frac{1}{p} \cdot \sum_{t=1}^{p} \frac{\left| Q_{ij}(t) - P_i(t) \right|}{P_i(t)} \tag{4.1}$$

where

P_i – assumed polynomial for i–th variety,

Q_{ij} – estimated polynomial for i–th variety and j–th repetition,

p – number of time points.

Polynomials estimated by both methods were compared using formula (4.1) and it was counted which method gives most frequently better results. Fig. 4 presents percentages (vertical axis) in which iterative method gives better estimators of polynomials in dependence of relative combined influence of concomitant variables on assumed polynomials.

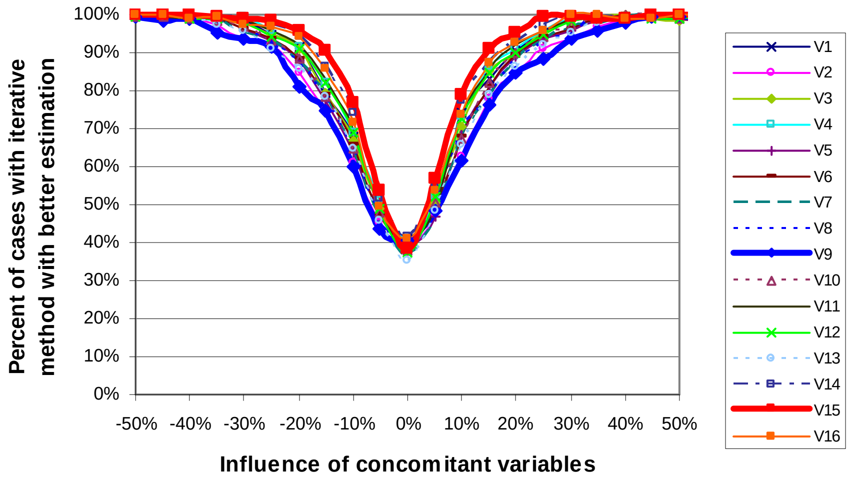


Fig. 4. Percentages of cases in which iterative method gives better results with dependency of combined influence of concomitant variables (for all 16 varieties)

Because assumed polynomials has different maximum value, for every repetition of simulation (with different values of concomitant values) and for each polynomial the proper relative combined influence of concomitant variables was calculated as average value obtained in each time point (horizontal axis). This influence shows how many percentages concomitant variables increased (positive influence) or decreased (negative influence) values of assumed polynomials. The calculation of the combined influence of concomitant variables $\gamma'\mathbf{X}$ was done with rounding it to precision of 5%. The two wider lines correspond to two polynomials for chosen varieties which are specially marked of fig.1. As one can see they determine limits for all other lines which are connected with polynomials laying between these extreme ones.

In considered situation, the estimators given by Potthoff–Roy's method are better than iterative ones in the cases when influence of concomitant variables is small – approximately in 60% cases if there is no combined influence of concomitant variables. If combined influence of concomitant variables reaches approximately 10% both methods give better estimators with the same frequency. Iterative method is better if influence of concomitant variables is larger (more than 10% of influence). In discussed simulation if relative influence of concomitant values exceed 20–25% of assumed polynomials values, then in 90% cases estimators given by iterative method is better fitted to assumed polynomials. It can also be seen that for variety 15th, which is more diverse in time than variety 9th, iterative method gives faster better results with increasing influence of concomitant values. Some of obtained numbers of cases in which specific method gives better estimation are listed in Table 1. Positive and negative influence is almost symmetrical.

Table 1. Number of cases in which chosen method gives better estimation of assumed polynomial

	Number of cases			
	Variety 9		**Variety 15**	
Variables influence	Iterative better	PR better	Iterative better	PR better
–30%	393	26	340	3
–25%	368	33	492	7
–20%	401	93	587	25
–15%	360	121	659	69
–10%	292	195	646	194
–5%	244	317	482	415
0%	211	309	348	557
5%	256	271	518	391
10%	318	199	618	167
15%	376	116	633	64
20%	351	63	582	29
25%	366	48	480	3
30%	354	24	358	1

In view of facts described by Bochniak and Wesołowska–Janczarek (2010) the iterative method has bad properties if variability of concomitant variables values i.e. the elements of matrix **X** has small differences in successive time points. Fig. 5 presents similar chart examining influence of concomitant variables but separate lines corresponds to simulations with different variability of **X**. Standard deviations of generated data x_{ij} was decreased or increased in comparison to the ones calculated for original experimental data by multiplication by following values: a) 0.1 – the least diversity; b) 0.5; c) 1 – diversity as in the original data; d) 2; e) 5 – the greatest diversity of concomitant variables values. The 9^{th} and 15^{th} varieties are only drawn for easier observation and only positive influence is shown because negative one is symmetrical. It can be easily seen that the lesser variability in concomitant variables values in time, the less exact estimation of the assumed curve in the iterative method with regard to the fixed regression dependence on concomitant variables i.e. is fixed elements in vector $\boldsymbol{\gamma}$.

