1.4. Variance of complex-valued variables in modern statistics

Complex situation with the use of complex-valued variables models is caused by the fact that the branch of mathematical statistics studying random complex-valued variables seems undeveloped so far (*Svetunkov, 2018*). Scientists paid attention to statistical treatment of complex-valued variables dynamics in the 50-60s of the 20th century. R. Wooding (Wooding, 1956) was the first who formulated this task and offered the approach to presenting the complex random value based on the normal distribution. R.Arens (*Arens, 1957*) and I.S.Reed (*Reed, 1962*) introduced major concepts and characteristics of the random normally distributed complex-valued variable such as mathematical mean value, moment coefficients (including correlation coefficient), covariance, variance, etc. It was a priori supposed that the situation of independent normally distributed real and imaginary parts was under discussion, and only N.R.Goodman formulated the hypothesis more clearly (*Goodman, 1963*). Since that time scientists began to consider the distribution of complex random value as an aggregate of two independent normally distributed random values – real and imaginary parts. It was W. Feller who systematized these statements (*Feller, 1966*). Today this assumption about the autonomy of real and imaginary parts of the complex-valued random variable serves as a key prerequisite for mathematical statistics of complex-valued random variable.

As the scope of functions of modeling using complex-valued variables in various scientific spheres widened, scientists faced the necessity to develop the body of mathematical statistics which allowed to do it. Since this function is of interest not only for economists but for the specialists dealing with complex-valued variables in other sciences as well, an adapted least square method (*Tavares, 2007*) was proposed based on the real part of the complex random value variance. However, at this point the development of the statistical tools of the complex random value stopped – no tools for calculating complex-related correlations were offered, no tools for determining the confidence limits or other instruments of statistical treatment of random complex-valued variables were proposed.

Nowadays mathematical statistics deals with variance of independent constituents – real and imaginary parts – rather than variance of the complex-valued variable in general. In this case statistical characteristics of the complex-valued random variable are regarded as real values. To calculate the real characteristics of the complex random value, this value is multiplied by the complex conjugate. This procedure, as it is known, allows to find out the real characteristic of the complex number. Variance of a complex-valued variable is presented as mathematical mean value of squared absolute value of the corresponding centered variable (*Bliss, 2013*; *Panchev, 2013*; *Steven, 2010*; *Tuelay, 2011*):

$$D(z) = M[|z|] = M[|x_r + ix_i|] = M[(x_r + x_i)(x_r - x_i)] = M[x_r^2] + M[x_i^2], \quad (1.4.1)$$

where $$M[x_r^2] = M[(x_r - \overline{x}_r)^2] = D(x_r), \quad (1.4.2)$$

$$M[x_i^2] = M[(x_i - \overline{x}_i)^2] = D(x_i). \quad (1.4.3)$$

Viz:

$$D(z) = D(x_r) + D(x_i). \quad (1.4.4)$$

However, such interpretation of the complex-valued variable imposes the limits to statistical data treatment. Let us illustrate it by determining the correlations between two complex-valued random variables. Actually, we can calculate the pair correlation coefficient of the variables by using the correlation moment and the variance:

$$r_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y}. \quad (1.4.5)$$

Let us take this approach to calculate the correlation between the complex-valued random variables. The correlation moment is represented as a real value using one of the variable in the conjugate form, whereas the variance is calculated as real characteristics in accordance with (1.4.1) (*Miyabe*, 2015; *Panchev*, 2013). It should be noted that the correlation moment calculated as

$$\mu_{XY} = M[(x_r + ix_i)(y_r - iy_i)] \quad (1.4.6)$$

is not the real number. It is the complex number as when multiplying and grouping the summands, we obtain:

$$\mu_{XY} = M[x_r y_r] + M[x_i y_i] + i(M[x_i y_r] - M[x_r y_i]) =$$
$$\mu_{x_r y_i} + \mu_{x_i y_i} + i(\mu_{x_i y_r} - \mu_{x_r y_i}). \quad (1.4.7)$$

And only in the case when z_X=z_Y, the latter summond (1.4.7) with the imaginary part equals zero, and the correlation moment becomes a real number. In all other cases the correlation moment will be of complex type, thus the pair correlation coefficient between two complex-valued random variables will be the complex value. Keeping this in mind, scientists claim that they calculate the absolute value of the correlation moment, i.e. instead of (1.4.6) they use $\mathrm{Re}(\mu_{XY})$ (*Miyabe*, 2015).

However, let us find the pair correlation coefficient by using (1.4.6).

