

Merce, Emilian; Drăghici, Manea; Berca, Mihai; Merce, Cristian C.; Necula, Raluca-Alexandra

Conference Paper

Method and program for autocorrelation distribution on influence factors

Provided in Cooperation with:

The Research Institute for Agriculture Economy and Rural Development (ICEADR), Bucharest

Suggested Citation: Merce, Emilian; Drăghici, Manea; Berca, Mihai; Merce, Cristian C.; Necula, Raluca-Alexandra (2018) : Method and program for autocorrelation distribution on influence factors, In: Agrarian Economy and Rural Development - Realities and Perspectives for Romania. 9th Edition of the International Symposium, November 2018, Bucharest, The Research Institute for Agricultural Economy and Rural Development (ICEADR), Bucharest, pp. 10-15

This Version is available at:

<https://hdl.handle.net/10419/205081>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

METHOD AND PROGRAM FOR AUTOCORRELATION DISTRIBUTION ON INFLUENCE FACTORS

EMILIAN MERCE¹, MANEA DRĂGHICI², MIHAI BERCA³,
CRISTIAN C. MERCE⁴, RALUCA-ALEXANDRA NECULA⁵

Summary: *The authors point out that autocorrelation is an accidental statistical phenomenon. The cause of the occurrence of collinearity is the incomplete data base and that ideally, the elimination of the occurrence of autocorrelation is achieved by knowing the effect stage for all possible combinations of the variants of the factors involved. In many cases, the practical realization of such a desideratum is impossible. Such a difficulty is especially specific to statistical processing in the economic, social and psychological field. Neither multi-factorial experiments of agrobiological nature are not avoided by such difficulties. Consequently, to the researcher remain at his disposal methods of distributing collinearity on influence factors using methods based on the calculation of partial correlation coefficients (Merce E., 1986; Moineagu C., 1974). With obvious computing facilities, compared to the evoked methods, the authors suggest using an original method based on the principle of proportional distribution of autocorrelation with the proportion of simple determinations, following the next six steps, the last four are solved instantly after the first two steps have been solved:*

1. *The calculation of the multiple correlation coefficient and simple correlation coefficients using the Regression function of the Data Analysis component of Microsoft Excel Program;*
2. *The recording of the multiple correlation coefficient and of the simple correlation coefficients;*
3. *The simple determination coefficients and the multiple determination coefficient calculation;*
4. *The sum of the simple determination coefficients;*
5. *The calculation of the simple determinations proportions, considering their sum equal to 100;*
6. *The determination of each factor influence, as a product of multiple determination and the proportion of simple determinations.*

Keywords: *autocorrelation, distribution of autocorrelation by factors, method and program*

JEL Classification: *C40*

INTRODUCTION

Collinearity is an objective reality in research of complex causal relationships. It is exteriorized, as illustrated in the literature [5; 6; 7; 10], whenever the causality complex database is incomplete. The presence of collinearity alters the accuracy of numerical determinations between factors, on the one hand, and the effect studied, on the other.

The ideal solution would be to use complete databases. However, this desideratum cannot be always achieved practically because of the complexity of the investigated causal relationships. Here, the research from economics, sociology, psychology, as well as the multi-factorial agrobiological experiments can be nominated. In all these situations, the researcher must assess the collinearity numerically and then proceed to correct the relationship between the factors studied and the effect. In order to achieve this goal, the old working methods [1; 8; 9] based on the calculation of the partial correlation coefficients can be used. The use of these methods is, however, rather cumbersome, requiring separate calculations to be made on a case-by-case basis.

MATERIAL AND METHOD

Whenever the magnitude of the causal complex makes it impossible to organize experiences that include all the possible combinations of the variants of the investigated factors, the

¹ Professor Ph.D., USAMV Cluj-Napoca, emerce@usamvcluj.ro

² Professor Ph.D., USAMV Bucharest, dmprofesor@hotmail.com

³ Professor Ph.D., USAMV Bucharest, prof.mihai.berca@gmail.com

⁴ Assistant Professor Ph.D., USAMV Cluj Napoca, merceccristian@hotmail.com

⁵ Lecturer Ph.D., USAMV Bucharest; PhD Student at Academy of Economic Studies Bucharest, raluca_nec@yahoo.com

phenomenon of autocorrelation appears. To illustrate the content of the method, we assume an experimental plan with three factors, each factor with 5 variants. That is: N (X_1) [0; 50; 100; 150; 200]; P (X_2) [0; 40; 80; 120; 160] and K (X_3) [0; 30; 60; 90; 120]. In this case, a complete experimental plan would include 125 variants in two or three rehearsals. Such an experimental plan is very difficult or even impossible to organize, to control and to complete. To illustrate the content of the proposed method, it was accepted that the experimental plan was a simplified one and contained only 60 combinations of the 125 possible (conventional data).

