University of Szeged
Faculty of Economics and Business Administration
Doctoral School in Economics

Examination of Multicollinearity in Linear Regression Models
Examination of PETRES' Red

Theses of PhD Dissertation

Written by:
Péter Kovács

Consultant:
Dr. Tibor Petres
associate professor

Szeged
2008

I. Problem definition, aims and hypotheses of the research

I. 1. Definition of the problem

In today's globalizing world decision makers have an increased need for information. The great increase in the quantity of data is not automatically accompanied by an appropriate increase in information. Actually, the problem that decision makers have to face today is not the lack but the abundance of information, but this huge amount of data frequently has only a little information content, which means that **redundancy** is high. Redundancy means "superfluous" data which do not convey new or noteworthy information in terms of the examination. For this reason **the information content of metric data is an essential issue** in empirical analyses. This is particularly true for the application of linear regression models. **In the case of linear regression models, multicollinearity can be interpreted as a type of redundancy.** With matrix algebraic notation this can be written in the form of $\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}$, where $\tilde{\mathbf{y}}$ is the n component column vector of the dependent variable; $\tilde{\mathbf{X}}$ is the matrix of explanatory variables consisting of row n and column (m+1), where the first column is always an $\tilde{\mathbf{x}}_0$ sum vector; $\tilde{\boldsymbol{\beta}}$ is the (m+1) component column vector of the model parameters unknown to us; m is the number of explanatory variables (explanatory variables); $\tilde{\boldsymbol{\varepsilon}}$ is the n component column vector of the error term.

The concept of multicollinearity is apparently uniform in literature. Definitions usually differ from each other in one word, but this entails significant changes in content.

Multicollinearity as an expression was first used by RAGNAR FRISCH. He used it for the description of cases in which one variable was present in several relations. In his examinations he did not distinguish dependent variables from explanatory variables. He supposed that the measurement of all variables was erroneous, the correlation between the actual values of the variables had to be estimated on this basis.

It is very superficial when multicollinearity is defined as the absence of the independence of explanatory variables. This definition is problematic because it is not defined unambiguously what the independence of the explanatory variables means. Does it mean their linear independence or possibly their independence in the statistical sense?

One of the primary conditions of the standard linear regression model is the linear independence of the explanatory variables (KENNEDY). Therefore in certain sources multicollinearity is interpreted as the absence of the linear independence of explanatory variables. This approach can be regarded as a special case of multicollinearity, which is called extreme multicollinearity. This case does not pose special problems in practice as it is easily manageable.

In the course of empirical analyses cases close to extreme multicollinearity are frequently encountered, when the variances of individual estimated parameters are considerably increased as compared to the variance of the error term. The great majority of literature on multicollinearity deals with this case. However, let me remark that multicollinearity could mean a much more general phenomenon, namely the covariance **of explanatory variables**. Naturally, the special cases of this definition would convey the content meant by multicollinearity to everybody.

The recognition of multicollinearity and the identification of its cause often present a serious problem in empirical examinations, as on the one hand **the negative consequences of multicollinearity do not always occur**, and on the other hand **multicollinearity can be caused not only by one variable but also by a group of variables.** Thus it can be suspected that **the indicators of multicollinearity do not always describe this phenomenon properly.**

The interpretation of the indicators of multicollinearity is frequently quite subjective. Firstly, most of the indicators give how much the data examined are not ideal, that is to what extent they deviate from the "ideal case" when each explanatory variable is linearly independent of each other. For some indicators there is no definite boundary for indicating the harmful extent of "deviation". Secondly, if the specification of the applied model is appropriate, multicollinearity is only the consequence of the lack of proper information.

The success of the methods used for reducing or eliminating the negative effects of multicollinearity can largely depend on the exact recognition of multicollinearity. Although the use of the majority of these methods decreases or may decrease the extent of the negative consequences of multicollinearity, this may be accompanied by other negative consequences – for example, by significant information loss or by the improper interpretability of the results.

The topicality of the subject is given by the fact that these problems are almost always encountered in the course of economic analyses. This is particularly true if there is a strong trend in the explanatory variables, or if the information available is too little for the examination of the effect of the explanatory variables on the dependent variable.