For the sake of simplicity, we assume that the discrete sequence of the complex random value z centred in relation to its arithmetic mean is under review, hence:

$$z - \overline{z} = x_r - \overline{x}_r + i(x_i - \overline{x}_i) \Leftrightarrow z = x_r + ix_i. \quad (1.4.8)$$

Sample value of the correlation coefficient (1.4.5) when using the real-valued variance (1.4.1) and the correlation moment (1.4.6) is of the form:

$$r = \frac{\mu_{XY}}{\sigma_X \sigma_Y} = \frac{\sum (y_i y_r + x_i x_r) + i(\sum (x_r y_i - y_r x_i))}{\sqrt{\sum (y_r{}^2 + y_i{}^2) \sum (x_r{}^2 + x_i{}^2)}} . \quad (1.4.9)$$

On the other hand, we should keep in mind that the pair correlation coefficient as related to real numbers was suggested in the 90s of the 19th century by K. Pearson to estimate linear interrelation of complex-valued variables. It was defined as geometric mean of regressions *y* by *x* and *x* by *y* (*Pearson, 2013)*:

$$r = \pm\sqrt{a_1 b_1} , \quad (1.4.10)$$

where the proportionality factors of simple regressions a_1 and b_1 are found by using the least square method.

To find the formula for calculating the sample value of the pair correlation coefficient for the case of the two complex-valued variables using K. Pearson’s approach (1.4.10), we shall handle that variant of the least square method which results from the assumption about real character of the complex-valued variable variance (*Tavares, 2007*). The complex regression coefficient of linear relationship between complex-valued variable *Y* and other complex-valued variable *X* by using this approach to the least square method will be calculated in the following way:

$$a = \frac{\sum (y_r + iy_i)(x_r - ix_i)}{\sum (x_r + ix_i)(x_r - ix_i)} = \frac{\sum (y_r + iy_i)(x_r - ix_i)}{\sum (x_r^2 + x_i^2)} . \quad (1.4.11)$$

Inverse relationship of complex-valued variable X to other complex random variable Y, represented in the linear form, has the following formula for calculating the complex regression coefficient found by using the least square method:

$$b = \frac{\sum (x_r + ix_i)(y_r - iy_i)}{\sum (y_r + iy_i)(y_r - iy_i)} = \frac{\sum (x_r + ix_i)(y_r - iy_i)}{\sum (y_r^2 + y_i^2)} . \quad (1.4.12)$$

By using these formulae of estimating the sample values of proportionality coefficients of regression lines *Y* by *X* and *X* by *Y* in (1.4.10), we obtain the formula for calculating the sample value of the complex pair correlation coefficient:

$$r = \sqrt{a_1 b_1} = \frac{\sum (x_r y_r + x_i y_i) + i\sum (x_i y_r - x_r y_i)}{\sqrt{\sum (y_r^2 + y_i^2) \sum (x_r^2 + x_i^2)}} . \quad (1.4.13)$$

Now let us compare formula (1.4.9) with formula (1.4.13). This should be one and the same coefficient which is calculated based on the same basic prerequisites. However, if the denominators of formulae (1.4.8) and (1.4.13) coincide, their numerators differ fundamentally from each other. These are different formulae that are used to calculate different coefficients, and these coefficients will give different values for one and the same sequence. This is why it seems unclear: should we use formula (1.4.9) or should we use formula (1.4.13) or none of these formulae can be used? The obtained result is contradictory, and it does not allow scientists to form the body of complex correlations.

Even if we agree with the suggestions made by the followers of the conception of real-value character of the complex-valued variables variance and use their real parts (*Miyabe, 2013*) instead of complex characteristics, the conflict will not be solved.

In reality, for (1.4.9) we will obtain:

$$\mathrm{Re}(\frac{\mu_{XY}}{\sigma_X\sigma_Y})=\mathrm{Re}(\frac{\sum(y_iy_r+x_ix_r)+i(\sum(x_ry_i-y_rx_i))}{\sqrt{\sum({y_r}^2+{y_i}^2)\sum({x_r}^2+{x_i}^2)}})=\frac{\sum(y_iy_r+x_ix_r)}{\sqrt{\sum({y_r}^2+{y_i}^2)\sum({x_r}^2+{x_i}^2)}}, \quad (1.4.14)$$

and for (1.4.13) we will obtain:

$$\mathrm{Re}(\sqrt{a_1b_1})=\mathrm{Re}(\frac{\sum(x_ry_r+x_iy_i)+i\sum(x_iy_r-x_ry_i)}{\sqrt{\sum(y_r^2+y_i^2)\sum(x_r^2+x_i^2)}})=\frac{\sum(x_ry_r+x_iy_i)}{\sqrt{\sum(y_r^2+y_i^2)\sum(x_r^2+x_i^2)}}. \quad (1.4.15)$$

As it can be seen from two results compared, different formulae and a contradictory result are obtained again. This is why P.Schreier and L. Scharf note that so far the research carried out in the area of the correlation of complex-valued random variables has produced deplorable results (*Schreier, 2010*).

