Available online at www.sciencedirect.com
ScienceDirect
IFAC PapersOnLine 54-1 (2021) 490–497

Maximal Correlation Techniques in Constructing Non-Linear Econometric Models

Kirill R. Chernyshov

V.A. Trapeznikov Institute of Control Sciences
65 Profsoyuznaya Street, Moscow 117997, Russia
(e-mail: myau@ipu.ru)

Abstract: Constructing non-linear stochastic econometric models is considered, from the point of view of revealing and proper reflecting quantitatively stochastic dependencies between input and output variables involved in the models. The dependencies are investigated by applying a corresponding measure of dependence, the maximal correlation. The approach enables one both to obtain corresponding non-linear transformations describing relationships between the input and output variables and to assure the existence of the model parameters estimates. As well, the approach provides obtaining the model identifiability conditions and quantitative description of the model non-linearity. The presentation is preceded with an analysis of a close approach available in the tutorial literature, with emphasizing certain delusions involved.

Keywords: Econometrics, Maximal correlation, Measures of dependence, Non-linear models, System identification.

1. PRELIMINARIES

Generically, econometrics approaches are concerned with the systematic study and appraisal of general principles, statistical procedures and modeling strategies, as well as methodological presuppositions that underlie econometric methods, with a view of evaluating their effectiveness in achieving the primary objective of "learning from data" about economic phenomena of interest (Spanos, 2012). The latter, learning from data, conditions the direct application of system identification techniques, meanwhile, the uncertainty conventionally involved in the model construction implies the consideration of namely stochastic approaches.

To solve stochastic systems identification problems, one is forced unavoidably to apply measures of dependence of random values. As the most frequent case within the problems, conventional linear covariance and correlation measures are applied. Their use directly follows from an identification problem statement itself when the statement is based on a quadratic type criterion. The main advantage of linear correlation-based measures of dependence is convenience of their use involving both a possibility to derive explicit analytical expressions in the course of determining required system characteristics and relative simplicity to construct their estimates within cases concerned with the availability of both independent and depended data. At the same time, the main drawback of the linear correlation-based measures of dependence is concerned with cases, when such measures may vanish within a stochastic and even deterministic dependence of considered random values.

In order to overcome such a possibility, involving more complicated, non-linear, measures of dependence is motivated to be applied within stochastic systems identification problems. Meanwhile, non-linear phenomena are considerably implied just by the complexity of the probabilistic dependence of variables involved in constructed econometric models.

Non-linear modeling-based approaches find increasing attention within the econometric literature, with dedicating, in particular, special journal issues to these studies (Rombouts et al., 2020, Songnian Chen and Qi Li, 2010, Zongwu Cai et al., 2018). As well, Kline and Tamer (2020) consider applying non-linear, in particular, linear-in-mean, models within econometric analysis of social interactions, and Bargigli (2017), within agent-based econometric modeling. In the present paper, a problem of non-parametric/parametric identification of non-linear models based on applying such a measure of dependence as the maximal correlation is considered. The approach enables one to construct a procedure that assures obtaining required non-linear transformations of the model variables and providing the existence of model parameters estimates.

The paper is organized as follows. In the next Section, issues concerned with the notion of consistent measures f dependence are presented. In Section 3, an analysis of certain particularities available in tutorial literature and concerned with applying consistent measures of dependence in the system identification is implemented. The main paper derivation is presented in Section 4, accompanied by a corresponding example in Section 5.

2. BASIC NOTES ON MEASURES OF DEPENDENCE

To characterize in concise and uniform manner requirements, which a measure of dependence is to meet to, in the paper of Rényi (1959b) seven axioms have been formulated and, finally, have been commonly accepted in the scientific

10.1016/j.ifacol.2021.08.057

community, despite the availability of certain discussion, as the most natural conditions describing the required properties of a measure of dependence $\mu(x,y)$ between two random values x and y.

A) $\mu(x,y)$ is defined for any pair of random values x and y, with no one of them being almost surely a constant.

B) $\mu(x,y)=\mu(y,x)$.

C) $0\le\mu(x,y)\le 1$.

D) $\mu(x,y)=0$ if and only if x and y are independent.

E) $\mu(x,y)=1$ if there exists a strict dependence between x or y, what means either $y=\varphi(x)$, or $x=\psi(y)$, with φ and ψ being some Borel measurable functions.

F) If φ and ψ are some one-to-one Borel measurable functions, then $\mu(\varphi(x),\psi(y))=\mu(x,y)$.

G) If joint probability distribution of x and y is Gaussian, then $\mu(x,y)=|\mathbf{corr}(x,y)|$, with $\mathbf{corr}(X,Y)$ being the ordinary correlation coefficient between x and y.

Among different measures of dependence, the ordinary correlation coefficient $\mathbf{corr}(x,y)$ is, of course, the most widely known and used. A more thin approach to elicit stochastic dependence between random values is concerned with applying the correlation ratio $\theta(x,y)$

$$\theta(x,y)=\frac{\mathbf{var}\left(\mathbf{E}(^{y}/_{x})\right)}{\mathbf{var}(y)},\mathbf{var}(y)>0, \tag{1}$$

and the maximal correlation coefficient $S(x,y)$ that has originally been introduced by Gebelein (1941) and subsequently investigated by Rényi (1959a, b) and Sarmanov (1963a, b),

$$S(x,y)=\sup_{B,C}\frac{\mathbf{cov}(B(y),C(x))}{\sqrt{\mathbf{var}(B(y))\mathbf{var}(C(x))}}, \tag{2}$$

$\mathbf{var}(B(y))>0,\mathbf{var}(C(x))>0$.

