Time Series Analysis and Forecasting with Complex Dynamic Models

Sergey Svetunkov and Ivan Svetunkov

2023-04-08

Contents

Introduction **5**

1 Introduction to theory of complex variables **9**
1.1 Theory of complex numbers 9
1.2 Complex random variable 21
1.3 Parametric Distributions of c.r.v. 27

2 Simple Complex Linear Regression **33**
2.1 Model formulation 33
2.2 Estimation 35
2.3 Comparing different estimators for SCLR 44

3 Correlation analysis of complex random variables **51**
3.1 Visualisation of relations 51
3.2 Types of correlation coefficients 54
3.3 Conjugate correlation 56
3.4 Direct correlation 58
3.5 Pearson's correlation 59
3.6 Correlation matrix 59

4 Multiple Complex Linear Regression **61**
4.1 Model formulation 61
4.2 Estimation 62
4.3 Inference 65
4.4 Diagnostics 67
4.5 Forecasting 67
4.6 Examples of application (Production functions) 67

Introduction

The theory of complex variables functions is actively used in a variety of disciplines, including modern physics and engineering sciences. It is relatively easy to describe the complex phenomena that are studied in these areas of science using the models and methods of this branch of mathematics. Social sciences, being much more complicated due to the unpredictability of human behaviour, tend to use simpler instruments for modelling complex processes (references). For example, there are only few scientific publications in which methods and models of the theory of complex variables functions are used in economics, and they typically use that instrument for diagnostics or statistical tests (references) rather than proper modelling.

In 2012, Springer published the monograph "Complex-valued Modeling in Economics and Finance" (Svetunkov, 2012), which presented the theory and methodology of modelling using complex variables in economics. Svetunkov (2012) have summarised the main principles of using complex variables in economics and discussed how to estimate some of those models. While, this was the first monograph that discussed the topic, the first paper in this direction was Ben Tamari (1997). The author first introduced Wealth as a complex variable consisting of Output (real part) and Money (imaginary part) and showed how, even with such an elementary representation, interesting new results can be obtained in the area of economics. Unfortunately, this work has gone unnoticed in scientific world. We became aware of this work only in 2016, when Ben Tamari had kindly sent Sergey Svetunkov his paper.

Since 2012, there has been some development in the area of modelling using complex variables, notably a paper by Svetunkov & Kourentzes (2022) on Complex Exponential Smoothing and Kourentez et al. (2019) on an error measure based on the idea of complex numbers. Up until now, the modelling with complex variables has been mainly picked up by academics working in the areas of forecasting and engineering. The latter group has been using complex autoregressions for a couple of decades, modelling and predicting signals. The former group has only started using the principles described in Svetunkov (2012). In other disciplines, complex variables are not used directly for model building. The probable reason for this is the lack of the communication between the disciplines and the inherited inertia of academia.

In a try to speed up the adoption of the new instrument, we have written this monograph, summarising the research that has been done in the area of dynamic models since 2012. We should clarify that the term "dynamic model" used in this monograph refers to models that have a structure that changes over time. The classical example of such a model is ARIMA (AutoRegression Integrated with Moving Average by Box & Jenkins, 1976), which allows producing forecasts for a variety of processes based on the existing historical time series. Another example is a multivariate version of ARMA, Vector ARMA or VARMA, which allows modelling the dynamics of several related processes simultaneously. For instance, in business and economics, these models are used for prices, demand forecasting and for capturing complex interactions between macroeconomic indicators over time. However, all autoregression models that exist today use real variables. In this monograph we will discuss their complex-valued counterparts, namely complex ARIMA and complex VARMA, and study their properties, showing how to identify their orders and how to estimate these models in practice. But before rushing into the discussion, we will explain the basics of random complex variables and complex-valued statistics, which has been developed in signal processing literature, but has not been used to its full potential.

Chapter 1 discusses the theory of random complex variables, conventional and complex-valued statistics, the complex least squares method, maximum likelihood and complex autocorrelation and partial autocorrelation functions. In Chapter 2, we will move to the discussion of simple dynamic model – complex AR, starting from its properties and slowly moving to its identification and estimation and then to forecasting. After that, in Chapter 3, we will move towards complex MA, again discussing its properties and how one can apply it in practice. The two parts will be united in a cARIMA model in Chapter 4, where we will also discuss seasonal counterpart of the model. Finally, we will move to vector models in Chapter 5, introducing cVAR, cVMA and cVARMA, showing their advantages in comparison with the conventional real valued models. All of this will be supported by examples in R, which will rely heavily on the package "complex", developed especially for this monograph. This means that anyone can then use the proposed models for purposes of time series analysis and forecasting. This work was partially supported by the Russian Foundation for Basic Research (grant No. 19-010-00610\19 "Theory, methods and methodologies of forecasting economic development by autoregression models of complex variables"). Thanks to this support, it became possible to carry out this research in general, forming a new scientific direction in statistical modelling and short-term forecasting.

The authors are grateful to those young scientists of St. Petersburg Polytechnic University of Peter the Great, who participated in this study, patiently testing the author's hypotheses. They helped forming the theory behind the models discussed in this monograph. We are pleased to mention their names: Evgeny Goltsev, Nikolai Pitukhin, Victoria Matskevich, Yulia Selivanova, Galina Siruk and Nazira Shaikhleeva.

this monograph relies heavily on the `complex` package for R. Many of the

examples in R will use functions from this package, so make sure to install it before running any R code (Svetunkov, 2022):

```
install.packages("complex")
```

Chapter 1

Introduction to theory of complex variables

1.1 Theory of complex numbers

This section sets out the basic concepts of the Theory of Complex Variables Functions (TCVF) with some historical overlook of the idea of complex numbers. Hopefully, this will help in understanding the main idea that we will use extensively in the following chapters.

The theory of complex numbers started in 1572 in the small Italian town of Kura. It was in this place and precisely in this year that the manuscript of an Italian mathematician Rafael Bombelli was published. The book was called Algebra. Rafael Bombelli showed in it how to solve the following cubic equation:

$$x^3 = 15x + 4. \tag{1.1}$$

The root of cubic equations at that time was calculated using the formula of Scipione del Ferro. With regards to the task at hand, finding the root should have been carried out as follows:

$$\begin{aligned} x = &\sqrt[3]{\frac{4}{2} + \sqrt{\left(\frac{4}{2}\right)^2 - \left(\frac{15}{3}\right)^3}} + \sqrt[3]{\frac{4}{2} - \sqrt{\left(\frac{4}{2}\right)^2 - \left(\frac{15}{3}\right)^3}} = \\ &\sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}. \end{aligned} \tag{1.2}$$

As can be seen from equation (1.2), there are several square roots, and it follows directly from the right hand side of this equation that to calculate the roots of equation (1.1), we need to extract the square root from the negative number, -121. This means that it will not be possible to find a solution to this equation in the domain of real numbers.