Table 1. Correspondence between the level of factors allocated and the average production per hectare

X_1	X_2	X_3	Y	X_1	X_2	X_3	Y	X_1	X_2	X_3	Y
0	0	0	4600	150	120	30	8490	100	40	90	7259
0	40	0	4945	150	160	30	8474	100	80	90	9039
50	40	0	5980	200	120	30	8614	100	120	90	9200
50	80	0	5865	200	160	30	8469	150	80	90	9313
100	40	0	6095	0	0	60	5217	150	120	90	9450
100	80	0	7590	0	40	60	5609	150	160	90	9432
100	120	0	7725	50	40	60	6783	200	120	90	9587
150	80	0	7820	50	80	60	6652	200	160	90	9426
150	120	0	7935	100	40	60	6913	0	0	120	5697
150	160	0	7920	100	80	60	8609	0	40	120	6125
200	120	0	8050	100	120	60	8762	50	40	120	7407
200	160	0	7915	150	80	60	8869	50	80	120	7264
0	0	30	4922	150	120	60	9000	100	40	120	7549
0	40	30	5291	150	160	60	8983	100	80	120	9401
50	40	30	6399	200	120	60	9130	100	120	120	9568
50	80	30	6276	200	160	60	8977	150	80	120	9685
100	40	30	6522	0	0	90	5478	150	120	120	9828
100	80	30	8121	0	40	90	5889	150	160	120	9809
100	120	30	8266	50	40	90	7122	200	120	120	9970
150	80	30	8367	50	80	90	6985	200	160	120	9803

In order to process such an incomplete database, compared to the total number of possible combinations, the literature has validated several methods, which offer the possibility of identifying collinearity and its distribution by factors. Each method is based on a certain hypothesis, the differences in the operability of the calculations may be substantial. The method proposed by the authors has as a working hypothesis the distribution of the total autocorrelation, respectively the sum of the squares of the deviations (SPA), on the factors of production, according to the principle of proportionality with the coefficients of simple correlation, using the distribution coefficient.

The method involves, in the case of three factors, the drawing up of a table (Table 2) comprising:

- Calculation of simple linear equations for each factor (Y_{X_1} ; Y_{X_2} ; Y_{X_3}) and multiple equation ($Y_{X_1X_2X_3}$), which was performed with Regression Function, Data => Analysis;

- Passing in the Table the SPA, by regression, for each factor ($SPAX_1 = 96151981$; $SPAX_2 = 85432839$; $SPAX_3 = 20253550$, $SPAX_1X_2X_3 = 120636154$ and $SPA_{total} = 136126417$);

- For the statistical highlighting of autocorrelation, the Durbin-Watson test for each factor was calculated: $DWX_1 = 0.689$; $DWX_2 = 0.543$; $DWX_3 = 0.365$. $DW_{theoretical}$ ($k = 1$; $n = 60$ is 1.184 and 2.03). It results the autocorrelation results for each factor ($0 < DW_{cal} < 1.184$) [10].

- The distribution of $SPAX_1X_2X_3 = 120636154$ on the three factors using the proportional distribution coefficient, in relation to the SPA sum, of the 3 factors;

Table 2. Linear equations, SPA and the calculation of the pure determination on each factor

	Linear equation	Simple SPA	Distribution coeff. of $SPAX_1X_2X_3$	Assigned SPA	R2 (determination)	R2(%)	r (correlation coeff.)
X1	$Y_{x1}=5743.05+19.17x_1$	96151981	0.48	57468781	0.4222	42.22	0.650
X2	$Y_{x2}=5609.84+24.58x_2$	85432839	0.42	51062091	0.3751	37.51	0.612

X3	$Y_{x3}=6918.98+13.69x_3$	20253550	0.10	12105282	0.0889	8.89	0.298
Sum SPA ($x_1+x_2+x_3$)		201838370	1.00				
$x_1.x_2.x_3$	$Y_{x_1x_2x_3}=4693.84+13.33X_1+9.65X_2+13.69X_3$	120636154		120636154	0.8862	88.62	0.941
Rest		15490263		15490263	0.1138	11.38	
Total;		136126417		136126417	1.0000	100.00	1.000

- The calculation of the determinations (R^2) of each factor, according to the SPA, resulted from the distribution of SPAX₁X₂X₃, by reference to the total SPA;

- The percentage calculation of the determination for each factor, which is also referred to as pure determination. The correlation factors corresponding to the pure determination were also calculated.