In (1) and (2) the supremum is taken over the set of Borel measurable functions; in turn $\mathbf{E}(^{\cdot}/_{\cdot})$ stands for the conditional mathematical expectation, $\mathbf{var}(\cdot)$, for the variance, and $\mathbf{cov}(\cdot)$, for the covariance.

Meanwhile, the paper of Rényi (1959b) shows the maximal correlation coefficient $S(x,y)$ in (2) to meet all the axioms presented only, while the ordinary correlation coefficient $\mathbf{corr}(x,y)$ and the correlation ratio $\theta(x,y)$ in (1) do not, in particular, axioms D, E, F are not met for the correlation coefficient, and axioms D, F, for the correlation ratio. Accordingly, in the light of the Rényi axioms, measures of dependence, which meet all of them, are natural to be referred to as consistent in the strict Rényi sense in the full analogy to the consistency in the sense of Kolmogorov, which presumes meeting axiom D at least only (Sarmanov and Zakharov, 1960).

To apply consistent measures of dependence is system identification problems one should take into account their certain particularities and features, neglecting which may result in inadequate inferences concerning the entity of solved problems; for instance, such neglecting may be revealed in books of Pashchenko (2001, 2006), and a corresponding example will be considered below as having a direct relation to the problem of deriving non-linear econometric models.

3. REVISING THE CONSISTENT "LEAST SQUARES METHOD" AS PER PASHCHENKO (2001, 2006)

Namely, within the consideration of a possibility of vanishing ordinary correlation coefficient under the dependence of corresponding random values, Pashchenko (2006) in Section 1 of this tutorial considers "a stochastic system of the form

$$y=a_{\cdot 1}x_1^2+a_{\cdot 2}x_2^2, \tag{3}$$

where x_1,x_2 are random input signals with Gaussian probability distributions $N_1(0,1), N_2(0,1)$.

The corresponding model is searched in the form

$$y=a_{\cdot 1}x_1+a_{\cdot 2}x_2.$$

The least squares estimates for this model have the form

$$A=(X^TX)^{-1}X^TY, \tag{4}$$

where

$$(X^TX)=\begin{bmatrix}K_{x_1x_1} & K_{x_1x_2}\\ K_{x_2x_1} & K_{x_2x_2}\end{bmatrix}; X^TY=\begin{pmatrix}K_{yx_1}\\ K_{yx_2}\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}.$$

Hence the unknown parameters estimate vector is equal to $(a_1,a_2)^T=(0,0)$; and $y=0$ for any magnitudes of x_1,x_2. From here it follows that input actions do not influence the output system signal, what does not correspond to reality. Analogous reasoning is valid, if the system is of the form

$$y=a_{\cdot 1}x_1^2+a_{\cdot 2}x_2^2+\xi,$$

where ξ is a random noise with the parameters $\sim N_\xi(0,1)$, being statistically independent of x_1 and x_2.

Thus, the least squares method widely used under different problems of processing experimental data in natural sciences, economics, sociology, and other sciences may give intrinsically wrong results characterizing an investigated phenomenon or system in the case of incorrect selecting the model structure, or incorrect the identification problem statement and selecting a solution" (pages 72, 73 in (Pashchenko, 2006)): end of the citation.

Regarding the citation from the tutorial of Pashchenko (2006) presented, the following should be pointed out. Firstly, the notation used in expression (4):

$X=(x_1\quad x_2), Y=y, K_{x_ix_j}=\mathbf{E}\left(x_ix_j\right), i,j=1,2, K_{yx_j}=\mathbf{E}\left(yx_j\right), j=1,2,$

where $\mathbf{E}(\cdot)$ stands for the mathematical expectation.

Secondly, estimate (4) is not related to the least squares. It is the mean squares minimization estimate followed from the criterion

$$\min_{a_1,a_2}\mathbf{E}\left((y-a_{\cdot 1}x_1+a_{\cdot 2}x_2)^2\right). \tag{5}$$

Expression (5) and estimate (4) implied by it are not the least squares!!! To refer to the least squares in various aspects and applications one can mention the books of Grabe (2005), Wolberg (2006), Zimmerman (2020), and many others.

Thirdly, the affirmation made that “input actions do not influence the output system signal, what does not correspond to the reality” is incorrect, since the consideration is concerned with a model constructed by the researcher, and, consequently, is concerned with the output signal of the model rather than the system (plant). Moreover, if one will consider the identification problem statement based on the mean square criterion, and whose solution gives rise namely to relationship (4), then namely zero magnitudes of the coefficients in this linear model just provide the minimal magnitude of the identification criterion. In other words, the identification problem solution is achieved: the *model* built,

$$y_M^0 \equiv 0,$$

is *linear*, and for *any other linear model* $\tilde{y}_M$ the inequality

$$\mathbf{E}(a_1 x_1^2 + a_2 x_2^2 - \tilde{y}_M)^2 > \mathbf{E}(a_1 x_1^2 + a_2 x_2^2 - y_M^0)^2$$

holds.