This does not mean that equation (1.1) does not have a solution. If you represent the problem graphically on a plane then it is easy to see that there is one. But it is impossible to find it arithmetically using the Scipione del Ferro formula, because there is a negative number in the expression on the right-hand side of (1.2), and back then, mathematicians were sure that the square root of a negative number does not exist.

The mathematicians of that time struggled to solve this problem, and Rafael Bombelli suggested to ignore the negative sign in the expression (Bombelli, 12). After all, the number -121 can be represented as a product of two numbers: -1 and 121. Then the square root of this number can be written as $\sqrt{-1} \times 11$. So, formula (1.2) becomes:

$$x = \sqrt[3]{2 + 11\sqrt{-1}} + \sqrt[3]{2 - 11\sqrt{-1}}. \tag{1.3}$$

Now it is possible to get a solution of the problem, and the root of the cubic equation is $x = 4$.

For a long time, Bombelli's approach was considered by scientists as a convenient mathematical trick. And the square root of minus one was called "magic unit", "vanishing unit", etc. Finally, the term "imaginary unit" was established by René Descartes in 1637 and any number multiplied by the imaginary unit would be called "imaginary number". Merging a real number with an imaginary number gives another number, which can be considered as a more general to the two, which is called "complex". It can be written as:

$$\underline{z} = x + iy, \tag{1.4}$$

where x is the real part, iy is the imaginary part of a complex number, x and y are real numbers and i is the imaginary unit that satisfies the equality:

$$i^2 = -1. \tag{1.5}$$

We use the symbol to denote a complex number.

With complex numbers, one can perform almost all the same operations as with the real ones. However, taking into account the properties of the imaginary unit, these operations can lead to results that are not common in the domain of real numbers.

The main problem that analysts face when trying to understand a complex number is the complexity of the interpretation of the imaginary part. The main question in this case can be formulated as: Where do imaginary numbers appear in practice? And what is the practical meaning of an imaginary unit? We have heard questions like that many times, and while they are valid, we think that they do not have appropriate answers and divert the analysts from what the modelling with complex variables brings.

We argue that the imaginary unit does not have any practical meaning: neither in economics, nor in sociology, nor in engineering, nor in physics. The imaginary

unit is a mathematical rule, and that is all. It is a mathematical abstract that has useful properties. A number $\sqrt{2}$ or $\ln 3$ do not have practical meaning either, but they are useful for modelling purposes and solving applied problems in a variety of disciplines. Similarly, an imaginary unit is convenient for solving a whole class of problems in different areas of application. With the help of the rules specified by conditions (1.4) and (1.5), it is possible to use new mathematical operations, obtain new mathematical results, and create new mathematical models.

Imaginary and complex numbers are mathematical instruments that can help in describing a real-life phenomenon. If a researcher decides to use complex variables for real processes modelling, they will need to predefine the rules according to which one component of a complex process is attributed to the real part, and another one is attributed to the imaginary part of the complex variable. These rules might seem arbitrary, because there is no "true" recipe, according to which the assignment to complex variable parts should be done. The main motivation for them is the efficiency of modelling (such as goodness of fit or forecasting accuracy) and nothing else.

In order to better understand how the modelling with complex variables can be done, we need to better understand how complex variables can be represented and interpreted. A real number represents visually a certain segment on the numerical axis, which has a zero point and a multitude of values from minus infinity to infinity. Any real number is characterized by the distance from zero to the value of that number. Negative numbers are located to the left of zero, while the positive ones are located to the right of it.

A complex number consists of two parts, which can be visualised on the plane with two perpendicular axes, where the real numbers are marked on the horizontal line, and the imaginary ones are on the vertical one. This is shown visually in Figure 1.1 with a complex number $\underline{z} = x + iy$.

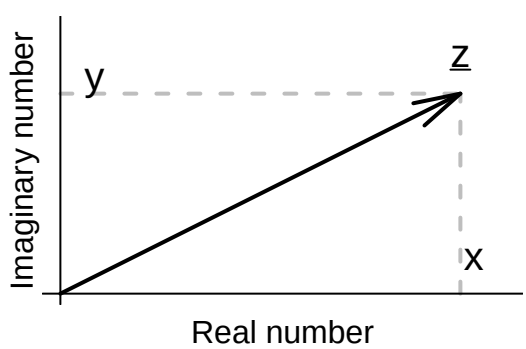

Figure 1.1: Visual presentation of a complex number on the complex plane.

Any point lying on the complex plane in Figure 1.1 characterises a complex number, even if that point lies on the axis of real numbers. In that specific case, we would be talking about a complex number with a zero imaginary part,

i.e. $\underline{z} = x + i0$.

Given that complex numbers can be represented on a plane (in so called Cartesian coordinates), the number (1.4) can be represented in a form of a vector that starts in the origin of coordinates and ends at the point (x, y). Then any complex number can be represented in form of polar coordinates using the magnitude of the vector and its polar angle:

$$\underline{z} = x + iy = r(\cos\phi + i\sin\phi), \tag{1.6}$$

where $r = |\underline{z}|$ is the magnitude, which is calculated as a Euclidean distance from the origin to the point (x, y) on the plane:

$$r = \sqrt{x^2 + y^2}, \tag{1.7}$$

and $\phi = \mathrm{Arg}(\underline{z})$ is the angle, which equals to:

$$\phi = \arctan\frac{y}{x} + 2\pi l, \tag{1.8}$$

where l is an integer number. While depending on a task, l can be set to a positive, a negative number, or a zero, for convenience, we will restrict it to $l = 0$, because all the other values will not be useful for the inference in following chapters. The polar angle is sometimes referred to as the argument of a complex number.