1.5. Complex-valued variance

Exactly the same problems arise in the field of statistical hypotheses and other branches of mathematical statistics of complex-valued variables which to certain extent rely upon an important variability measure – variance. Since economists come across the problem of interrelation (direct or indirect) between the factor and the indicator when considering some object or phenomenon, the assumption that the variance of real and imaginary parts are not interrelated is rarely met. Thus, the hypothesis saying that the variance of the complex random value should always be real cannot be taken as a basis in econometrics. On the contrary, the variance of economic complex-valued random variable should be presented as a complex characteristic of the variability of a random complex sequence (Svetunkov, 2018).

Then the complex variance of the complex random value can be represented in the following way:

$$D_c(z)=M[z^2]=M[|z|e^{i2\theta}]=M[|z|\cos 2\theta]+iM[|z|\sin 2\theta], \quad (1.5.1)$$

where

$$\theta=arctg\frac{x_i}{x_r}+2\pi k, k=0,1,2,...$$

As will readily be observed, variance (1.4.1) is a special case of variance (1.5.1), namely – when vectorial angle θ between real part and imaginary part of the complex-valued variable is equal to $\theta = \pi k, k = 0,1,2,...$, i.e. real part and imaginary part are not interdependent.

How can the assumed hypothesis about the relationship between the real and imaginary parts, and that the variance of the complex-valued variable should be considered as the complex value, help in solving applied econometric problems? To answer this question let us turn back to the calculation of correlations between complex random values using two methods, which resulted in an impasse if we assume that the variance of the complex-valued variable is a real number.

We shall consider all the characteristics of the complex random value as complex numbers. This is why we shall not resort to their artificial transformation into real numbers of these characteristics by multiplying the complex number by its conjugate. Let us represent the correlation moment of two random complex-valued variables as a complex number:

$$\mu_{XY} = M[x_r y_r] - M[x_i y_i] + i(M[x_i y_r] + M[x_r y_i]) = \mu_{x_r y_i} - \mu_{x_i y_i} + i(\mu_{x_i y_r} + \mu_{x_r y_i}). \quad (1.5.2)$$

If we apply the values of complex variance (1.4.14) and complex correlation moment (1.5.2) to the formula for calculation of pair correlation coefficient (1.4.5), we obtain:

$$r_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y} = \frac{\sum (x_r + ix_i)(y_r + iy_i)}{\sqrt{\sum (x_r + ix_i)^2 \sum (y_r + iy_i)^2}} = \frac{\sum (x_r y_r - x_i y_i) + i \sum (x_r y_i + x_i y_r)}{\sqrt{\sum (x_r + ix_i)^2 \sum (y_r + iy_i)^2}}. \quad (1.5.3)$$

The obtained formula for the calculation of the sample value of the pair correlation coefficient of two random variables (1.4.7) does not coincide with any of the previously derived formulae (1.4.9) and (1.4.13), when all the characteristics were considered to be real ones.

We shall calculate the complex pair correlation coefficient using the second method – as the geometric mean of sample values of the regression coefficients. In order to find this coefficient, let us formulate the criterion of least square method. It is to be recalled that assuming that the variance of the random complex-valued variable is a real value, the least square method means in fact searching for such regression coefficients whereby:

$$M[(\varepsilon_r + i\varepsilon_i)(\varepsilon_r - i\varepsilon_i)] = M[\varepsilon_r^2 + \varepsilon_i^2] \to \min. \quad (1.5.4)$$

where $\varepsilon_r = y_r - \hat{y}_r$, $\varepsilon_i = y_i - \hat{y}_i$ - are approximation errors.