Thus the affirmation of Pashchenko (2006) that the least squares method does not give “intrinsically wrong results” is wrong itself, while, otherwise, it is an efficient indicative tool, indicating the researcher to the necessity of changing the initial identification problem statement, if the solution obtained is not acceptable for the researcher due to other characteristics that had not been formalized in the initial problem statement..

As to the tutorial of Pashchenko (2006), the approach to the “solution” of the “problem” stated in the example represented by expressions (3), (4) is presented in a section entitled “Consistent least squares method”, with the frivolity of the terminology used causing an unpleasant surprise and confusion (see also (Chernyshov, 2006, 2007): “The use of the apparatus of generalized correlation functions and generalized correlations leads to creating a consistent least squares method.

Let us consider a static system described by the non-linear equation

$$Y = F(\bar{x}), \tag{6}$$

where $\bar{x}$ is an n-dimensional vector. The system model is searched in the class of semi-linear systems

$$B(y) = AC(x), \tag{7}$$

where B and C are some non-linear transformers, A is an unknown parameters vector” (p. 272 in (Pashchenko, 2006)).

Besides the incorrectness of the notations used in the tutorial of Pashchenko (2006), one should also point out the unsuccessful term of “semi-linear systems”.

In (7), the transformations B and C are determined “as the solution of the following problem

$$(B, C) = \underset{B,C}{\operatorname{argsup}} \frac{\mathbf{cov}(B(y), C(x))}{\sqrt{\mathbf{var}(B(y))\mathbf{var}(C(x))}}, \tag{8}$$

$$\mathbf{var}(B(y)) = \mathbf{var}(C(x)) = 1, \mathbf{E}(B(y)) = \mathbf{E}(C(x)) = 0.$$

Then, the vector of unknown parameters A, corresponding to the mean square error minimum, under the transformers B and C calculated, meets the equation

$$A = ([C(x)]^T [C(x)])^{-1} [C(x)]^T B(y), \tag{9}$$

where the matrix $([C(x)]^T [C(x)])$ has the form

$$([C(x)]^T [C(x)]) = \begin{pmatrix} 1 & r^{\Phi}_{x_1 x_2} & \cdots & r^{\Phi}_{x_1 x_n} \\ r^{\Phi}_{x_2 x_1} & 1 & \cdots & r^{\Phi}_{x_2 x_n} \\ \cdots & \cdots & \cdots & \cdots \\ r^{\Phi}_{x_n x_1} & r^{\Phi}_{x_n x_2} & \cdots & 1 \end{pmatrix}, \tag{10}$$

while the matrix $[C(x)]^T B(y)$ is the column-vector of the normalized maximal correlation coefficients

$$[C(x)]^T B(y) = \left(r^{\Phi\max}_{y x_1}, \quad r^{\Phi\max}_{y x_2}, \quad \ldots, \quad r^{\Phi\max}_{y x_n} \right)^T. \tag{11}$$

In the general case, when the transformations B and C are searched in the form

$$(B, C) = \arg \frac{\mathbf{cov}(B(y), C(x))}{\sqrt{\mathbf{var}(B(y))\mathbf{var}(C(x))}}, \tag{12}$$

the matrix $[C(x)]^T B(y)$ takes the form:

$$[C(x)]^T B(y) = \left(r^{\Phi}_{y x_1}, \quad r^{\Phi}_{y x_2}, \quad \ldots, \quad r^{\Phi}_{y x_n} \right)^T, \tag{13}$$

where $r^{\Phi}_{y x_i}$ is the normalized functional correlation coefficient between the output. y, and i-th input, x_i, variables of the modeled system.

Taking into account that the functional correlation functions and coefficients are always non-zero, when the values x_i and y are statistically independent, the method proposed is a consistent variant of the least squares method. The problem presented in the above example has been solved” (pp. 272-273 in (Pashchenko, 2006)).

Leaving out of the consideration scopes “general case” (12), (13), one should, in the first turn, specify the notations used in formulae (7)-(11). If one will be based on model (7) and expressions (9)-(11), then $C(x)$ is a vector. If one will be based on expressions (8), then $C(x)$ is a scalar. Thus, and taking into account designation (10), (11), one may conclude that expression (8) represents, perhaps, some simplified record of the following criterion

$$(B, C_i) = \underset{B,C}{\operatorname{argsup}} \frac{\mathbf{cov}(B(y), C_i(x_i))}{\sqrt{\mathbf{var}(B(y))\mathbf{var}(C_i(x_i))}},$$

$$\mathbf{var}(B(y)) = \mathbf{var}(C_i(x_i)) = 1, \mathbf{E}(B(y)) = \mathbf{E}(C_i(x_i)) = 0,$$
$$i = 1, \ldots, n,$$

in other words, in (8) the subscript i under the designation of the transformation C_i and input variable x_i, as well as the right-hand part of expression (7) should be understood as

$\sum_{i=1}^{n} a_i C_i(x_i)$,

where a_i are the vector A coefficients.