Using the magnitude and the polar angle, we can also represent any complex number in the exponential form, which was first proposed in 1748 by Euler, in his book "Introduction to the Infinitesimal Analysis" **(Euler, 1961, pp. 118-119)**, which:

$$\underline{z} = re^{i\phi}, \tag{1.9}$$

where e is the Euler's constant. Equation (1.9) is nowadays also called "Euler's form". Comparing @ref{eq:complexNumberTrigonometric} with @ref{eq:complexNumberExponential}, we can conclude that:

$$e^{i\phi} = \cos\phi + i\sin\phi, \tag{1.10}$$

which gives the connection between the linear, trigonometric and exponential forms of a complex number. In the exponential form (1.9), a positive real number has $\phi = 0$, while a negative one has $\phi = \pi$, and all imaginary numbers have an angle dividable by $\frac{\pi}{2}$. So, for example:

$$\begin{aligned}
41 =& 41e^{i0} \\
i41 =& 41e^{i\frac{\pi}{2}} \\
-41 =& 41e^{i\pi} \\
-i41 =& 41e^{i\frac{3\pi}{2}}
\end{aligned}$$

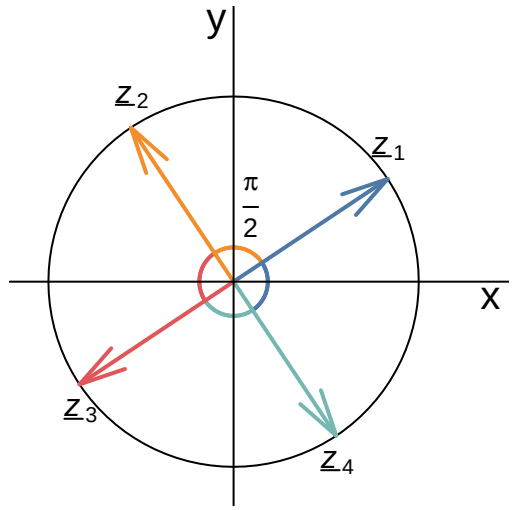

Figure 1.2: Multiplication of a complex number by i implies the rotation of it by $\frac{\pi}{2}$.

In fact, multiplication of any complex number by the imaginary unit implies the rotation of complex vector by $\frac{\pi}{2}$, as shown in Figure 1.2, where a complex number $\underline{z}_1 = x_1 + iy_1$ becomes $\underline{z}_2 = \underline{z}_1 \times i = x_2 + iy_2$ etc.

In a way, any mathematical operation with a complex number implies a change in magnitude and angle of the number, while the multiplication and division by a number with non-zero imaginary part will always lead to the rotation of the vector. Operations of addition and subtraction are simpler in the linear form:

$$\underline{z}_3 = \underline{z}_1 + \underline{z}_2 = x_1 + x_2 + i(y_1 + y_2),$$

while operations of multiplication and division are easier to do in either exponential or trigonometric forms of complex numbers:

$$\underline{z}_3 = \underline{z}_1 \times \underline{z}_2 = r_1 r_2 e^{i\phi_1 + \phi_2}$$

or:

$$\underline{z}_3 = r_1 r_2 \left(\cos(\phi_1 + \phi_2) + i \sin(\phi_1 + \phi_2)\right).$$

Furthermore, given that any complex number can be represented in the exponential form (1.9), it can also be visualised on a polar coordinates plane, where the magnitude is marked on the x-axis, and the polar angle is marked on the y-axis. This is shown visually in Figure 1.3.

The usefulness of the polar coordinates presentation becomes apparent when a set of complex numbers is considered, because then in some cases it becomes possible to see some relations that are not obvious on the Cartesian plane. Figure

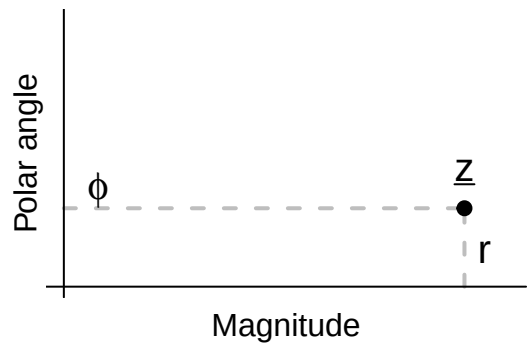

Figure 1.3: Visual presentation of a complex number in the polar coordinates.

1.4 shows an example of a set of complex random numbers, for which the real and imaginary parts do not seem to have any obvious linear relation, but the magnitude and the angle have a negative relation.

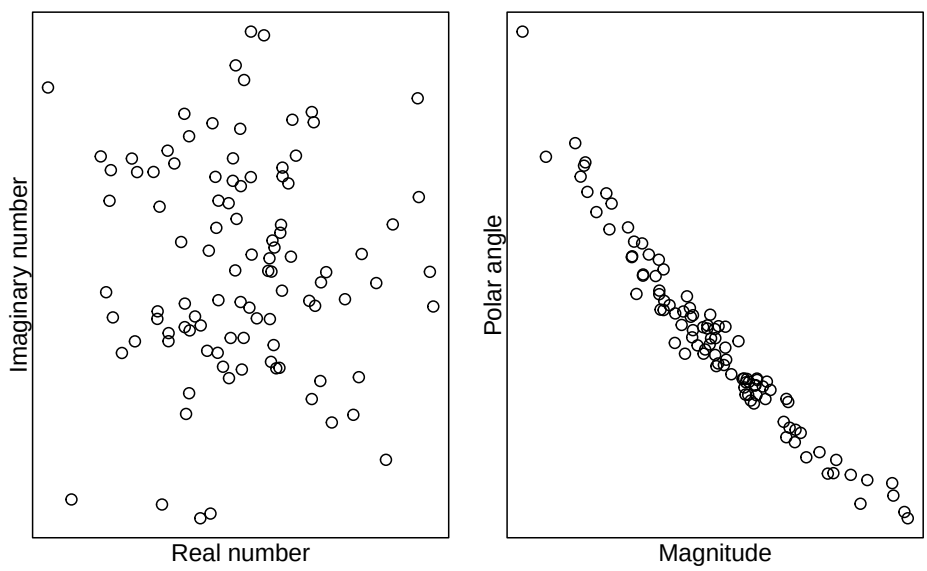

Figure 1.4: Visualisation of a set of complex numbers on Cartesian and on polar coordinates planes.

In this case, modelling can be done in the exponential form of complex numbers, which might allow capturing complicated non-linear relations between variables.

