1.9. Confidence limits of a complex random variable

We consider the problems and tasks of constructing econometric models exclusively to the conditions of reversible processes - random and normally distributed. And this means that the researcher is dealing with sample values of random variables by which he judges the general population as a whole. Since sample values are being evaluated, it is necessary to determine how much these sample values can be trusted, that is, to assess how close they are to their true value, namely, to the mathematical expectation.

It is clear that if the researcher is faced with the task of studying a simple stationary process of a random real variable, the one which is represented by some sample from the general population, then, assuming the normal distribution of this variable Y_i, its arithmetic mean should be calculated first:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad , \qquad (1.9.1)$$

and, having calculating the variance of the actual observations` deviations from this mean σ^2, it is possible to determine the interval in which the true value of Y is located:

$$\bar{Y} - t_\alpha \frac{\sigma}{\sqrt{n}} \leq Y \leq \bar{Y} + t_\alpha \frac{\sigma}{\sqrt{n}} \; . \qquad (1.9.2)$$

Here t_α *is* the value of the Student's *t*-statistics.

As can be seen from (2.8.2), the confidence limits for real one-dimensional variables are a segment on the numerical axis, within which random variables can be located with a given probability.

If, instead of the real case, we consider a complex-valued variable, then the course of reasoning should not, at first glance, be violated – the arithmetic means for a complex random variable are calculated (which is identical to the arithmetic means calculation for the real and imaginary parts separately), variances are determined for them, after which, using the standard approach (1.9.2), the confidence limits are determined. Then the confidence limits of the two components values of a random complex variable should be defined as follows: (1.9.3)

$$(\bar{y}_r + i\bar{y}_i) - \frac{t_\alpha}{\sqrt{n}}(\sigma_{y_r} + i\sigma_{y_i}) \leq y_r + iy_i \leq \bar{y}_r + \frac{t_\alpha}{\sqrt{n}}(\sigma_{y_r} + i\sigma_{y_i}). \qquad (1.9.3)$$

But the meaning of this method of determining the confidence limits of a complex random variable will be revealed if it is written as a system of two conditions for changing the confidence limits separately for the real and imaginary parts:

$$\begin{cases} \bar{y}_r - t_\alpha \dfrac{\sigma_{y_r}}{\sqrt{n}} \le y_r \le \bar{y}_r + t_\alpha \dfrac{\sigma_{y_r}}{\sqrt{n}}, \\ \bar{y}_i - t_\alpha \dfrac{\sigma_{y_i}}{\sqrt{n}} \le y_i \le \bar{y}_i + t_\alpha \dfrac{\sigma_{y_i}}{\sqrt{n}}. \end{cases} \quad (1.9.4)$$

It clearly follows from this that on the complex plane of the random complex variable, the confidence region will be a rectangle outlined by the sides determined by the confidence limits (1.9.4), and the sides of this rectangle being parallel to the axes of the complex plane. The center of this rectangle, and hence the confidence region in the form of a rectangle, will be a point on the complex plane determined by the coordinates of the arithmetic mean of the complex random variable $(\bar{y}_r, \bar{y}_i)$.

As it has been shown in paragraph 1.3 of this monograph, the confidence region of a complex random variable should be a scattering cloud of acceptable values, and not a rectangle. In addition, this cloud should have the shape of an ellipse, the axes of which are parallel to the axes of the complex plane only if the real and imaginary parts of the complex variable do not depend on each other. And in the case of their dependence on each other, and this is the case that we are considering, the axes of the scattering ellipse will not be parallel to the axes of the complex plane (Fig. 3).

Thus, the procedure for finding confidence limits using the rule (1.9.3), which at first sight seems correct, turns out to be a very rough approximation to reality and it can be used only for the purpose of finding very approximate boundaries of the confidence region.

Therefore, the standard approach, which seems to be so obvious, turns out to be wrong. For scientific and practical research, it is necessary to use a confidence region, which represents an ellipse, inside of which there are those points that enter the confidence region, and outside of which there are points that go beyond the confidence region.

Let us use the scattering ellipse equation. With regard to our problem, the confidence region must be inside this ellipse, that is, the following condition must be met:

$$\frac{(y_r - m_{y_r})^2}{\sigma_{y_r}^2} - 2\frac{r_{y_r y_i}(y_r - m_{y_r})(y_i - m_{y_i})}{\sigma_{y_r}\sigma_{y_i}} + \frac{(y_i - m_{y_i})^2}{\sigma_{y_i}^2} \le s_{\alpha,n} \,. \quad (1.9.5)$$

Here $s_{\alpha,n}$ is some number defining the limits of the confidence region. This number depends on the confidence probability level α and the number of degrees of freedom n

For sample values, when instead of mathematical expectations we know their estimate - the arithmetic means and sample variance values, the equation of the confidence region ellipse of the complex random variable will look like this:

$$\frac{(y_r - \bar{y}_r)^2}{\sigma_{y_r}^2} - 2\frac{r_{y_r y_i}(y_r - \bar{y}_r)(y_i - \bar{y}_i)}{\sigma_{y_r}\sigma_{y_i}} + \frac{(y_i - \bar{y}_i)^2}{\sigma_{y_i}^2} \le s_{\alpha,n}$$

(1.9.6)

We failed to find an analytical relationship between $s_{\alpha,n}$ and t_α, and this will be the task of our further research study. But, since modern computing technology allows to perform numerous simulation experiments as well as computer tests, this approach makes it possible to find a tabular relationship between them. Table 1.4 shows the recommended values of $s_{\alpha,n}$ depending on the of confidence probability level α and the number of degrees of freedom n, which were obtained during such machine experiments.