Correspondingly, the "normalized functional correlation coefficient between the output, y, and i-the input, x_i, variables of the modeled system" is defined by the relationship

$$r_{yx_i}^{\Phi} = \frac{\mathbf{cov}\left(B(y), C_i(x_i)\right)}{\sqrt{\mathbf{var}\left(B(y)\right)\mathbf{var}\left(C_i(x_i)\right)}}.$$

And after that, one should point out that the approach considered (presented in Section 4.3.1 of the tutorial (Pashchenko, 2006)) is not related to the problem represented by the example (expressions (3), (4)), since expressions (3), (4) are concerned with the problem of *parametric* identification under a *given* model structure, namely, linear model. Thus, in the example represented by expressions (3), (4) and in the approach, represented by expressions (6), (7) and subsequent, are principally different, since there are principally different both the models subject to the identification (compare expressions of the form $y = Ax$ for the example, represented by expressions (3), (4), and expression $B(y) = AC(x)$ (7)) and the identification criteria (within expressions (3), (4) the unknown parameters vector is subject to the identification only, in the "consistent least squares method" there are to be determined the non-linear transformations and parameters vector).

4. CONSTRUCTING AN ADDITIVE MODEL

In the present Section, an approach to construct non-linear additive models by the use of such a consistent measure of dependence as maximal correlation (2) is considered. Namely, a model output variable y, and n model input variables $x_1, \dots, x_n$ are considered for sake of simplicity but without the loss of generality as zero-mean unit variance random values. At the first stage of constructing the additive model, local relationships of the form

$$B_i(y) = \nu_i C_i(x_i) + e_i, \quad i = 1, \dots, n \tag{14}$$

are considered.

In (14), $B_i(y), C_i(x_i)$ are unknown non-linear Borel measurable functions, meeting the conditions

$$\mathbf{E}\left(B_i(y)\right) = \mathbf{E}\left(C_i(x_i)\right) = 0, \quad \mathbf{var}\left(B_i(y)\right) = \mathbf{var}\left(C_i(x_i)\right) = 1, \quad i = 1, \dots, n, \tag{15}$$

ν_i are unknown scalar coefficients, e_i and play the role of fraction-not-explained, $i = 1, \dots, n$. Within the problem statement, these e_i are considered as mutually stochastically independent, and stochastically independent of $x_1, \dots, x_n$.

In turn, $B_i(y), C_i(x_i)$, $i = 1, \dots, n$ in (14) are subject to the identification in accordance to the criterion of the fraction-not-explained variance minimization:

$$\mathbf{var}(e_i) \to \inf_{B, C_i}, \quad i = 1, \dots, n. \tag{16}$$

By virtue of model (14) and conditions (15), one can finally write

$$\mathbf{var}(e_i) = 1 - \nu_i \mathbf{E}\left(\left(B(y) C_i(x_i)\right)\right) + \nu_i^2, \quad i = 1, \dots, n,$$

and, hence,

$$\mathbf{var}(e_i) \geq 1 - \nu_i S_i^{abs}(x_i, y) + \nu_i^2$$
$$\forall \alpha_i : \ \mathrm{sign}(\nu_i) = \mathrm{sign}\left(S_i^{abs}(x_i, y)\right)$$
$$i = 1, \dots, n,$$

where

$$S_i^{abs}(x_i, y) = \sup_{B, C_i} \left| \mathbf{E}\left(\left(B(y) C_i(x_i)\right)\right) \right| \tag{17}$$

is the maximal (in the absolute value) correlation coefficient of the output variable y and input variables x_i, $i = 1, \dots, n$ calculated under conditions (15). In other words,

$$\left|S_i^{abs}(x_i, y)\right| = S_i(x_i, y),$$

where $S_i(x_i, y)$ is the maximal correlation coefficient between the random values x_i and y defined by (2).

Namely, in the comparison to (Gebelein, 1941, Rényi, 1959a,b), the maximal in the absolute value correlation coefficient has been introduced in the fullness of time by Sarmanov (1963a,b) and investigated in a number of papers.

Let

$$\left(B_i^{abs}, C_i^{abs}\right) = \underset{B, C_i}{\mathrm{argsup}} \left| \mathbf{E}\left(\left(B(y) C_i(x_i)\right)\right) \right|, \tag{18}$$

$$\left(B_i^{*}, C_i^{*}\right) = \underset{B, C_i}{\mathrm{argsup}} \mathbf{E}\left(\left(B(y) C_i(x_i)\right)\right), \tag{19}$$

$i = 1, \dots, n.$

Then, summarizing models (14) over $i = 1, \dots, n$ at $B = B_i^{abs}$ and $C_i = C_i^{abs}$ leads to the total model

$$\mathbf{B}^{abs}(y) = \sum_{i=1}^{n} a_i C_i^{abs}(x_i) + \varepsilon, \tag{20}$$

where

$$\mathbf{B}^{abs}(y) = \frac{1}{n}\sum_{i=1}^{n} B_i^{abs}(y),$$

$$a_i = \frac{\nu_i}{n}, i = 1, \dots, n,$$

and

$$\varepsilon = \frac{1}{n}\sum_{i=1}^{n} e_i$$

is the total fraction-not-explained.

In total model (20), the parameters $a_i, i = 1, \dots, n$ are subject to the identification in accordance to the criterion of the total fraction-not-explained variance, $\mathbf{var}(\varepsilon)$, minimization:

$$\mathbf{var}(\varepsilon) \to \inf_A, \tag{21}$$

where $A = (a_1, \dots, a_n)^T$.