From this calculation, the YX₁X₂X₃ multiple regression equation explained 88.62% of factor and production relationship. Factors influenced as follows: X₁ (N) = 42.22%; X₂ (P) = 37.51%; X₃ (K) = 8.89%. Compared to 100% total influence, it remains an unexplained 11.38% rest.

In order to increase the speed of the calculation method of the autocorrelation distribution while preserving the principle of distribution, we suggest a method of calculation starting from the simple correlation coefficients. Practically, it starts from the individualization of the correlation coefficients of each factor, calculated using the features offered by Microsoft Excel as a spreadsheet work program. The calculation method, of course, assures the calculation of the pure determination of each factor by simply deciding the multiplication correlation coefficient and the coefficients of the simple correlation respectively.

RESULTS AND DISCUSSIONS

The distribution of autocorrelation on factors of influence implies the preliminary determination of the multiple correlation coefficient and of the simple correlation coefficients in the hypothesis of a certain theoretical regression model. Given the nature of the database presented in Table 1, a three-factorial linear model, a second-order tri-factorial model and a linear tri-factorial model were used to express the causal relationship between the three factors and the average production, with the combined influence of factors.

All calculations were performed using the Regression and Correlation functions of the Data =>

Analysis component of Microsoft Excel.

Through these calculations, we exemplify by calculating the pure determination for a trifactorial experience that is modeled by a linear equation, a second degree equation and a linear equation with the combined effect of factors.

The following concrete situations resulted:

A. Calculation of the pure determination in the linear tri-factorial model:

- It is calculated with the Regression function of the Data => Analysis component, the multiple equation: $Y_{(X_1.X_2.X_3)} = 4693.8+13.334X_1+9.648X_2+13.694X_3$ and the multiple correlation coefficient $r_{y.x_1x_2x_3}=0.94139$;

- It is calculated with Correlation Function in Data => Analysis, the simple correlation coefficients: $r_{Y.X_1}=0.84044$; $r_{Y.X_2}=0.79221$; $r_{Y.X_3}=0.38573$;

- For the statistical highlighting of autocorrelation, the Dubrin-Watson test for each factor was calculated: $DWX_1 = 0.689$; $DWX_2 = 0.543$; $DWX_3 = 0.365$. $DW_{theoretical}$ (k = 1; n = 60 is 1.184 and 2.03). It results positive autocorrelation for each factor ($0 < DW_{cal} < 1.184$) [10].

- By registering the coefficient of the multiple correlation and the coefficients of the simple correlation in the centralizing table (Table 3), the calculation steps are taken of the pure determination by factors, which leads to the individualization of the influence of each factor, and which are highlighted in the table as pure determination on factors.

Table 3. The case of a linear multifactorial model

Correlation and determination		Correlation coefficients	Determination coefficients	The simple determination proportion	The pure determination on factors
The sum of simple determination		*	148.27	100.00	*
Simple correlation	X ₁	0.84044	70.63	47.64	42.22
	X ₂	0.79221	62.76	42.33	37.51
	X ₃	0.38573	14.88	10.03	8.89
	X ₄	0.0000	0.00	0.00	0.00
	X ₅	0.0000	0.00	0.00	0.00
Multiple (X ₁ .X ₂ .X ₃)		0.9414	88.62	*	88.62

From the analysis of the data in Table 3, the linear regression equation Y (X₁.X₂.X₃) shows that the influence explained by the three factors is by 88.62%. This is explained by the influence of factor X₁ (N) by 42.22%, the influence of factor X₂ (P) by 37.51% and factor X₃ (K) by 8.89%. There is an unexplained influence of 11.38%.

B. The pure determination calculation of the second degree tri-factorial model:

- It is calculated with the Regression function of the Data => Analysis component, the multiple equation:

$$Y_{(X1*X1.X1*X2.X2.X2.X3.X3.X3)} = 4001.96 + 25.18X_1 - 0.0609X_1*X_1 + 21.001X_2 - 0.0658X_2*X_2 + 17.11X_3 - 0.0284 X_3*X_3$$

and the multiple correlation coefficient $r_{y.X_1.X_1*X_1.X_2.X_2*X_2.X_3.X_3*X_3} = 0.94242$;

- It is calculated with the Correlation function in the Data => Analysis component, the simple correlation coefficients (Table 4);