In sum, in empirical analyses it frequently happens that not all the data have a useful content in respect of the examination, in other words the database is redundant. In a multivariate linear regression calculation multicollinearity can be interpreted as a type of redundancy. Therefore during regression analysis it is essential to know the proportion of the data with a useful content in respect of the estimator $\hat{\tilde{\boldsymbol{\beta}}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$, but its proper measurement poses a problem. It is questionable what the indicators of multicollinearity indicate, or how the negative consequences of the presence of multicollinearity can be decreased.

I. 2. Aim of the dissertation

PETRES' *Red* is one possibility for measuring the proportion of data with a useful content in respect of the estimator $\hat{\tilde{\boldsymbol{\beta}}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$. PETRES' *Red* is a new possible indicator of redundancy and thus of multicollinearity. The *Red* indicator is defined by using the eigenvalues λ_j $(j=1,2,\ldots,m)$ of the correlation matrix $\mathbf{R}$ of the explanatory variables. The *Red* indicator is based on the following train of thought. If the database serving as the source of the explanatory variables is redundant in respect of estimator $\tilde{\boldsymbol{\beta}}$, that is if the covariance of the data is considerable, not all the data will have a useful content. The smaller the proportion of the data with a useful content is, the greater the extent of redundancy will be. The greater the dispersion of the eigenvalues is, the greater the covariance of the explanatory variables in the database will be. There are two extreme cases: all the eigenvalues are equal to each other (that is their value is one), or all the eigenvalues with the exception of one equal zero. The extent of dispersion can be quantified with the relative dispersion of the eigenvalues or with their dispersion (being equal in this case).

$$v_\lambda = \frac{\sigma_\lambda}{\bar{\lambda}} = \frac{\sqrt{\frac{\sum_{j=1}^{m}(\lambda_j - \bar{\lambda})^2}{m}}}{\frac{\sum_{j=1}^{m}\lambda_j}{m}} = \frac{\sqrt{\frac{\sum_{j=1}^{m}(\lambda_j - \bar{\lambda})^2}{m}}}{\frac{m}{m}} = \sqrt{\frac{\sum_{j=1}^{m}(\lambda_j - 1)^2}{m}} = \sigma_\lambda$$

In order to make the redundancy of various databases comparable, the above indicator has to be normalized. As the eigenvalues are nonnegative, normalization is carried out with value $\sqrt{m-1}$ because of the relationship $0 \le v_\lambda \le \sqrt{m-1}$ concerning relative dispersion.

The indicator obtained in this way can be used to quantify the extent of redundancy, and the *Red* indicator can be defined with its help as follows.

$$Red = \frac{v_\lambda}{\sqrt{m-1}}$$

In the case of the absence of redundancy the value of the above indicator is zero or zero percent, while in the case of maximum redundancy it is one or one hundred percent.

The *Red* indicator measures the redundancy of the examined database of the given size. When the redundancies of two or more databases of different sizes are compared, the *Red* indicators can only be used to determine how redundant individual databases are, but one cannot make a direct statement as to which of these has more useful data.

I. 3. Structure of the dissertation

The aim of my dissertation **is to examine the properties of the *Red* indicator and to compare it with other indicators in a multivariate linear regression model.** In accordance with its aim, **my dissertation is organized as follows.**

In Chapter I the problem, the tasks and the aims of the dissertation are laid out. In order to do so the fundamentals of regression analysis essential to understanding the dissertation are summarized in short.

In Chapter II the literature on multicollinearity is surveyed. Several known and less-known indicators of multicollinearity, its method of detection, its potential consequences and also the possibilities to decrease the negative effects thereof are discussed in this chapter.

Procedures of detection and indicators discussed

- Examination of the correlation matrix of explanatory variables
- KLEIN'S rule of thumb
- MASON'S and PERREAULT'S proposal
- M_1 indicator
- *M in*dicator
- FARRAR–GLAUBER test
- WILKS test
- Examination of the differences of the correlation coefficients and partial correlation coefficients
- FRISCH'S bunch maps method
- *VIF* indicator
- BELSLEY'S gamma indicator
- FELLMAN'S *L* indicator
- MAHAYAN'S and LAWLES'S M_1 indicator
- THISTED'S *mci* (multicollinearity index) és *pmci* (predicted multicollinearity index)
- *ISRM* (Index of Stability of Relative Magnitudes)
- *DEF* indicator (Direct Effect Factor)

Procedures discussed for decreasing the adverse effects of multicollinearity

- Omission of explanatory variables from the model
- Increasing the number of elements in the sample
- Use of external information
- Use of the MOORE–PENROSE inverse
- Principal component analysis

– Ridge regression

– *Nested estimate* procedure

– Examination of the orthogonality of explanatory variables

As the conclusion of the chapter, I illustrated the procedures and indicators mentioned with an example.