Then, from model (20) and criterion (21) it directly follows

$$\mathbf{E}\left(\mathbf{B}^{abs}(y)\begin{pmatrix} C_1^{abs}(x_1) \\ \vdots \\ C_n^{abs}(x_n)\end{pmatrix}\right) = \begin{pmatrix} 1 & r_{x_1x_2}^{abs} & \cdots & r_{x_1x_n}^{abs} \\ r_{x_2x_1}^{abs} & 1 & & \vdots \\ \vdots & & \ddots & r_{x_{n-1}x_n}^{abs} \\ r_{x_nx_1}^{abs} & \cdots & r_{x_nx_{n-1}}^{abs} & 1 \end{pmatrix} A, \tag{22}$$

where

$$r_{x_ix_j}^{abs} = \mathbf{E}\left(C_i^{abs}(x_i)C_j^{abs}(x_j)\right), \quad i,j = 1, \dots, n \tag{23}$$

Accordingly, in the expression (22) left-hand part the column-vector components, $\mathbf{s}_{yx_i}^{abs}, i = 1, \dots, n$, are of the form

$$\mathbf{s}_{yx_i}^{abs} = \frac{1}{n}\left(S_i^{abs}(y, x_i) + \sum_{j\neq i}^n \mathbf{E}\left(B_j^{abs}(y)C_i^{abs}(x_i)\right)\right), \quad i = 1, \dots, n. \tag{24}$$

If all $B_i^{abs}, i = 1, \dots, n$ in (18) are equal,

$$B_i^{abs} = B^{abs}, i = 1, \dots, n,$$

then model (20) directly takes the form

$$B^{abs}(y) = \sum_{i=1}^n a_i C_i^{abs}(x_i) + \varepsilon. \tag{25}$$

The feature of approach (14)-(21) is to ensure, from the one hand side, the optimal selection of the corresponding non-linear transformations, and, from another hand side, the uniqueness of at least one component of the parameter vector A estimate.

Regarding the existence of transformations (18), let $p(x_i)$, $p(y)$, $p(y, x_i)$ be correspondingly the marginal and joint probability distribution densities of the input x_i and output y random values, the square of the stochastic kernel $p(y, x_i)/\sqrt{p(x_i)p(y)}$ be integrable

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{p^2(y,x_i)}{p(y)p(x_i)} dy dx_i < \infty. \tag{26}$$

Let, again, from the condition

$$B(y) - C_i(x_i) \equiv 0 \text{ a.s.}$$

it follows that B and C_i are identically zero transformation. Then, in accordance to (Rényi, 1959b, Sarmanov and Zakharov, 1960) there exist transformations B_i^{abs}, C_i^{abs} meeting (18).

In turn, the uniqueness of at least one component of the parameter vector A estimates is achieved on the basis of non-vanishing at least one of the equation (22) left-hand part column vector components. As some of $S_i^{abs}(x_i, y)$ in (17) ((24)) vanishes, all $\mathbf{E}\left(B_j^{abs}(y)C_i^{abs}(x_i)\right)$, $j \neq i$ in (24) vanish. Accordingly, vanishing all $S_i^{abs}(x_i, y)$, $i = 1, \dots, n$ in (17) ((24)) results in the non-identifiability of the model described by expressions (14)-(20) at all.

From another hand side, the case when at least one of $S_i^{abs}(x_i, y)$, $i = 1, \dots, n$ is non-zero, but all $\mathbf{s}_{yx_i}^{abs}, i = 1, \dots, n$ in (24) are zeros looks very doubtfully, but, nevertheless, possible from a formal point of view. Let the case described is hold for some κ-th of the input variables $x_i, i = 1, \dots, n$ of model (20) with corresponding optimal output transformation B_κ^{abs}. Then, within such a case, model (20) can be rewritten in the form

$$\frac{1}{n}B_\kappa^{abs}(y) = \sum_{i=1}^n a_i C_i^{abs}(x_i) - \frac{1}{n}\sum_{i\neq\kappa}^n B_i^{abs}(y) + \varepsilon. \tag{27}$$

Within model description (27) and criterion (21), problem statement (20), (21) is equivalent to the problem statement with the model description

$$\frac{1}{n}B_\kappa^{abs}(y) = \sum_{i=1}^n a_i C_i^{abs}(x_i) - \frac{a_{n+1}}{n}\sum_{i\neq\kappa}^n B_i^{abs}(y) + \varepsilon \tag{28}$$

and the identification criterion

$$\mathbf{var}(\varepsilon) \to \inf_{\tilde{A}:\, a_{n+1}=1}, \tag{29}$$

where $\tilde{A} = (a_1, \dots, a_n, a_{n+1})^T$.

Thus, constrained optimization problem (28), (29) leads to the equation, being an equivalent of that of (22),

$$\mathbf{E}\left(\frac{B_\kappa^{abs}(y)}{n}\begin{pmatrix} C_1^{abs}(x_1) \\ \vdots \\ C_n^{abs}(x_n) \\ -\frac{1}{n}\sum_{i\neq\kappa}^n B_i^{abs}(y)\end{pmatrix}\right) = \begin{pmatrix} 1 & r_{x_1x_2}^{abs} & \cdots & r_{x_1x_n}^{abs} & \mathbf{r}_{x_1y}^{abs} \\ r_{x_2x_1}^{abs} & 1 & & \vdots & \mathbf{r}_{x_2y}^{abs} \\ \vdots & & \ddots & r_{x_{n-1}x_n}^{abs} & \vdots \\ r_{x_nx_1}^{abs} & \cdots & r_{x_{n-1}x_n}^{abs} & 1 & \mathbf{r}_{x_ny}^{abs} \\ \mathbf{r}_{x_1y}^{abs} & \mathbf{r}_{x_2y}^{abs} & \cdots & \mathbf{r}_{x_ny}^{abs} & \mathbf{r}_{yy}^{abs} \end{pmatrix} \tilde{A}. \tag{30}$$