When it comes to comparing complex numbers, mathematically we will say that two complex numbers $\underline{z}_1 = x_1 + iy_1$ and $\underline{z}_2 = x_2 + iy_2$ are equal to each other if

and only if their real and imaginary parts are equal:

$$\underline{z}_1 = \underline{z}_2 \iff \begin{cases} x_1 = x_2 \\ y_1 = y_2 \end{cases},$$

which is equivalent to saying that their magnitudes and polar angles are equal:

$$\underline{z}_1 = \underline{z}_2 \iff \begin{cases} |\underline{z}_1| = |\underline{z}_2| \\ \mathrm{Arg}(\underline{z}_1) = \mathrm{Arg}(\underline{z}_2) \end{cases}.$$

Unfortunately, given that complex numbers are two dimensional, it is not possible to say in general whether one number is greater or less than the other. However, we could use magnitude to compare complex numbers, to say which one lies further away from the origin than the other. In that case, the number with a larger magnitude could be considered as a greater than the one with the smaller magnitude. While we could compare the polar angles as well, they only give us information about rotation of a complex number and thus do not give any meaningful information about the comparison of numbers. So, two complex numbers in Figure 1.5 would have the same magnitude, but will have different angles, and it is not possible to say whether one number is greater than the other in this case. In order to conclude that, we would need to devise additional criteria for comparison of numbers (e.g. positive numbers are "better" than the negative ones, thus $\phi_1 = \pi$ and $\phi_1 = -\pi$ is "worse" than $\phi_2 = 0$).

Continuing the discussion of different forms of a complex number, its magnitude can be represented in exponential form:

$$r = e^{\ln(r)},$$

where ln is the natural logarithm. This can then be used to rewrite the exponential form of a complex number as:

$$\underline{z} = e^{\ln r + i\phi}, \tag{1.11}$$

And this form allows calculating logarithms of a complex number:

$$\ln \underline{z} = \ln\left(e^{\ln r + i\phi}\right) = \ln r + i\phi.$$

This operation will become useful when we discuss transformations of complex-valued functions, but it also shows that a logarithm of a negative real number is a complex number, because in that case $\phi = \pi$. This demonstrates that the field of complex numbers is complete: any mathematical operation of a complex number will give another complex number.

One of the important definitions in theory of complex numbers is the conjugate complex number. It is the number, for which the imaginary part has the sign opposite to the one of the original variable. For example, the conjugate of the

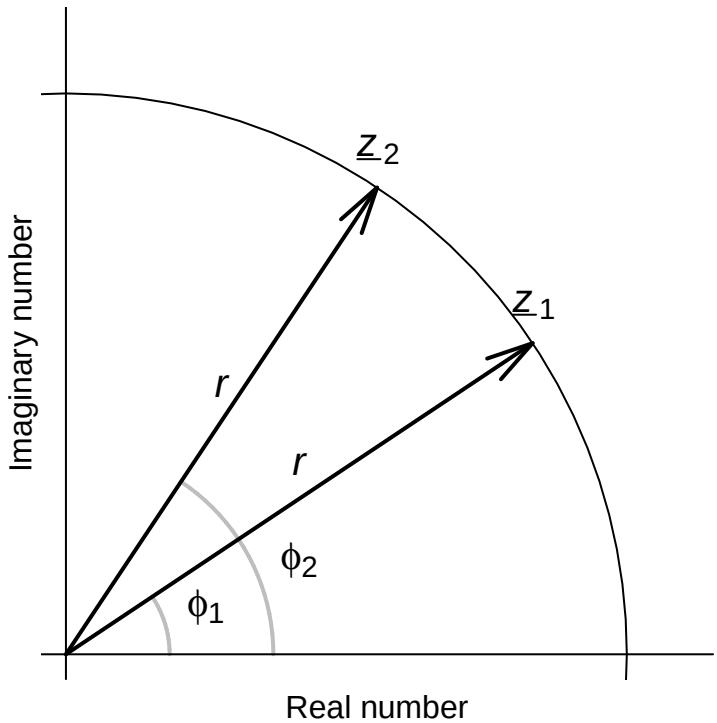

Figure 1.5: Comparison of two complex numbers $\underline{z}_1$ and $\underline{z}_2$ that have the same magnitude r, but different angles ϕ_1 and ϕ_2.

complex variable $\underline{z} = x + iy$ is $\underline{\tilde{z}} = x - iy$. These numbers are useful because the multiplication of a complex number by its conjugate gives a real number:

$$\underline{z} \times \underline{\tilde{z}} = (x + iy)(x - iy) = x^2 + y^2, \tag{1.12}$$

which is the square of the magnitude (1.7) of both complex variables $\underline{z}$ and $\underline{\tilde{z}}$. Conjugation is distributive over addition, subtraction, multiplication and division, meaning that:

$$\begin{aligned} \widetilde{x + y} &= \tilde{x} + \tilde{y} \\ \widetilde{x - y} &= \tilde{x} - \tilde{y} \\ \widetilde{x \times y} &= \tilde{x} \times \tilde{y} \\ \widetilde{\left(\frac{x}{y}\right)} &= \frac{\tilde{x}}{\tilde{y}}. \end{aligned} \tag{1.13}$$

Finally, given that any complex number can be represented visually in a form of a vector, we can write it mathematically either like vector:

$$\boldsymbol{z} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{1.14}$$

or like a matrix:

$$\underset{\sim}{\boldsymbol{z}} = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}, \tag{1.15}$$

where the symbold $\sim$ below the letter denotes a matrix presentation of a complex number. In the matrix notation, any real number can be represented as a diagonal matrix, while any imaginary one is a matrix with zero diagonal:

$$\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} \\ \begin{pmatrix} 0 & -y \\ y & 0 \end{pmatrix}.$$

All the mathematical operations done with the vector and matrix representations would correspond to the ones for the conventional complex numbers. Furthermore, the multiplication by the transpose of the original object in both of these cases is equivalent to the multiplication by the conjugate value (1.12). For the vector:

$$\boldsymbol{z}\boldsymbol{z}' \begin{pmatrix} x & y \end{pmatrix} \times \begin{pmatrix} x \\ y \end{pmatrix} = \left(x^2 + y^2\right) \tag{1.16}$$

and for the matrix:

$$\underset{\sim}{\boldsymbol{z}}\underset{\sim}{\boldsymbol{z}}' = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \times \begin{pmatrix} x & y \\ -y & x \end{pmatrix} = \begin{pmatrix} x^2 + y^2 & 0 \\ 0 & x^2 + y^2 \end{pmatrix}. \tag{1.17}$$

These transpositions will be called in this monograph "conjugate transpositions". The vector and matrix representations become especially useful when we work

with complex variables to construct complex models. Finally, in case of matrix form, it is possible to multiply complex variable by itself without the conjugation, which results in:

$$\underset{\sim}{\boldsymbol{z}}\underset{\sim}{\boldsymbol{z}}^{\top} = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \times \begin{pmatrix} x & -y \\ y & x \end{pmatrix} = \begin{pmatrix} x^2 - y^2 & -2xy \\ 2xy & x^2 - y^2 \end{pmatrix}. \tag{1.18}$$

In this monograph, we will call the multiplications (1.17) and (1.18) "conjugate" and "direct" respectively.