Tabelul 4. The simple correlation coefficients calculation

	Y	X ₁	X ₁ *X ₁	X ₂	X ₂ *X ₂	X ₃	X ₃ *X ₃
Y	1						
X ₁	0.840442	1					
X ₁ *X ₁	0.733032	0.954316	1				
X ₂	0.792212	0.823662	0.774358	1			
X ₂ *X ₂	0.698539	0.76979	0.765195	0.96272	1		
X ₃	0.385726	0	-1.3E-17	0	0	1	
X ₃ *X ₃	0.361809	0	1.36E-17	0	0	0.958927	1

- For the statistical highlighting of autocorrelation, the Dubrin-Watson test for each factor was calculated: DWX₁=0.689; DWX₁*X₁=0.1229; DWX₂=0.543; DWX₂*X₂=0.1253; DWX₃=0.365; DWX₃*X₃=0.1802. DWtheoretical (k = 1; n = 60 is 1.184 and 2.03). It Results positive autocorrelation for each factor (0 <DWcal <1.184) [10].

- By registering the coefficient of multiple correlation, respectively the coefficients of simple correlation in the centralizing table (Table 5), the calculation steps are taken of the pure determination on factors, linear and quadratic, and finally leading to the individualization of the influence of each factor, are passed in the table as pure determination by factors.

Table 5. The case of a second degree multifactorial model

Correlation and determination		Correlation coefficients	Determination coefficients(%)	Proportion of simple determinations	Determination of factors by degree of equation	Pure determination on factors
Simple determination Sum		*	263.89	100.00	*	*
Simple correlation	X ₁	0.84044	70.63	26.77	25.23	44.42
	X ₁ *X ₁	0.73303	53.73	20.36	19.19	
	X ₂	0.79221	62.76	23.78	22.41	

	X2*X2	0.69854	48.80	18.49	17.43	
	X3	0.38573	14.88	5.64	5.31	9.99
	X3*X3	0.36181	13.09	4.96	4.68	
	X4	0.0000	0.00	0.00	0.00	0.00
	X4*X4	0.0000	0.00	0.00	0.00	
	Multiple	0.9708	94.24	*	94.24	94.24

From the data analysis in Table 5, it follows that through the quadratic regression equation $Y_{(X_1.X_1*X_1.X_2.X_2*X_2.X_3.X_3*X_3)}$, the influence explained by the three factors is 94.24%. This is explained by the influence of factor X_1 (N) by 44.42%, the influence of factor X_2 (P) by 39.84% and the influence of factor X_3 (K) by 9.99%. There is an unexplained influence of 5.76%.

C. The pure determination calculation in the linear tri-factorial model with the combined influence of the factors:

- It is calculated with the Regression function of the Data => Analysis component, the multiple equation:

$$Y_{(X_1.X_2.X_3.X_1*X_2.X_1*X_3.X_2*X_3)} = 4286.84 + 20.69X_1 + 20.864X_2 + 9.758X_3 + 0.114X_1*X_2 + 0.024X_1*X_3 + 0.017X_2*X_3 \text{ and } r_{Y(X_1.X_2.X_3.X_1*X_2.X_1*X_3.X_2*X_3)} = 0.9681;$$

- It is calculated with the Correlation function of Data => Analysis component, the simple correlation coefficients (Table 6);

Table 6. Calculation of simple correlation coefficients

	Y	X1	X2	X3	x1x2	x1x3	x2x3
Y	1						
X1	0.84044	1					
X2	0.79221	0.823662	1				
X3	0.38573	0	0	1			
X1*X2	0.75991	0.91565	0.91795	0.00000	1		
X1*X3	0.79396	0.60357	0.49714	0.67347	0.553	1	
X2*X3	0.75114	0.46824	0.56848	0.71780	0.522	0.907336	1

- For the statistical highlighting of autocorrelation, the Durbin-Watson test for each factor was calculated: $DWX_1=0.689$; $DWX_2=0.543$; $DWX_3=0.365$; $DWX_1*X_2=$; 0.1169; $DWX_1*X_3=0.1314$; $DWX_2*X_3=0.1312$. $DW_{theoretical}$ ($k = 1$; $n = 60$ is 1.184 and 2.03). Resulting positive autocorrelation for each factor ($0 < DW_{cal} < 1.184$) [10].

- By registering the coefficient of the multiple correlation and the coefficients of the simple correlation in the centralizing table (Table 7), the calculation steps were taken of pure determination on factors, linear and the combined influence of the factors are completed, which ultimately leads to the individualization of the influence of each factor and influences combined, as well as the calculation of pure determination on factors.