Table 1.4.

Critical points of distribution $s_{\alpha,n}$

Number of degrees of freedom n	Significance level α			
	0,10	0,05	0,02	0,01
1	19,908	80,645	506,256	2028,846
2	2,842	6,163	16,194	32,802
3	1,381	2,528	5,153	8,526
4	0,907	1,546	2,812	4,233
5	0,673	1,100	1,536	2,707
6	0,538	0,857	1,409	1,966
7	0,477	0,696	1,125	1,531
8	0,384	0,593	0,935	1,254
9	0,335	0,511	0,795	1,056
10	0,298	0,452	0,692	0,914
11	0,270	0,403	0,617	0,806
12	0,244	0,366	0,552	0,716
13	0,224	0,333	0,502	0,647
14	0,207	0,305	0,458	0,592
15	0,191	0,284	0,423	0,544
16	0,180	0,264	0,392	0,502

17	0,168	0,247	0,367	0,467
18	0,158	0,232	0,342	0,437
19	0,150	0,218	0,323	0,409
20	0,143	0,208	0,305	0,387
21	0,134	0,197	0,289	0,364
22	0,129	0,186	0,274	0,346
23	0,122	0,177	0,260	0,329
24	0,117	0,170	0,248	0,314
25	0,112	0,163	0,238	0,299
26	0,108	0,157	0,228	0,286
27	0,104	0,150	0,218	0,274
28	0,100	0,145	0,209	0,263
29	0,096	0,140	0,202	0,254
30	0,095	0,138	0,199	0,250
40	0,094	0,136	0,195	0,243
60	0,093	0,133	0,190	0,236
120	0,092	0,131	0,186	0,229

Let us show how to use this table and the conditions (1.9.6) on a specific example.

We have at our disposal data on the results of daily quotations on the world commodity exchanges of two commodities - *Brent* crude oil and natural gas from January 4, 2010 to August 9, 2013. Since these two products reflect the situation on the world market of organic fuel, they can be represented as one random complex variable:

$$Z_t = y_{rt} + iy_{it}. \quad (1.9.7)$$

where y_{rt} is the price for a barrel of *Brent* crude oil, and y_{it} is the price of a cubic meter of natural gas.

Since the dimensions and scale of these variables are different, they must be reduced to the same dimension and to the same scale. The easiest way to do this is to divide each value of value series of the price for a barrel of oil by this indicator of the first observation value dated January the 4th, 2010, and each value of value series of the gas price to divide by the first value of the gas price dated January the 4th, 2010. Then dimensionless quantities comparable both to each other and in scale will be obtained.

For the obtained series of more than 900 observations, the arithmetic mean has been found, which is equal to:

$$\bar{y}_{rt} + i\bar{y}_{it} = 1,259 + i0,643. \quad (1.9.8)$$

For the same series, the variances and their standard deviation (SD) have been calculated, which are equal to: $\sigma_r=0,00616$ and $\sigma_i=0,00427$. The pair correlation coefficient between the real and imaginary parts of the complex random variable has also been calculated, which turned out to be equal to *r=-0,57001*. This, by the way, once again reiterates our conviction that such economic indicators, combined into one complex variable, in no case can be considered as independent of each other and the variance of such a complex random variable should be considered as a complex value.

Now it is possible to substitute these values into the condition (1.9.6) and to determine the region of confidence limits for the values of this series of complex random variable:

$$26334,55(y_r-1,259)^2+43293,6(y_r-1,259)(y_i-0,643)+54765,12(y_i-0,643)^2 \le s_{\alpha,n} . \quad (1.9.9)$$

Let us find the answer to such a question - whether the number (1.40+i0.80) falls into the confidence parameter region with a probability of 0.95?

To answer this question, the specified values of a complex random variable should be substituted in (1.9.9) and the value of the left side of the inequality should be calculated. Let us do this. As a result of calculations, the number s = 2831.366 was obtained. It is significantly higher than the critical value, the one which, as can be seen from Table 1.4, for more than 900 observations and the significance level $\alpha=0,05$ is equal to $s_{\alpha,n}=0,131$. Therefore, the specified number *(1,40+i0,80*) goes beyond the range of confidence parameter.

Now let us find the answer to another question – whether the number *(1,26+i0,64)* falls into the range of confidence parameters with the same probability of 0.95?

Substituting the values of this complex number in (1.9.6) and calculating the value of *s* (the left side of the inequality (1.9.6)), we obtain *s=0,018*. It can be seen that the calculated value of the left side of the inequality is less than the critical one and the inequality is satisfied: $s < s_{\alpha,n}=0,131$. Therefore, the number in question *(1,26+i0,64)* is inside the confidence region.

Thus, the proposed procedure for determining confidence limits for a complex random variable can be used for scientific and practical research in the field of complex-valued econometrics.

It can also be used to estimate the confidence limits of other sample complex variables, for example, complex coefficients of regression models.

We will not consider the issue of estimating the confidence limits for the complex coefficient of pair correlation here, realizing that the approach for estimating the confidence limits of a complex random variable, considered in these paragraphs, is universal, and it can also help to estimate the confidence limits of sample values of the complex coefficient of pair correlation.