1.1.1 Complex variables

Similarly to any other variable, complex variable is a place holder for values, with the difference that it represents two parts of a number instead of one. Given the properties of complex numbers discussed in the previous section, any complex variable can be represented in linear, trigonometric and exponential forms. But in addition, it can also be represented in the form of a set of equations, i.e. if $y_r + iy_i = 3 + 2i$ then:

$$\begin{aligned} y_r =& 3 \\ y_i =& 2 \end{aligned}, \tag{1.19}$$

where the subscript r or i refers to respectively the real or imaginary parts of the complex number. In fact, any function of a complex variable can be represented as system of two equations, so that for a function

$$y_r + iy_i = f(x_r + ix_i) \tag{1.20}$$

we have:

$$\begin{aligned} y_r =& \mathcal{R}(f(x_r + ix_i)) \\ y_i =& \mathcal{I}(f(x_r + ix_i)) \end{aligned}, \tag{1.21}$$

where $\mathcal{R}()$ and $\mathcal{I}()$ are respectively the real and the imaginary parts of a resulting complex variable. For example, a simple linear function:

$$y_r + iy_i = a_{0,r} + ia_{0,i} + (a_{1,r} + ia_{1,i})(x_{1,r} + ix_{1,i}),$$

which can be written after opening the brackets and regrouping elements as:

$$y_r + iy_i = a_{0,r} + a_{1,r}x_{1,r} - a_{1,i}x_{1,i} + i\left(a_{0,i} + a_{1,i}x_{1,r} + a_{1,r}x_{1,i}\right)$$

and is equivalent to the system of the following two equations:

$$\begin{aligned} y_r =& a_{0,r} + a_{1,r}x_{1,r} - a_{1,i}x_{1,i} \\ y_i =& a_{0,i} + a_{1,i}x_{1,r} + a_{1,r}x_{1,i} \end{aligned}.$$

Given the vector and matrix representation of complex variables, the same linear function can be written as:

$$\begin{pmatrix} y_r \\ y_i \end{pmatrix} = \begin{pmatrix} a_{0,r} \\ a_{0,i} \end{pmatrix} + \begin{pmatrix} a_{1,r} & -a_{1,i} \\ a_{1,i} & a_{1,r} \end{pmatrix} \begin{pmatrix} x_{1,r} \\ x_{1,i} \end{pmatrix},$$

which is convenient in some situations. Presenting the system of equations (1.21) in the linear form (1.20) is often convenient. In fact, any function of complex variables can be represented as a system of two functions. Even if the real or the imaginary part of the output complex variable $y_r + iy_i$ is equal to zero, we can still use a system of two equations, where one of the equations becomes just a constraint for the function. For example, if we assume that $y_i = 0$, then the system (1.21) becomes:

$$\begin{array}{l} y_r = \mathcal{R}(f(x_r + ix_i)) \\ \mathcal{I}(f(x_r + ix_i)) = 0 \end{array}.$$

This property becomes especially useful if one needs to construct a model with linear relations between the input variables (this is discussed in Chapter ??? of **?**).

Given the discussion about the vector and matrix forms of complex numbers, the system of equations (1.21) can be written as:

$$\begin{pmatrix} y_r \\ y_i \end{pmatrix} = \begin{pmatrix} \mathcal{R}(f(x_r + ix_i)) \\ \mathcal{I}(f(x_r + ix_i)) \end{pmatrix}. \tag{1.22}$$

or:

$$\begin{pmatrix} \mathcal{R}(f(x_r + ix_i)) & -\mathcal{I}(f(x_r + ix_i)) \\ \mathcal{I}(f(x_r + ix_i)) & \mathcal{R}(f(x_r + ix_i)) \end{pmatrix}. \tag{1.23}$$

The representations (1.20), (1.21), (1.22) and (1.23) are equivalent and can be used interchangeably depending on the task at hand.

Finally, in terms of definitions, the function (1.20) is often referred to as a "complex-valued" function. It maps a certain values of the input variable $x_r + ix_i$ with a specific value of the output variable $y_r + iy_i$. If there is only a one-to-one relation between the input and output variables then such function is called "univalent". When several different values of the input variable leads to one and the same value of the output variable then the function is called "multivalent". An example of the former is the linear function:

$$y_r + iy_i = (a_{1,r} + ia_{1,i})(x_{1,r} + ix_{1,i}),$$

while for the latter, an example is

$$y_r + iy_i = (a_{1,r} + ia_{1,i})(x_{1,r} + ix_{1,i})^2,$$

where two different input complex numbers (for example $0+i$ and $0-i$) correspond to the same output number $(-1 + i0)$. While there are examples of functions like that in the real domain, in the complex one, they appear even more often due to the rotation property discussed in the previous Subsection.