After surveying the literature and on the basis of empirical examples, I made the following findings.

1. **The increase in the variance of the estimated parameters is the most frequently mentioned negative consequence of multicollinearity, but what should be considered is not their absolute value but the extent of their "inflation" compared to the variance of the error term.**

2. **Several methods are known for detecting and measuring multicollinearity but only few of them are widely accepted, partly because it is often very difficult to detect multicollinearity, and partly because the interpretation of the majority of the indicators is quite subjective. Usually, some of the indicators and procedures only detect multicollinearity without – generally due to their synthetic nature – localizing the problem. In contrast with this, a group of the indicators and procedures tries to localize multicollinearity – with more or less success.**

3. **The great disadvantage of indicators using the reciprocals of eigenvalues is that their interpretation is subjective, which means that there is no definitive threshold value indicating strong multicollinearity. The values of the indicators are not comparable with each other. Moreover, the values of these indicators mainly depend only on the smallest eigenvalue.**

4. **The presented indicators describe multicollinearity from different aspects.**

5. **There is no generally valid procedure for decreasing the negative consequences of multicollinearity, that is each procedure may have adverse effects – from different aspects.**

6. **As a summary of the described and applied indicators, ideas and algorithms it can be stated that the indicators and procedures mentioned are not generally**

valid in the sense that only in special cases do they describe or handle the phenomenon of multicollinearity properly.

In Chapter III the methods applied during my research and their results are presented. The major characteristics of the *Red* indicator are examined. The results of other, similar examination methods are also presented here and compared with my results.

The other chapters of my dissertation contain the evaluation of my research activity and results, the list of the literature, figures and tables used, the outcome of longer computerized analyses and the list of my publications.

I. 4. Research hypotheses

The problems and hypotheses detailed in the following are investigated in order to achieve the aim of my dissertation.

1. **Calculation of the *Red* indicator in a different way.**

In accordance with its definition, the *Red* indicator can be calculated on the basis of the eigenvalues of the correlation matrix. The question may arise whether the value of the indicator can be calculated without knowing the eigenvalues, merely from the elements of the correlation matrix of the explanatory variables. I examined the following hypothesis in Chapter III.1.

Hypothesis 1: The Red indicator can be expressed without knowing the eigenvalues of the correlation matrix of the explanatory variables, merely from the correlation coefficients of the pairs.

2. **Generalization of the examination method of multicollinearity.**

In my opinion not only the covariance of the variable pairs but also the covariance of the variable groups may pose a problem during the examination of multicollinearity. However, no detailed methodology has been worked out for this yet. I think a possible solution to the problem could be the use of canonical correlation analysis, a special case of which can be examined with the help of the *Red* indicator. I examined the following hypothesis in Chapter III.1.

Hypothesis 2: The covariance of two groups of explanatory variables can be examined in special cases with the help of the Red indicator.

3. Examination of new modelling possibilities of multicollinearity.

A possible method for modelling multicollinearity is to examine the orthogonality of explanatory variables, that is the "stretching" of the space of explanatory variables. The question rightly arises whether multicollinearity can be modelled in a different way. I examined the following hypothesis in Chapter III.2.

Hypothesis 3: The elliptical model of multicollinearity can be formulated on the basis of the Red indicator as a new approach.

4. Searching for some relationship between the variances of the estimated regression parameters and the *Red* indicator.

As one of the most frequently mentioned negative consequences of multicollinearity is the increase in the variance of the estimated regression parameters and in their inflation, it is advisable to examine the relationship between the *Red* indicator and the variances of the estimated regression parameters. I examined the following hypothesis in Chapter III.3.

Hypothesis 4: A critical value of the Red indicator can be given as the precondition for the variances of the estimated parameters not to be infinite.