The κ-th component of the equation (30) left-hand part column vector is

$$\frac{S_\kappa^{abs}(y,x_\kappa)}{n} \neq 0,$$

what ensures the uniqueness of the estimate of $\tilde{A}$ in (29) or, equivalently, uniqueness of A in (22). In equation (30),

$$\mathbf{r}_{x_iy}^{abs} = -\frac{1}{n}\sum_{j\neq\kappa}^n \mathbf{E}\left(B_j^{abs}(y)C_i^{abs}(x_i)\right),$$

$$\mathbf{r}_{yy}^{abs} = n^{-2}\left(n-1+\sum_{\substack{i\neq j\\ i,j\neq \kappa}}^{n} \mathbf{E}\left(B_i^{abs}(y)B_j^{abs}(y)\right)\right).$$

5. SOME PROPERTIES OF THE MAXIMAL CORRELATION AND AN EXAMPLE

Due to condition (26), for the probability distribution density $p(y,x)$ the bilinear expansion converging in the mean is valid (Sarmanov and Zakharov, 1960)

$$p(y,x) = p(y)p(x)\left(1+\Re_{yx}\right),$$

and

$$\Re_{yx} = S^{abs}(y,x)B^{abs}(y)C^{abs}(x) + \\ +\sum_{k=2}^{\infty} S_k(y,x)B^{(k)}(y)C^k(x)\Big),$$

where

$$S^{abs}(y,x) \geq S_2(y,x) \geq \cdots \geq 0,$$

are absolute values of the eigenvalues of the stochastic kernel $p(y,x)/\sqrt{p(x)p(y)}$, and $B^{(k)}(y), C^{(k)}(x)$, $k = 2,3,...$; are the pairs of eigenfunctions corresponding to these eigenvalues S_k, $k = 2,3,....$ Meanwhile, accordingly, $S^{abs}(y,x)$ is the largest (in the absolute value) eigenvalue of the stochastic kernel, corresponding to the first eigenfunctions $B^{abs}(y), C^{abs}(x)$ of the stochastic kernel. The process of determination of the pair of the operators $B^{abs}(y), C^{abs}(x)$ is natural, in accordance to paper (Sarmanov and Zakharov, 1960), to be referred to as the maximal arithmetization of the probability distribution defined by the probability distribution density $p(y,x)$.

As pointed out above, the maximal correlation is a consistent measure of dependence of random values, which exhaustively characterizes their mutual interconnection in the stochastic sense, while correlation ratio (1) and, all the more so, the ordinary correlation may give an understated value of the connection measure of random processes, and, as known, there exist examples when the correlation function may vanish even under the availability of a deterministic functional connection between random values (Rajbman, 1981, Rényi, 1959b).

Moreover, there exist examples, when even in the case of the linear regression of the output variable on the input one and wise versa, the dependence between the variables is non-linear and is adequately characterized by the maximal correlation only. Indeed, let the joint probability distribution of the input x and output y variables of a system be given by the probabilistic distribution density (Sarmanov and Bratoeva, 1967)

$$p^F(x,y) = \frac{1}{3\sqrt{3}\pi}\times \\ \times\left\{\exp\left(-\frac{2}{3}(x^2+y^2+xy)\right)+2\exp\left(-\frac{2}{3}(x^2+y^2-xy)\right)\right\} = \\ = \frac{e^{-\frac{x^2+y^2}{2}}}{2\pi}\left\{1+\sum_{k=1}^{\infty}\frac{2+(-1)^k}{3\cdot 2^k}P_k^H(x)\,P_k^H(y)\right\} = \\ = \frac{e^{-\frac{x^2+y^2}{2}}}{2\pi}\{1+\sum_{k=1}^{\infty} c_k P_k^H(x)\,P_k^H(y)\}, \quad (31)$$

where

$$P_k^H(u) = \frac{(-1)^k}{\sqrt{k!}} e^{u^2/2}\frac{d^k}{du^k}e^{-u^2/2} = \\ = \frac{1}{\sqrt{k!}}\left\{u^k - \frac{k(k-1)}{1!\cdot 2}u^{k-2}+\cdots\right\} \quad (32)$$

are the Hermite polynomials. In this case, the correlation is linear; the correlation coefficient

$$c_1 = \mathbf{corr}(x,y) = {}^1\!/_6, \quad (33)$$

the regression lines have the equations

$$\mathbf{E}({}^y\!/_x) = {}^x\!/_6,\ \mathbf{E}({}^x\!/_y) = {}^y\!/_6,$$

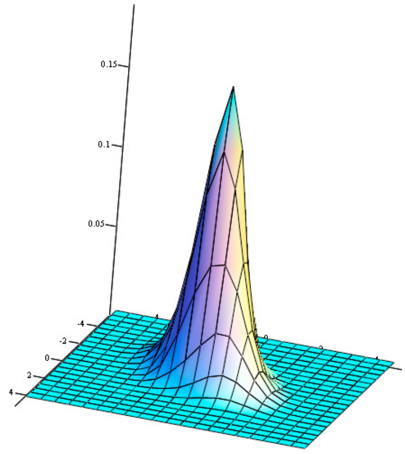
but $P_1^H(x) = x$ is not the first eigenfunction, and ordinary correlation coefficient (33) is not the first eigenvalue. The first eigenvalue