1.1.2 Vectors of complex variables

Last but not least, for what follows, we introduce vectors and matrices of complex variables. There are different ways how one can represent a vector of complex

variables. The simplest and most direct one is:

$$\underline{\mathbf{x}} = \begin{pmatrix} x_{r,1} + ix_{i,1} \\ x_{r,2} + ix_{i,2} \\ x_{r,3} + ix_{i,3} \end{pmatrix}, \tag{1.24}$$

where the index in subscript (1, 2, 3) represents the observed first, second and third values of the complex variable. Alternatively, given that any complex variable can be represented as a vector (1.14), the same complex vector can be written as a matrix:

$$\mathbf{x} = \begin{pmatrix} x_{r,1} & x_{i,1} \\ x_{r,2} & x_{i,2} \\ x_{r,3} & x_{i,3} \end{pmatrix}, \tag{1.25}$$

Finally, given the connection of complex numbers and matrices (1.15), the same complex vector x can be written as a matrix:

$$\underset{\sim}{\mathbf{x}} = \begin{pmatrix} x_{r,1} & -x_{i,1} \\ x_{i,1} & x_{r,1} \\ x_{r,2} & -x_{i,2} \\ x_{i,2} & x_{r,2} \\ x_{r,3} & -x_{i,3} \\ x_{i,3} & x_{r,3} \end{pmatrix}. \tag{1.26}$$

We use the symbol $\sim$ below a character to denote the complex variable in the matrix form (1.26). Each of the forms (1.24), (1.25) and (1.26) might be useful in different circumstances, their usage should be dictated by the modelling purpose.

Finally, note that the transposition of these three forms will imply different things. As discussed earlier, in vector and matrix forms transposition of a complex number implies the conjugation. However, transposing (1.24) will not have the same effect - it will only switch rows and columns of the complex vector:

$$\underline{\mathbf{x}}^\top = \begin{pmatrix} x_{r,1} + ix_{i,1} & x_{r,2} + ix_{i,2} & x_{r,3} + ix_{i,3} \end{pmatrix}, \tag{1.27}$$

If we then switch to either the form (1.14) or (1.15), we will not get the vector of conjugate complex numbers, but rather an object of the original complex variables with the switched rows and columns:

$$\mathbf{x}^\top = \begin{pmatrix} x_{r,1} & x_{r,2} & x_{r,3} \\ x_{i,1} & x_{i,2} & x_{i,3} \end{pmatrix} \tag{1.28}$$

and

$$\underset{\sim}{\mathbf{x}}^\top = \begin{pmatrix} x_{r,1} & -x_{i,1} & x_{r,2} & -x_{i,2} & x_{r,3} & -x_{i,3} \\ x_{i,1} & x_{r,1} & x_{i,2} & x_{r,2} & x_{i,3} & x_{r,3} \end{pmatrix}. \tag{1.29}$$

We will denote the operation of transposition above with the symbol $\top$ and will call it just "transposition". The transposition that produces the conjugate complex numbers is called "conjugate transposition" and will be denoted as $\prime$ in

this monograph. For the three examples above, the conjugate transposition will be:

$$\underline{\mathbf{x}}' = \begin{pmatrix} x_{r,1} - ix_{i,1} & x_{r,2} - ix_{i,2} & x_{r,3} - ix_{i,3} \end{pmatrix}, \tag{1.30}$$

$$\mathbf{x}' = \begin{pmatrix} x_{r,1} & x_{r,2} & x_{r,3} \\ -x_{i,1} & -x_{i,2} & -x_{i,3} \end{pmatrix} \tag{1.31}$$

and

$$\underset{\sim}{\mathbf{x}}' = \begin{pmatrix} x_{r,1} & x_{i,1} & x_{r,2} & x_{i,2} & x_{r,3} & x_{i,3} \\ -x_{i,1} & x_{r,1} & -x_{i,2} & x_{r,2} & -x_{i,3} & x_{r,3} \end{pmatrix}. \tag{1.32}$$

1.2 Complex random variable

In statistics, a random variable is a variable, the value of which depends on random events. For example, for the classical coin tossing experiment, x would be considered as a random variable if it is equal to 1 in case of heads and is equal to zero otherwise. In case of complex random variables (c.r.v.), the logic is similar, but we would typically deal with two-dimensional cases (e.g. tossing two coins simultaneously). We accept that in some cases, some parts of the c.r.v. might become non-random (e.g. the real part becomes equal to some fixed number). As such, we would be saying that $x = x_r + ix_i$ is a complex random variable if at least one part of it follows some distribution. Furthermore, while random variables can be measured in a variety of scales (coin tossing represents the nominal scale), in this monograph we focus the discussion on continuous numerical variables.

As any other random variable, c.r.v. can be characterised by its moments. However, some of its moments differ from the conventional ones for real variables and are based on the properties of complex numbers.

1.2.1 First moment

The first moment of any c.r.v. is relatively straightforward. For the variable $\underline{x} = x_r + ix_i$ it is:

$$\underline{\mu} = \mathrm{E}(\underline{x}) = \mathrm{E}(x_r + ix_i) = \mathrm{E}(x_r) + i\mathrm{E}(x_i) = \mu_r + i\mu_i, \tag{1.33}$$

due to the law of total expectation and because i is a constant. In (1.33), E(.) is the expectation of a variable, $\underline{\mu}$ is the first moment of a complex variable and μ_r and μ_i are the respective real and imaginary parts of the first moment. In sample, this moment can be calculated as:

$$\underline{\hat{\mu}} = \hat{\mu}_r + i\hat{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} x_{r,j} + i\frac{1}{n}\sum_{j=1}^{n} x_{i,j}, \tag{1.34}$$

where n is the sample size and $\hat{\mu}$ is the estimate of μ. The properties of the first moment of c.r.v. are similar to the properties of any other random variable. The

only thing to note is that the moment can also be represented in a vector or matrix form (see Subsection 1.1.1):

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_r \\ \mu_i \end{pmatrix} \tag{1.35}$$

or

$$\underset{\sim}{\boldsymbol{\mu}} = \begin{pmatrix} \mu_r & -\mu_i \\ \mu_i & \mu_r \end{pmatrix}. \tag{1.36}$$

In R, the first moment can be obtained via the `mean()` function from `stats` package:

```
# Create a random variable
x <- complex(real=rnorm(100, mean=50, sd=10),
             imaginary=rnorm(100, mean=100, sd=10))
# Calculate the first moment
mean(x)
```

```
## [1] 50.88343+97.36909i
```

1.2.2 Second moment

While there are raw moments for c.r.v., here we will focus on the centred ones, i.e. moments for $\underline{x} - \bar{\underline{x}}$. There are several instruments related to the second moment that could characterise a complex random variable. The first one is called "variance" and is based on the multiplication by a conjugate number (**?**). So for the c.r.v. $x_r + ix_i$ the variance is:

$$\begin{aligned} \mathrm{V}(\underline{x}) = \sigma_x^2 = & \mathrm{E}((\underline{x} - \underline{\mu})(\tilde{\underline{x}} - \tilde{\underline{\mu}})) = \\ & \mathrm{E}\left(((x_r - \mu_r) + i(x_i - \mu_i))((x_r - \mu_r) - i(x_i - \mu_i))\right) = \\ & \mathrm{E}((x_r - \mu_r)^2) + \mathrm{E}((x_i - \mu_i)^2) \end{aligned} \tag{1.37}$$

or taking that the variances of the real and the imaginary parts are $\sigma_{x_r}^2$ and $\sigma_{x_i}^2$ respectively:

$$\sigma_x^2 = \sigma_{x_r}^2 + \sigma_{x_i}^2. \tag{1.38}$$

Multiplication by conjugate in (1.37) allows keeping the size of variability for both real and imaginary parts, but at the same time masks the individual contribution of each part in the overall variance and removes the covariance between the parts. The resulting values shows in a way the overall variability of a c.r.v. along a diagonal line as shown in Figure 1.6.