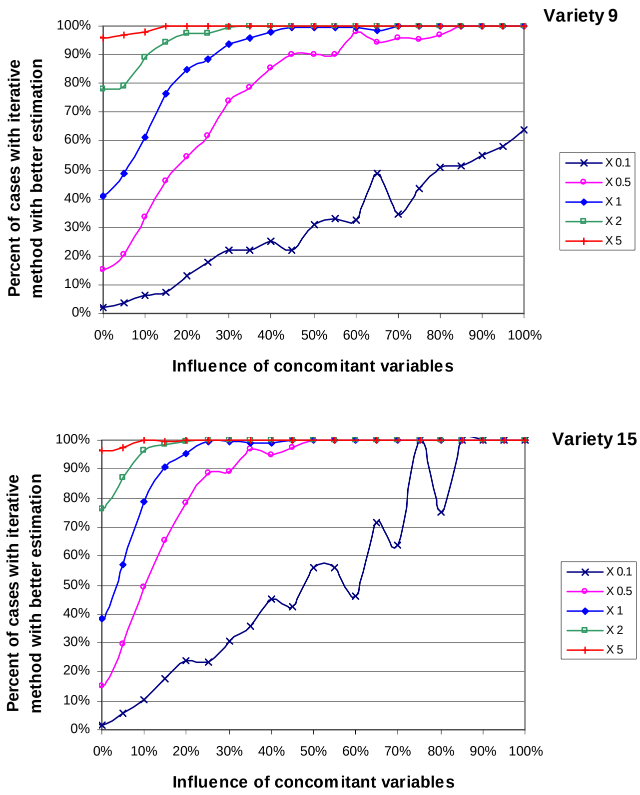
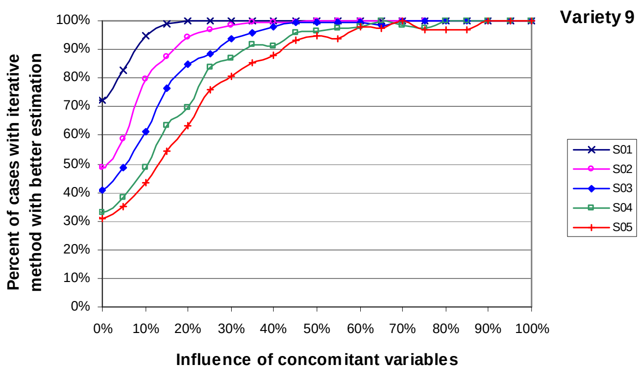


Fig. 5. Percentages of cases in which iterative method gives better results with dependency of influence of concomitant variables and different diversity of **X**

The influence of variability of generated observation values (matrix **Y**) in time was also examined. This diversity was changed by generating smaller or greater elements in covariance matrix **Σ**. The results of this study is shown on fig.6 where this time lines corresponds for different covariance matrix. This matrix was generated under assumption of specific error for single generated element of **Y**. The standard deviation of this error is assumed as: 0.1; 0.2; 0.3; 0.4 and 0.5. Values of concomitant values has the same diversity as in the original data. Also charts only for the two extreme varieties are shown here. The lesser variability in polynomials values in time increases exactness of estimation of the assumed curve in the iterative method.



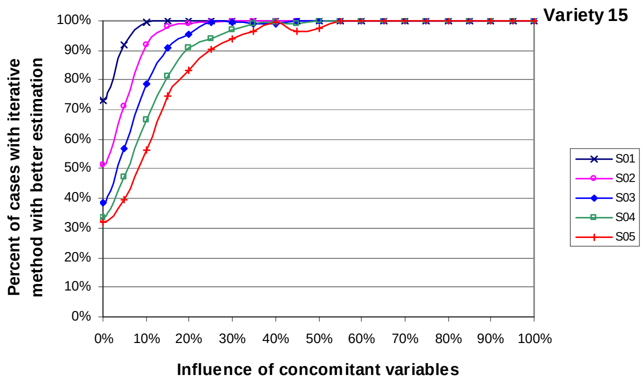


Fig. 6. Percentages of cases in which iterative method gives better results with dependency of influence of concomitant variables and different diversity of $\mathbf{\Sigma}$

5. Conclusions

The paper presents the results obtained in studies which aim to bring the solution of the problem if values of concomitant variables must always be considered and what does the precision of growth curve estimation depend on. The conclusions based on the studies that have been carried out so far are following:

1) Growth curves estimation using iterative method is better than Potthoff–Roy's method if influence of concomitant variables increases (Fig. 4).

2) The exactness of growth curves estimation using iterative method is increased with enlarged difference between concomitant variables values in successive time points (Fig. 5).
3) If concomitant variables values in successive time points are constant or little diverse then iterative method should not be used because of very bad behaviour in some cases (Bochniak and Wesołowska–Janczarek 2010, Fig. 5 X 0.1). Unfortunately the exact reason of this situation has not been solved yet.
4) If polynomial values in successive time points are strongly differentiated, then exactness of curve estimation obtained by iterative method is better (variety 9^{th} and 15^{th} in Fig. 4, Fig.5, Fig.6).
5) Iterative method estimation in comparison to Potthoff–Roy's method enlarges its precision if values of observations for varieties are less diverse (Fig. 6).
6) Further studies are necessary to determine the minimum value of variance of concomitant variables values in time, and possibly to omit these variables in growth curve analysis.

References

Bochniak A., Wesołowska–Janczarek (2010). On influence of variability in concomitant variables values on estimation of polynomial coefficients in growth curves models with concomitant variables changing in time and the same values for all experimental units. *Colloquium Biometricum* 40, 135–145.

Gupta A.K., Nagar D.K. (1999). *Matrix Variate Distributions*. Chapman&Hall/CRS.

Kshirsagar A.M. (1988). A note on multivariate linear models with non–additivity. *Austral. J. Statist.* 30(3), 292–298.

Potthoff R.F., Roy S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problem. *Biometrika* 51, 313–326.

Wesołowska–Janczarek M., Fus L. (1996). Parameters estimation in the growth curves model with time–changing concomitant variables. *Colloquium Biometryczne* 26, 263–277.

Wesołowska–Janczarek M. (2009). Selected models and methods of parameter estimation in growth curves with concomitant variables. *Colloquium Biometricum*, 39, 21–31.