5. Examination of the distribution of the *Red* indicator.

In Chapter III.4 I tried to prepare the empirical distribution function of the *Red* indicator and to determine its theoretical distribution.

6. Examination of the application possibilities of the *Red* indicator.

An interesting question is in what fields the *Red* indicator can be used. I examined the following hypothesis in Chapter III.5.

Hypothesis 5: The KMO index used in factor analysis can be expressed on the basis of the Red indicator.

7. Identification of an indicator similar to the *Red* indicator.

As the *Red* indicator is a normal relative dispersion calculated on the basis of the eigenvalues of the correlation matrix of the explanatory variables, I think that multicollinearity can be measured with some other dispersion indicator of the eigenvalues, which is a conception similar to that of the *Red* indicator. I examined the following hypothesis in Chapter III.6.

Hypothesis 6: The GINI coefficient of the eigenvalues of the correlation matrix of the explanatory variablesis an indicator of multicollinearity similar in conception to the definition of the Red indicator.

II. Research results and findings

Chapter III of the dissertation presents the new results of the paper. Part of the examinations is based on theoretical considerations, while for the other part various samples had to be created and their results had to be analysed. SPSS 13.0 and Microsoft Excel programs were used for the analyses. The geometric depiction was made with Derive 6.0. In sum, my dissertation contains the following theses.

Thesis 1: The *Red* indicator can be expressed without knowing the eigenvalues of the correlation matrix of the explanatory variable s, merely as the quadratic mean of the correlation coefficients of the pairs.

I could express the *Red* indicator – without knowing the eigenvalues – as the quadratic mean of the elements outside the main diagonal of the correlation matrix of the explanatory variable s. This means that this indicator shows not only the proportion of the data with a useful content in respect of the estimator $\tilde{\boldsymbol{\beta}}$ but also the mean covariance of the explanatory variable s. This result has been acknowledged in several international conferences and has also been cited in illustrious international journals.

Thesis 2: The examination of the covariance of two groups of explanatory variablesis possible with the *Red* indicator in the case of groups with one element

each, and with the harmonic mean of the VIF_j values in the case of groups with one–(*m*–1) elements.

I found that multicollinearity can be caused not only by variables but also by groups of variables. As this does not have abundant literature, later on I am going to examine the effect of the covariance of the variable groups. **I established that one special case of this can be measured with the *Red* indicator, while another special case with the help of the harmonic mean of the VIF_j values.**

Thesis 3: The elliptical model of multicollinearity can be formulated on the basis of the *Red* indicator as a new approach.

As a new approach, I formulated the elliptical model of multicollinearity. Parallel with the increase in the extent of the mean covariance of the variables, the "possible eigenvalues" are situated on an m-dimensional sphere with a greater radius**. The "possible eigenvalues" are situated on a segment of the m-dimensional sphere in such a way that with a fixed *Red* value they are located on an (*m*–1)-dimensional ellipsoid.**

Unfortunately, the higher the dimension number of the model is, the more conditions have to be given for determining and studying the range of "possible eigenvalues". Therefore the detailed examination of this range and of the elliptical curves was carried out only for three explanatory variables. I determined the possible values of the *Red* indicator as the function of one eigenvalue, and I could give the possible values of each eigenvalue depending on the value of the *Red* indicator. I compared how the ellipses and the lines containing the identical-value quotients of the highest and lowest values of the eigenvalues "move along" the range of the "possible eigenvalues".

Later I will try to improve the model and to extend the examination to higher dimensions.

Thesis 4: A critical value of the *Red* indicator can be given as the necessary precondition for the variances of the estimated parameters not to be infinite.

As the *Red* indicator is a synthetic indicator, it cannot be connected separately to the variances of the estimated regression parameters. I found that not the absolute value of the variances of the estimated regression parameters is to be examined but their "inflation" compared to the variance of the error term. The sum and mean of these depend in turn on the reciprocal sum of the eigenvalues. I proved that the product of the harmonic mean of the eigenvalues and of the arithmetical mean of the estimated regression parameters equals the variance of the error term, and also that the product of the harmonic mean of the eigenvalues and of the arithmetical mean of the VIF_j values equals one.