Table 7. The case of a multifactorial linear model with the combined influence of the factors

Correlation and determination		Correlation coefficients	Determination coefficients (procentual)	Simple determination share	Determination of factors after the combination of factors	Pure determination on factors
Determination Sum		*	3.25477	100	*	*
Simple correlation	X1	0.84044	0.70634	21.7	0.20	44.14
	X2	0.79221	0.62760	19.3	0.18	39.03
	X3	0.38573	0.14878	4.6	0.04	10.56
	X1*X2	0.75991	0.57746	17.7	0.17	
	X1*X3	0.79396	0.63037	19.4	0.18	
	X2*X3	0.75114	0.56421	17.3	0.16	
Multiple		0.96812	0.93725	x	0.937	93.73

From the analysis of the data in Table 7, it follows that by the quadratic regression equation $Y_{(X_1.X_1X_2.X_2X_3.X_3X_3)}$, the influence explained by the three factors is 93.73%. This is explained by the influence of each factor ($X_1(N)=0.20\%$; $X_2(P)=0.18\%$; $X_3(K)=0.04$), and the combined influence of factors $X_1 * X_2$ (NP) factor $X_1 * X_3$ (NK) by 0.18% and $X_2 * X_3$ (PK) = 0.16%. This influence was also allocated to factors by the linear proportional method resulting that the influence of X_1 (N) = 44.14%; X_2 (P) = 39.03% and X_3 (K) = 10.56%. There is an unexpected influence of 6.27%.

CONCLUSIONS AND PROPOSALS

1. The working method aims to calculate the pure influence of each factor investigated according to the principle of the proportionality of the autocorrelation in relation to the coefficients of the simple determination of the factors;

2. The first step in making the calculations is to calculate the multiple correlation coefficient and the simple correlation coefficients using the Regression and Correlation function of the Microsoft Excel Data Analysis component

3. By registering the multiple correlation coefficient, respectively the coefficients of simple correlation, the program automatically carries out successive steps of repartition of the autocorrelation according to the principle of proportionality;

4. To help the user understand how the calculations work, we show the steps that the program goes through instantly:

- Calculation of the simple determination coefficients and the multiple determination coefficient by reference to 100%;
- Sum of coefficients of simple determination;
- Percentage calculation of simple determinations, considering their sum equal to 100;
- Calculation of the influence of the investigated factors on the effect, by multiplying the simple determination of factors with multiple determinations.

5. We are aware that the extraordinary advances in computer science will allow, in the near future, calculating the pure determination of factors, by various methods developed by mathematicians, through a single function of a program- product.

BIBLIOGRAPHY

1. Merce E. – Statistică – aplicații practice, Universitatea Babeș-Bolyai, Facultatea de științe economice, Cluj-Napoca, 1986, p.92
2. Merce E. – Cu privire la calculul coeficientului corelației parțiale (I). Definiere, conținut. Studia 1/1989, Universitatea “Babeș-Bolyai”, Cluj-Napoca, pag. 51-60.
3. Merce E., V. Pârv – O nouă metodă de determinare a coeficientului corelației parțiale. Revista de Statistică nr. 3/1991, pag. 32-39.
4. Merce E., B. Pârv, Flavia Laun - Calculul coeficienților determinației parțiale. Metodă și program. Revista română de statistică nr. 6, 1992, pag. 29-37.
5. Merce E., C. C. Merce (2003) – Eliminarea autocorelației dintre variabilele cauzale prin întregirea bazei de date, în volumul Specializare, dezvoltare și integrare; ISBN; 973-86547-4-2; pag.275-281, Universitatea Babeș-Bolyai, Facultatea de Științe Economice, Cluj-Napoca;
6. Merce E., C. C. Merce, Cristina B. Pocol (2017): *A u t o c o r e l a ț i a - prevalența identificării cauzei coliniarității - Vol. Economia agrară și dezvoltarea rurală - realități și perspective pentru România, 2017*, Editura ASE; ISSN-L 2247 – 7187; 7 pagini;
7. Merce E., C. C. Merce – Statistică – paradigme consacrate și paradigme întregitoare, Editura AcademicPres, Cluj-Napoca, 2009, p.311;
8. Moineagu C. - Modelarea corelațiilor în economie, Editura științifică, București, 1974.
9. Drăghici M., Berca M., 1971, Studiu comparativ privind posibilitățile de interpretare prin diferențele polifactoriale in: "Analele IANB". Seria A, XIV, 1971
10. Tănăsioiu O. E., Andreea Iluia Iacob, 2005, Econometrie. Studii de caz, Editura ASE, București.