Figure 1.6 demonstrates graphically the relation between standard deviations of real, imaginary and the overall variance of a complex variable according to the formula (1.38).

? note that the moment (1.38) is not sufficient to entirely describe the second order statistics of a c.r.v. This is because it ignores the potential covariance

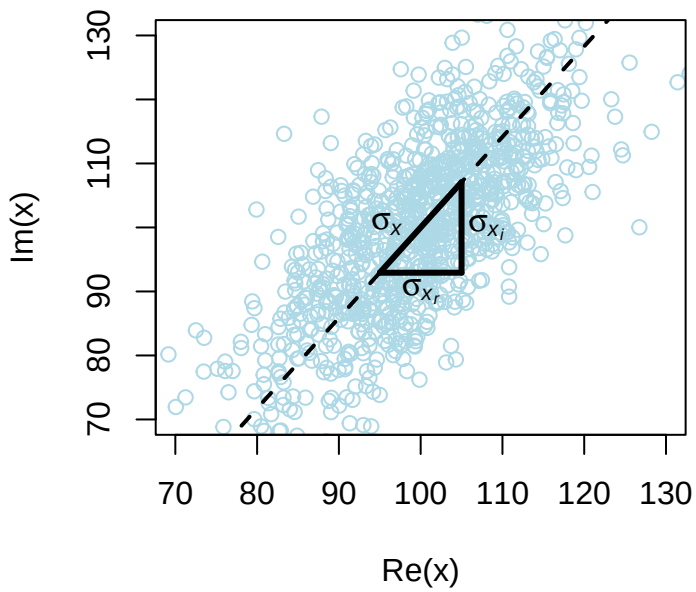

Figure 1.6: Visual representation of variance of a complex random variable.

between the parts of a variable and only describes their average variability. Another measure of variability of a complex random variable is the so called "pseudo-variance" (**?**), which can be obtained by applying the conventional formula of variance directly to the c.r.v. without the multiplication by conjugate:

$$\begin{aligned}\mathcal{V}(\underline{x}) = \varsigma_x^2 =& \mathrm{E}((\underline{x}-\underline{\mu})^2) = \mathrm{E}\left(((x_r-\mu_r)+i(x_i-\mu_i))^2\right) = \\ & \mathrm{E}((x_r-\mu_r)^2) - \mathrm{E}((x_i-\mu_i)^2) + i2\mathrm{E}((x_r-\mu_r)(x_i-\mu_i))\end{aligned} \tag{1.39}$$

or using the notation σ_{x_r,x_i} for covariance between the real and imaginary parts:

$$\varsigma_x^2 = \sigma_{x_r}^2 - \sigma_{x_i}^2 + i2\sigma_{x_r,x_i}. \tag{1.40}$$

In the literature, the moment (1.40) is called pseudo-variance (**?**), because it does not measure the variability of a variable, but rather gives a different information: it shows whether the real and imaginary variances are similar and what the covariance between them is. If both real and imaginary parts of (1.40) are equal to zero then it is said that the distribution of the complex variable x is spherical (**?**), i.e. the variances are similar and the real and imaginary parts are not linearly related. Furthermore, we should point out that while minimising (1.38) implies minimising both variances of the real and imaginary parts of x (ignoring the covariance between them), minimising (1.40) is not as straightforward. At very least, we could tell that it implies making the distribution of x closer to the spherical one, but we cannot provide any thorough insight about this.

We should note that both measures can be considered as complex, and the main distinction between them is the multiplication by the same c.r.v or by its

conjugate. As such we propose to call them respectively "conjugate variance" and "direct variance".

In R, the conjugate and direct variances are available in `cvar()` function of `complex` package.

```
# Create a random variable
x <- complex(real=rnorm(100, mean=50, sd=5),
             imaginary=rnorm(100, mean=100, sd=10))
# Calculate the conjugate variance
cvar(x, method="conjugate") |>
    setNames("Conjugate variance")
```

```
## Conjugate variance
## 129.4027+0i
```

```
# Calculate the direct variance
cvar(x, method="direct") |>
    setNames("Direct variance")
```

```
##     Direct variance
## -80.02685-20.40511i
```

As we see from the output above, the direct variance has the negative real part, which indicates that the imaginary part has higher variance than the real one. The imaginary part of the direct variance shows the double covariance between the real and imaginary parts. Finally, the conjugate variance should be $5^2 + 10^2 = 125$, but due to small sample (100 observations in the generated random variable above), it will be equal to a number close to it.

Given that any complex variable can be represented in a vector form, the c.r.v. x can be treated as a bivariate random variable, for which a covariance matrix can be calculated via:

$$\mathbf{\Sigma}_x = \begin{pmatrix} \sigma^2_{x_r} & \sigma_{x_r,x_i} \\ \sigma_{x_r,x_i} & \sigma^2_{x_i} \end{pmatrix}. \tag{1.41}$$

The minimisation of the matrix (1.41) does not make sense, but instead it is possible to minimise the determinant of that matrix, which is called "Generalised Variance" (GV):

$$\text{GV} = |\mathbf{\Sigma}_x| = \sigma^2_{x_r}\sigma^2_{x_i} - \sigma^2_{x_r,x_i}. \tag{1.42}$$

In R, this is implemented in `covar()` function from the `complex` package:

```
covar(x)
```

```
##           x_r       x_i
## x_r  24.68794 -10.20256
## x_i -10.20256 104.71479
```