After the refutation of a previous statement of mine I gave a critical value of the *Red* indicator as the necessary precondition for the variances of the estimated parameters not to be infinite, and similarly critical values which are the preconditions for the number of zero eigenvalues to be lower than k. As this is of slight practical importance as such, further detailed examinations are necessary. However, I performed these examinations for three explanatory variablesby using the elliptical model. I found that the reciprocal sum of the eigenvalues increases when moving farther from the lower boundary of the range of "possible eigenvalues". Based on this, I determined the smallest and – if possible – the greatest extent of the "inflation" of the sum of the variances of the estimated parameters as compared to the variance of the error term, in the function of the *Red* indicator. On the basis of this, in the function of the *Red* indicator, a critical value can be given which is the precondition for the sum of the variances of the estimated parameters not to be "inflated" to an extent greater than set beforehand, compared to the variance of the error term.

In the course of the examination of the distribution of the *Red* indicator I prepared the empirical distribution function in a few dimensions. During the analysis I examined only existing correlation structures. I used an algorithm written by myself for generating the "possible eigenvalues" and for preparing the distribution of the *Red* indicator. The essence of this algorithm is that all the possible eigenvalue combinations are prepared with a given accuracy. The analysis was made more difficult by the number of generated eigenvalues, which may be hundreds of thousands or even hundreds of millions even in a rough approximation. **The identification of the distribution of the *Red* indicator was unsuccessful.** High-performance computers would be needed for further examinations.

Thesis 5: The *KMO* index used in factor analysis can be expressed on the basis of the *Red* indicator.

I proposed an application possibility for the *Red* indicator. The *KMO* index used in factor analysis can be expressed on the basis of the *Red* indicator. Based on this I established that the mean covariance of the partial correlation coefficients cannot be smaller than the mean covariance of the correlation coefficients.

Thesis 6: The GINI coefficient of the eigenvalues of the correlation matrix of the explanatory variablesis an indicator of multicollinearity similar in conception to the definition of the *Red* indicator is.

I determined another possible indicator of multicollinearity based on a train of thought similar to that of the *Red* indicator. This indicator is the GINI coefficient of the eigenvalues. I identified an easily manageable manner of calculation for the indicator. I examined the behaviour of the indicator – in the case of three explanatory variables– in the range of the "possible eigenvalues". The behaviour of the indicator necessitates further detailed studies.

II. 1. Future research directions

As a conclusion to my dissertation, I am going to summarize my planned future research directions in the order corresponding to the structure of the dissertation.

1. The possibility to decrease the negative consequences of multicollinearity is a very important practical problem. Therefore
    a. on the one hand I would like to examine whether some optimal estimation can be made for the distort parameter used in ridge regression on the basis of the value of the *Red* indicator.
    b. On the other hand, I would like to prepare a variable selection procedure based on the value of the *Red* indicator by defining the indicator also partially, as a explanatory variable, as the mean covariance of the given explanatory variable with all the other explanatory variables.

2. I would like to continue to examine the extension of multicollinearity, that is how the covariance of a group consisting of two or an optional number of explanatory variables can be measured, and what negative consequences the phenomenon has.
3. Later on I would like to reveal further properties of the elliptical model both in the case of three explanatory variables and in higher dimensions.
4. I am planning to examine the relationship of the *Red* indicator and the *Red* indicator to be defined partially with the "inflation" of the estimated regression parameters in more detail.
5. The definition of the theoretical distribution and the empirical distribution of the *Red* indicator may pose an immense task in future.
6. I would like to prepare some statistical test concerning the hypothetical value of the *Red* indicator.
7. I would like to extend the range of application of the *Red* indicator both during theoretical methods and economic studies.

III. Publications and conference lectures

<u>*Reviewed scientific publications*</u>

[1] SZONDI I. – KOVÁCS P. – IDOVIKA B. [2002]: A családok helyzete Szeged város lakótelepein *(The Situation of Families in the Housing Estates of the City of Szeged)*, *ACTA JURIDICA ET POLITICA*, Tomus LXII. Fasc. 18., Szeged, 30 oldal.

[2] FÜLÖP V. – SZONDI I. – KOVÁCS P. [2003]: Lakáscélú állami támogatások és egyéb, a lakáshoz jutást segítő ellátási formák *(State Housing Supports and Other Forms of Housing Provisions)*, *PUBLICATIONES DOKTORANDUM JURIDICORUM*, Tomus II. Fasc. 6. Szeged, 30 oldal.