In case of the complex variance of the complex random value the least square method reduces to searching for other coefficients whereby (*Svetunkov, 2012,* p. 83):

$$M[(\varepsilon_r + i\varepsilon_i)^2] = M[\varepsilon_r^2 + i2\varepsilon_r\varepsilon_i - \varepsilon_i^2] \to \min. \quad (1.5.5)$$

One can pay attention here to the interrelation between criteria (1.5.4) and (1.5.5). With this aim in view let us present the complex approximation error in the exponential form:

$$\varepsilon_r + i\varepsilon_i = Re^{i\theta} = \sqrt{\varepsilon_r^2 + \varepsilon_i^2}\, e^{iarctg\frac{\varepsilon_i}{\varepsilon_r}} . \quad (1.5.6)$$

Taking it into account, criterion (1.5.4) takes the form:

$$M[(\varepsilon_r + i\varepsilon_i)(\varepsilon_r - i\varepsilon_i)] = M[R^2] \to \min , \quad (1.5.7)$$

and criterion (1.5.5) is as follows:

$$M[(\varepsilon_r + i\varepsilon_i)^2] = M[R^2 e^{i2\theta}] \to \min , \quad (1.5.8)$$

Therefore, criterion of the least square method (1.5.4) suggested by G.N.Tavares and L.M.Tavares, is a special case of criterion (1.5.5) – when the vectorial angle of the complex approximation error is equal to zero.

Now, using criterion (1.5.5) in relation to the complex regression coefficient of complex number *X* by complex number *Y* denoted as *a*, we will obtain such formula using the least square method with criterion (1.5.5) (*Svetunkov, 2012, p. 103 – 112*):

$$a = \frac{\sum (x_r + ix_i)(y_r + iy_i)}{\sum (x_r + ix_i)(x_r + ix_i)} . \quad (1.5.9)$$

The complex coefficient of proportionality *b* of the inverse regression can also be calculated by using criterion (1.5.5) as:

$$b = \frac{\sum (x_r + ix_i)(y_r + iy_i)}{\sum (y_r + iy_i)(y_r + iy_i)} . \quad (1.5.10)$$

Now when plugging these coefficients in the formula for calculation of the pair correlation coefficient (1.4.10), we obtain:

$$r_{XY} = \sqrt{a_1 b_1} = \frac{\sum (x_r + ix_i)(y_r + iy_i)}{\sqrt{\sum (x_r + ix_i)^2 \sum (y_r + iy_i)^2}} = \frac{\sum (x_r y_r - x_i y_i) + i\sum (x_r y_i + x_i y_r)}{\sqrt{\sum (x_r + ix_i)^2 \sum (y_r + iy_i)^2}} . \quad (1.5.11)$$

As it can be seen, the same formula of the complex pair correlation coefficient (1.5.3) as in the case of its calculation through the complex correlation moment (1.5.1) is obtained. This means that the obtained result is not contradictory. Both pair correlation coefficient (1.5.3) between two random complex-valued variables calculated through variance and correlation moment and the pair correlation coefficient calculated through the geometric mean of linear regression have one and the same form. This in turn means that our hypothesis about the need to use complex variance and other complex characteristics of the complex-valued variables in statistics of the complex-valued random variables, is confirmed.

Let us turn now to the analysis of the properties of complex variance of complex random value (1.4.14), the use of which in econometrics of the complex-valued random variables has just been justified. To illustrate it, let us write complex variance in the arithmetic form:

$$D_c(z) = M[z^2] = M[x_r^2] - M[x_i^2] + i2M[x_r x_i] . \quad (1.5.12)$$

Depending on what form the real and imaginary parts have, complex variance can be a complex, real or imaginary value – the variety of its values correspond to the variety of the properties of the complex-valued random variable. In addition, complex variance can be both positive and negative. Let us consider these options and properties of the complex random value sequence, for which these options of complex variance are valid.

Firstly, let us pay attention to the imaginary part of complex variance (1.5.12):

$$\mathrm{Im}[D_c(z)] = 2M[x_r x_i] . \quad (1.5.13)$$

It has a simple meaning – it is double covariance between the real and imaginary parts of the random complex-valued variable. If there is no correlation between variables, the variable covariance is equal to zero. This means that the imaginary part of the complex variance serves the basis to assume on the presence or absence of correlation between the real and imaginary parts of the random complex-valued variable.

The real part of complex variance of the random complex value is also meaningful for the researcher:

$$\mathrm{Re}[D_c(z)] = M[x_r^2] - M[x_i^2] . \quad (1.5.14)$$

As it can be seen, it characterizes the degree of distinction between the variance of the real part of the random complex-valued variable and the variance of the imaginary part of the given variable. This is why in case when both types of variance are equal to each other, the real part of the complex variance is equal to zero. If the variance of the real part of the complex-valued variable is larger than that of the imaginary part of the complex-valued variable, real part (1.5.12) of the complex variance will be positive. Otherwise, it will be negative.

It is noteworthy that justifying of complex character of the complex random value does not refuse the possibility to use variance in real form – it can be applied as an additional characteristic of the process under research, since real variance characterizes the variability measure of the absolute value of complex variance.