$$c_2 = S^{abs}(x,y) = {}^1\!/_4 \quad (34)$$

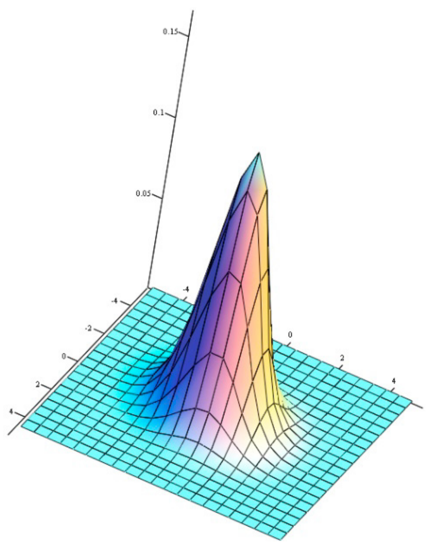
is the maximal correlation coefficient, and $P_2^H(x)$ is the first eigenfunction (Sarmanov and Bratoeva, 1967). From (32) it directly follows the first eigenfunction to have the form

$$\frac{u^2-1}{\sqrt{2}} \quad (35)$$

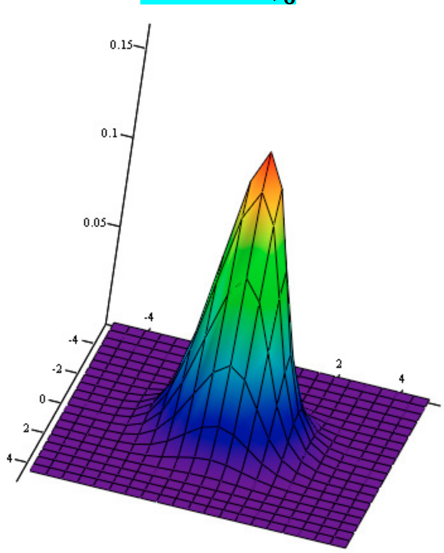
Fig. 1 display the shape of probability distribution density $p^F(x,y)$ in the comparison with the Gaussian one with correlation coefficient (33), as well in the comparison with Gaussian zero correlation probability distribution density. It can be easily seen that density $p^F(x,y)$ differs more significantly from the bivariate Gaussian one with the correlation coefficient $c_1 = \mathbf{corr}(x,y) = {}^1\!/_6$ than the latter from the Gaussian zero correlation probability distribution density. This is a manifestation of the internal dependence structure of density $p^F(x,y)$ defined by the specific eigenfunction expansion.



Density $p^F(x,y)$ (31)



Gaussian bivariate density with the correlation coefficient $c_1 = \mathbf{corr}(x,y) = {}^1/_6$



Gaussian bivariate zero correlation density with Laplace marginals

Fig. 1. Comparison of shapes of some probability distribution densities.

Let now, y be output variable of system with two input variables, x_1 and x_2, meanwhile, the joint probability distribution densities $p(y,x_1)$, $p(y,x_2)$, $p(x_1,x_2)$ (being, of course, not known to the researcher) be of the following form: $p(y,x_1) = p^F(y,x_1)$, with $p^F(y,x_1)$ being defined by (31), $p(y,x_2)$, $p(x_1,x_2)$ be joint Gaussian with Laplace marginal distributions and the correlation coefficients ρ_{yx_2} and $\rho_{x_1x_2}$ correspondingly.

Then on the basis of the theoretical deriving presented in the preceding Section and the properties of probability distribution density (31), one can write the following model

$$\frac{1}{2}\left(B_1^{abs}(y) + B_2^{abs}(y)\right) = a_1 C_1^{abs}(x_1) + a_2 C_2^{abs}(x_2) + \varepsilon. \quad (36)$$

In model (36), the transformations B_1^{abs}, C_1^{abs} are of the form:

$$B_1^{abs}(u) = C_1^{abs}(u) = \frac{u^2 - 1}{\sqrt{2}}$$

due to properties (32)-(35) of density (31) and normalization conditions (15).

Again, the transformations B_2^{abs}, C_2^{abs} in model (36) are identical ones,

$$B_2^{abs}(u) = C_2^{abs}(u) = u,$$

due to the bivariate Gaussian distribution properties.