[3] KOVÁCS P. – SZONDI I. [2003]: Úton az információs társadalom felé *(On the Way to the Informational Society)*, *ACTA JURIDICA ET POLITICA*, Tomus LXIII. Fasc. 13., Szeged, 20 oldal.

[4] KOVÁCS P. – PETRES T. – TÓTH L. [2004]: Adatállományok redundanciájának mérése *(Measurement of the Redundancy of Databases)*, *Statisztikai Szemle*, Budapest, 82. évfolyam 6.-7. szám, 595-604. oldal.

[5] GYÉMÁNT R. – PETRES T.– KOVÁCS P. [2005]: A Szandzsák, az egyedülálló vallási régió *(Sanjak, the Unique Region of Religion)*, *Területi statisztika*, 8. (45.) évfolyam 3. szám 278-287. oldal.

[6] KOVÁCS P. – PETRES T. – TÓTH L. [2005]: A new measure of multicollinearity in linear regression models, *International Statistical Review (ISR)*, Volume 73 Number 3, Voorburg, The Netherlands, 405-412. oldal.

[7] KOVÁCS P. – PETRES T. – TÓTH L. [2006]: Válogatott fejezetek Statisztikából, Többváltozós statisztikai módszerek *(Selected Chapters in Statistics, Multivariate Statistical Methods)*, JATEPress, Szeged, 167 oldal.

[8] KOVÁCS P. [2008]: A multikollinearitás vizsgálata lineáris regressziós modellekben *(Examination of Multicollinearity in Linear Regression Models)*, *Statisztikai Szemle*, Budapest, 86. évfolyam 1 szám, 38-67. oldal.

[9] LUKOVICS M.– KOVÁCS P. [2008]: Eljárás a területi versenyképesség mérésére *(Procedure for the Measurement of Regional Competitiveness)*, *Területi statisztika*, KSH, 11. (48.) évfolyam, 20 oldal, (megjelenés alatt).

[10] VILMÁNYI M. – KOVÁCS P. [2008]: Egyetemi-ipar együttműködések teljesítménye és lehetséges vizsgálati módszere *(Performance and Possible Examination Method of University-Industrial Co-operations)*, *Kérdőjelek a régiók gazdasági fejlődésében* (szerk. LENGYEL I. – LUKOVICS M.), *JATEPress*, Szeged, 25 oldal, (megjelenés alatt).

[11] KOVÁCS P. [2008]: Az információs társadalom szerinti területi egyenlőtlenségek mérése *(Measurement of Regional Inequalities According to the Informational Society)*, *Kérdőjelek a régiók gazdasági fejlődésében* (szerk. LENGYEL I. – LUKOVICS M.), *JATEPress*, Szeged, 11 oldal, (megjelenés alatt).

Educational materials, lecture notes

[1] KOVÁCS P. – PETRES T. [2004]: Statisztika Feladatgyűjtemény (közgazdász hallgatók számára) *[Collection of Exercises in Statistics (for students of economics)]*, SZTE GTK, 120 oldal.

[2] KOVÁCS P. – PETRES T. [2004]: Statisztika Feladatgyűjtemény *(Collection of Exercises in Statistics)*, Dunaújvárosi Főiskola, Dunaújváros, 284 oldal.

[3] KOVÁCS P. – PETRES T. [2004]: Statisztika Képletgyűjtemény *(Collection of Formulae in Statistics)*, Dunaújvárosi Főiskola, Dunaújváros, 50 oldal.

[4] KATONA T. – KOVÁCS P. – PETRES T. [2006]: Általános statisztika, tankönyv *(General Statistics, Coursebook)*, JATEPress, Szeged, 225 oldal.

[5] KOVÁCS P. – PETRES T. [2006]: Általános Statisztika Feladatgyűjtemény (joghallgatók részére) *[Collection of Exercises in General Statistics (for students of law)]*, JATEPress, Szeged, 2005, 132 oldal.

[6] KOVÁCS P. [2006]: Általános statisztikai alapismeretek, EU távoktatás elektronikus jegyzet *(Fundamentals in General Statistics, electronic lecture notes for EU distant teaching)*, 80 oldal.

[7] KOVÁCS P. — PETRES T. [2007]: Tanulási útmutató a főiskolák és egyetemek Általános statisztika című tantárgyához *(Learning Guide for the Subject of General Statistics in Universities and Colleges)*, Dunaújvárosi Főiskola, Dunaújváros, 190 oldal.

[8] KOVÁCS P. — PETRES T. [2008]: Szoftverek alkalmazása az üzleti életben: statisztikai programok *(Application of Softwares in Business Life: Statistical Programs)*, Dunaújvárosi Főiskola, Dunaújváros, 83 oldal.

[9] KOVÁCS P. — PETRES T. [2008]: Tanulási útmutató a Szoftverek alkalmazása az üzleti életben: statisztikai programok tantárgyához *(Learning Guide for the Subject of Application of Softwares in Business Life: Statistical Programs)*, Dunaújvárosi Főiskola, Dunaújváros, 78 oldal.

<u>*Foreign language conference publications*</u>

[1] KOVÁCS P. – SZONDI I. [2006]: E-europe- E-Hungary, Ungarn auf der Schwelle in die EU, A Pólay Elemér Alapítvány Könyvtára, sorozatszerkesztő: Balogh Elemér, Szeged, 29.-48. oldal.

[2] KOVÁCS P. – PETRES T. [2006]: A New Measure of Multicollinearity in Linear Regression Models, International Conference Applied Statistics (2006, Ribno, Slovenia), Program and Abstract, Statistical Society of Slovenia, Ljubljana.

[3] KOVACS P. – LUKOVICS M. [2006]: Classifying Hungarian sub-regions by their competitiveness, Globalization Impact on Regional and Urban Statistics, 25th

SCORUS Conference on Regional and Urban Statistics Research, Wroclaw, Poland, http://www.scorus2006.ae.wroc.pl, 12 oldal.

[4] KOVÁCS P. – PETRES T. [2007]: Measure of Multicollinearity with a New, Original Indicator (PETRES' *Red*) in Linear Regression Models, International Conference on Mathematics & Statistics, Athens Institute for Education Research, Athens, (KIADÁS ALATT)

Hungarian language conference publications

[1] KOVÁCS P. – LAMPERTNÉ A. I. – PETRES T. [2005]: A multikollinearitás mérése lineáris regressziós modellekben *(Measurement of Multicollinearity in Linear Regression Models)*, A Dunaújvárosi Főiskola Közleményei XXVI/II., Dunaújváros, 355-365. oldal.

[2] KOVÁCS P. [2005]: Statisztikai mintákat generáló algoritmusok *(Algorithms Generating Statistical Samples)*, A Dunaújvárosi Főiskola Közleményei XXVI/II., Dunaújváros, 347-354. oldal.

[3] KOVÁCS P. [2005]: Az informatika alkalmazása a közgazdasági képzésben *(Application of Informatics in Economist Education)*, Informatika a felsőoktatásban konferencia 2005 CD-melléklete, Debreceni Egyetem Informatikai Kar, Debrecen, 6 oldal.

[4] KOVÁCS P. [2005]: Az informatika oktatása és lehetőségei a jogászképzésben *(Teaching and Possibilities of Informatics in Law Education)*, Informatika a felsőoktatásban konferencia 2005 CD-melléklete, Debreceni Egyetem Informatikai Kar, Debrecen, 6 oldal.

[5] KOVÁCS P. – PETRES T. [2006]: A PETRES-féle *Red*-mutató eloszlásának vizsgálata *(Examination of the Distribution of PETRES' Red)*, A Dunaújvárosi Főiskola Közleményei XXVII/II., Dunaújváros, 2006, 521-530. oldal.

[6] KOVÁCS P. – PETRES T. – LUKOVICS M. [2006]: A PETRES-féle *Red*-mutató alkalmazásának lehetőségei *(Application Possibilities of PETRES' Red)*, *A Dunaújvárosi Főiskola Közleményei XXVIII.*, Dunaújváros, 304-316. oldal.