Then model (36) takes the form

$$\frac{1}{2}\left(\frac{y^2-1}{\sqrt{2}} + y\right) = a_1 \frac{x_1^2-1}{\sqrt{2}} + a_2 x_2 + \varepsilon.$$

Accordingly, equation (22) to determine the parameters a_1 and a_2 will look as follows

$$\begin{pmatrix} \frac{1}{8} \\ \frac{\rho_{yx_2}}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix},$$

since

$$\mathbf{E}\left(y(x_1^2 - 1)\right) = \mathbf{E}\left((y^2 - 1)x_2\right) = \mathbf{E}\left((x_1^2 - 1)x_2\right) = 0. \quad (37)$$

So, by virtue of (37) the final model calculated is of the form

$$\frac{1}{2}\left(\frac{y^2-1}{\sqrt{2}} + y\right) = \frac{1}{8} \cdot \frac{x_1^2-1}{\sqrt{2}} + \frac{\rho_{yx_2}}{2} \cdot x_2 + \varepsilon.$$

To estimate quantitatively the model non-linearity nature, a corresponding degree of non-linearity can be applied. In the fullness of time, the maximal correlation application approach to the non-linearity degree has been proposed by (Durgaryan and Pashchenko, 1985). Applying it, for models of the form of (20), the following non-linearity degree may be introduced:

$$\varpi = \max_{i=1,\dots,n} \sqrt{1 - \left(\frac{\mathbf{corr}(x_i,y)}{S^{abs}(x_i,y)}\right)^2}, \quad (38)$$

with

$\frac{0}{0} = 0.$

Of course, definition (38) has a sense if at least one of $S^{abs}(x_i, y)$, $i = 1, \dots, n$ is *non-zero*. Otherwise, the model should be considered as non-identifiable. With regard to the example considered, the non-linearity degree is

$$\varpi = \frac{\sqrt{5}}{3} \approx 0.745.$$

6. CONCLUSIONS

In the paper, a non-parametric/parametric distribution-free approach to constructing non-linear econometric models has been presented. Its feature is applying a consistent measure of dependence, the maximal correlation coefficient. Within the approach, determining the maximal correlation between the input and output variables directly implies determining corresponding non-linear transformations of these variables, with accompanying the procedure with estimating parameters of the linear part of the model. Meanwhile, the possibility to obtain conditions providing the existence of the parameters estimates is assured by the condition of non-vanishing of at least one of the maximal correlation coefficients, being the condition of the model identifiability.

REFERENCES

Bargigli, L. (2017). "Methods for Agent-Based Models", *Introduction to Agent-Based Economics*, Chapter 8, pp. 163-189.

Chernyshov, K.R. (2006). "A review on the book of I.V. Prangishvili, F.F. Pashchenko, B.P. Busygin "System Laws and Regularities in Electrodynamics, Nature, and Society", *Information and Control Systems*, no. 5, pp. 51-54. (in Russian)

Chernyshov, K.R. (2007). "Towards the education process support in the systems modeling branch", *Quality. Innovations. Education*, no. 9, pp. 39-50. (in Russian)

Durgaryan, I.S. and F.F. Pashchenko (1985). "Non-Parametric Identification of Nonlinear Systems", *IFAC Proceedings Volumes*, vol. 18, no. 5, pp. 433-437.

Gebelein, H. (1941). "Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung", *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 21, no. 6, pp. 364-379. (in German)

Grabe, M. (2005). *Measurement Uncertainties in Science and Technology*, Springer, 269 p.

Kline, B. and E. Tamer (2020). "Econometric analysis of models with social interactions", *The Econometric Analysis of Network Data*, Chapter 7, pp. 149-181.

Pashchenko, F.F. (2001). "Determining and modeling regularities via experimental data", In: *System Laws and Regularities in Electrodynamics, Nature, and Society*. Chapter 7, "Nauka", Moscow, pp. 411-521. ISBN 5-02-01308-5 (in Russian)

Pashchenko, F.F. (2006). *Introduction to consistent methods of systems modeling. Mathematical foundations of systems modeling*. Finansy i statistika Publ., Moscow, 328 p. ISBN 978-5-279-02922-8 (in Russian)

Rényi, A. (1959a). "New version of the probabilistic generalization of the large sieve", *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 10, no. 1–2, pp. 217-226.

Rényi, A. (1959b). "On measures of dependence", *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 10, no. 3–4, pp. 441-451.

Rombouts, J.V.K., Scaillet, O., Veredas D., and J.-M. Zakoian (2020). "Nonlinear financial econometrics JoE special issue introduction", *Journal of Econometrics*, vol. 217, no. 2, pp. 203-206.

Sarmanov, O.V and E.K. Zakharov (1960). "Measures of dependence between random variables and spectra of stochastic kernels and matrices", *Matematicheskiy Sbornik*, vol. 52(94), pp. 953-990. (in Russian)

Sarmanov, O.V. (1963a). "The maximum correlation coefficient (nonsymmetric case)", *Sel. Trans. Math. Statist. Probability*, vol. 4, pp. 207-210.

Sarmanov, O.V. (1963b). "Investigation of stationary Markov processes by the method of eigenfunction expansion", *Sel. Trans. Math. Statist. Probability*, vol. 4, pp. 245-269.

Songnian Chen and Qi Li (2010). "Annals Journal of Econometrics: Nonlinear and Nonparametric Methods in Econometrics", *Journal of Econometrics*, vol. 157, no. 1, pp. 3-5.

Spanos, A. (2012). "Philosophy of Econometrics", *Philosophy of Economics*, pp. 329-393.

Wolberg, J. (2006). *Data Analysis Using the Method of Least Squares. Extracting the Most Information from Experiments*, Springer, 250 p.

Zimmerman, D.L. (2020). *Linear Model Theory. With Examples and Exercises*, Springer, 504 p.

Zongwu Cai, Yongmiao Hong, and Shouyang Wang (2018). "Econometric Modeling and Economic Forecasting", *Journal of Management Science and Engineering*, vol. 3, no. 4, pp. 178-182.