Other studies

[1] KOVÁCS P. [2006]: A statisztika oktatásának és oktatásmódszertanának reformálása a saját gyakorlatomban *(Reform of the Teaching and Teaching Methodology of Statistics in My Own Practice)*, „A felsőoktatás szerkezeti és tartalmi fejlesztése” tárgyú Humánerőforrás-fejlesztési Operatív Program (HEFOP 3.3.) „Partnerközpontú önértékelési modell megalkotása és továbbképzések a felsőoktatási intézmények humánerőforrásainak fejlesztéséért” tanulmányainak CD gyűjteménye, Dunaújváros, 52 oldal.

Foreign language conference lectures

[1] KOVÁCS P. – SZONDI I.: eEurope, eHungary, Társadalmi és gazdasági kihívások az Eu-csatlakozás küszöbén *(Social and Economic Challenges upon the Imminent Accession to the European Union)*, angol nyelvű, nemzetközi konferencia előadás, Szeged, 2004. június 12.

[2] KOVÁCS P. – LUKOVICS M.: Classifying Hungarian sub-regions by their competitiveness, Globalization Impact on Regional and Urban Statistics, 25th SCORUS Conference on Regional and Urban Statistics Research, Wroclaw, Poland, 2006. augusztus 30.-szeptember 1.

[3] KOVÁCS P. – PETRES T.: A new measure of multicollinearity in linear regression models, Applied Statistics 2006 International Conference, Ribno (Bled), Slovenia, 2006. szeptember 17.-20.

[4] KOVÁCS P. – PETRES T.: Measure of Multicollinearity with a New, Original Indicator (PETRES’ *Red*) in Linear Regression Models, International Conference on Mathematics & Statistics, ATINER, 2007. június 11.

Hungarian language conference lectures

[1] KOVÁCS P. – LAMPERTNÉ A. I. – PETRES T.: A multikollinearitás mérése lineáris regressziós modellekben *(Measurement of Multicollinearity in Linear Regression Models)*, DUF Közgazdasági szimpózium, Dunaújváros, 2004. november.

[2] KOVÁCS P.: Statisztikai mintákat generáló algoritmusok *(Algorithms Generating Statistical Samples)*, DUF Informatikai szimpózium, Dunaújváros, 2004. november.

[3] KOVÁCS P.: Az informatika alkalmazása a közgazdasági képzésben *(Application of Informatics in Economist Education)*, Informatika a felsőoktatásban konferencia 2005, B szekció, Debreceni Egyetem Informatikai Kar, Debrecen, 2005. augusztus 24.-26.

[4] KOVÁCS P.: Az informatika oktatása és lehetőségei a jogászképzésben *(Teaching and Possibilities of Informatics in Law Education)*, Informatika a felsőoktatásban konferencia 2005, F szekció, Debreceni Egyetem Informatikai Kar, Debrecen, 2005. augusztus 24.-26.

[5] KOVÁCS P. – PETRES T.: A PETRES-féle *Red*-mutató eloszlásának vizsgálata *(Examination of the Distribution of PETRES' Red)*, Magyar Tudomány Hete a Dunaújvárosi Főiskolán Közgazdasági és menedzsment Konferencia, Dunaújváros, 2005. november 22.

[6] KOVÁCS P. – PETRES T. – LUKOVICS M.: A PETRES-féle *Red*-mutató alkalmazásának lehetőségei *(Application Possibilities of PETRES' Red)*, Magyar Tudomány Hete a Dunaújvárosi Főiskolán Közgazdasági és menedzsment Konferencia, Dunaújváros, 2006. november 16.

[7] KOVÁCS P. – PETRES T.: A PETRES-féle *Red*-mutató ismertetése *(Presentation of PETRES' Red)*, VI. Természet-, Műszaki- és Gazdaságtudományok Alkalmazása Nemzetközi Konferencia, Szombathely, 2007. május 18.

[8] VILMÁNYI M. – KOVÁCS P.: Egyetemi-ipar együttműködések teljesítménye és lehetséges vizsgálati módszere *(Performance and Possible Examination Method of University-Industrial Co-operations)*, *"Kérdőjelek a régiók gazdasági fejlődésében"* Konferencia, Szeged, 2007. november 12.

[9] KOVÁCS P. : Az információs társadalom szerinti területi egyenlőtlenségek mérése *(Measurement of Regional Inequalities According to the Informational Society)*, *"Kérdőjelek a régiók gazdasági fejlődésében"* Konferencia, Szeged, 2007. november 13.