The minimisation of GV, as can be seen from the formula (1.42), implies the simultaneous minimisation of variances of real and imaginary parts and maximisation

of the square of covariance between them, making the resulting distribution compacter and emphasising the potential relations between the real and imaginary parts of a c.r.v. For our example in R, the GV equals to:

```
covar(x) |> det()
```

```
## [1] 2481.1
```

All the three moments discussed in this subsection rely on variances of real and imaginary parts of a c.r.v. and on a covariance between those parts. These moments in turn can be calculated using the conventional formulae, correcting for the potential small sample bias (**?**):

$$\begin{aligned}
\hat{\sigma}^2_{x_r} =& \frac{1}{n-k}\sum_{j=1}^{n}(x_{r,j}-\bar{x}_{r,j})^2 \\
\hat{\sigma}^2_{x_i} =& \frac{1}{n-k}\sum_{j=1}^{n}(x_{i,j}-\bar{x}_{i,j})^2 \\
\hat{\sigma}_{x_r,x_i} =& \frac{1}{n-k}\sum_{j=1}^{n}(x_{r,j}-\bar{x}_{r,j})(x_{i,j}-\bar{x}_{i,j})^2,
\end{aligned} \tag{1.43}$$

where k is the number of estimated parameters in a model.

Finally, similar to the variance it is possible to calculate second moments between two complex random variables $x = x_r + ix_i$ and $y = y_r + iy_i$ (**?**). The "conjugate" covariance is calculated similarly to (1.37) via the multiplication by conjugate:

$$\begin{aligned}
\sigma_{x,y} =& \mathrm{E}((\tilde{x}-\tilde{\mu}_x)(y-\mu_y)) = \\
& \mathrm{E}\left(((x_r-\mu_{x,r}) - i(x_i-\mu_{x,i}))((y_r-\mu_{y,r}) + i(y_i-\mu_{y,i}))\right) = \\
& \mathrm{E}((x_r-\mu_{x,r})(y_r-\mu_{y,r})) + \mathrm{E}((x_i-\mu_{x,i})(y_i-\mu_{y,i})) + \\
& i\left(\mathrm{E}((x_r-\mu_{x,r})(y_i-\mu_{y,i})) - \mathrm{E}((x_i-\mu_{x,i})(y_r-\mu_{y,r}))\right)
\end{aligned} \tag{1.44}$$

where μ_x and μ_y are the respective first moments of c.r.v. x and y. Using the notations above, the same covariance can be rewritten as:

$$\sigma_{x,y} = \sigma_{x_r,y_r} + \sigma_{x_i,y_i} + i(\sigma_{x_r,y_i} - \sigma_{x_i,y_r}), \tag{1.45}$$

which similarly to the variance (1.38) mixes the moments of the real and imaginary parts of the two random variables. Note though that if the conjugate of y is used instead of the conjugate of x, the imaginary part of the covariance (1.45) will change to $\sigma_{x_i,y_r} - \sigma_{x_r,y_i}$, giving potentially different information about the relation between the variables.

In order to have more information about a c.r.v., we also need to consider the "direct" covariance (also known in the literature as pseudo-covariance, **?**), which can be shown to be equal to:

$$\varsigma_{x,y} = \sigma_{x_r,y_r} - \sigma_{x_i,y_i} + i(\sigma_{x_i,y_r} + \sigma_{x_r,y_i}). \tag{1.46}$$

Note that none of these moments on its own gives enough information about the relation between two complex variables, so they need to be used jointly. Alternatively, using vector representation, a covariance matrix between the two c.r.v. can be used to get a better understanding about the relations between them:

$$\mathbf{\Sigma}_{x,y} = \begin{pmatrix} \sigma^2_{x_r} & \sigma_{x_r,x_i} & \sigma_{x_r,y_r} & \sigma_{x_r,y_i} \\ \sigma_{x_r,x_i} & \sigma^2_{x_i} & \sigma_{x_i,y_r} & \sigma_{x_i,y_i} \\ \sigma_{x_r,y_r} & \sigma_{x_i,y_r} & \sigma^2_{y_r} & \sigma_{y_r,y_i} \\ \sigma_{x_r,y_i} & \sigma_{x_i,y_i} & \sigma_{y_r,y_i} & \sigma^2_{y_i} \end{pmatrix}. \tag{1.47}$$

In R, the complex covariances are implemented in `ccov()` function from the `complex` package:

```
# Create a random variable x
x <- complex(real=rnorm(100, mean=50, sd=5),
             imaginary=rnorm(100, mean=100, sd=10))
# Create a random variable y
y <- (1.5 + 3i) + (0.5 - 0.75i) * x +
            complex(real=rnorm(100, mean=0, sd=10),
                    imaginary=rnorm(100, mean=0, sd=10))
# Calculate the conjugate variance
ccov(x, y, method="conjugate") |>
    setNames("Conjugate covariance")
```

```
## Conjugate covariance
## 68.76492-90.30221i
```

```
# Calculate the direct variance
ccov(x, y, method="direct") |>
    setNames("Direct covariance")
```

```
##   Direct covariance
## -29.06618+38.12778i
```

The functions `cvar()` and `ccov()` also accept a matrix instead of vector `x`, in which case they will produce a matrix of moments with complex variances on diagonal and complex covariances on the off-diagonals.

```
ourData <- cbind(x,y)
cvar(ourData, method="direct")
```

```
##                      x                   y
## x -57.52798- 3.58447i -29.06618+38.12778i
## y -29.06618+38.12778i  32.43004+69.16716i
```

While there exist higher order moments for complex random variables, we do not discuss them in this book.

1.3 Parametric Distributions of c.r.v.

Similarly to the real valued random variables, complex random variables can follow some parametric distributions. In this section, we discuss several simple ones, some of which will be used later in this monograph.

In general, it is reasonable to assume that a c.r.v. has a distribution of a shape of a circle or ellipse on the complex plane. This becomes more apparent if we consider the exponential form of a c.r.v. (1.9) and consider both magnitude and the angle as random variables. In the simplest case, when they are not correlated and we consider a variable that is centred around the origin, the randomness coming from both of these components will result in a circle on the complex plane, see the following example in R (and Figure 1.7):

```
# Generate magnitudes from the uniform distribution
R <- runif(1000,0,10)
# Generate angles from the uniform distribution
phi <- runif(1000,-pi,pi)
# Create c.r.v in exponential form
x <- R * exp(complex(imaginary=phi))
# Plot the variable
plot(x)
```

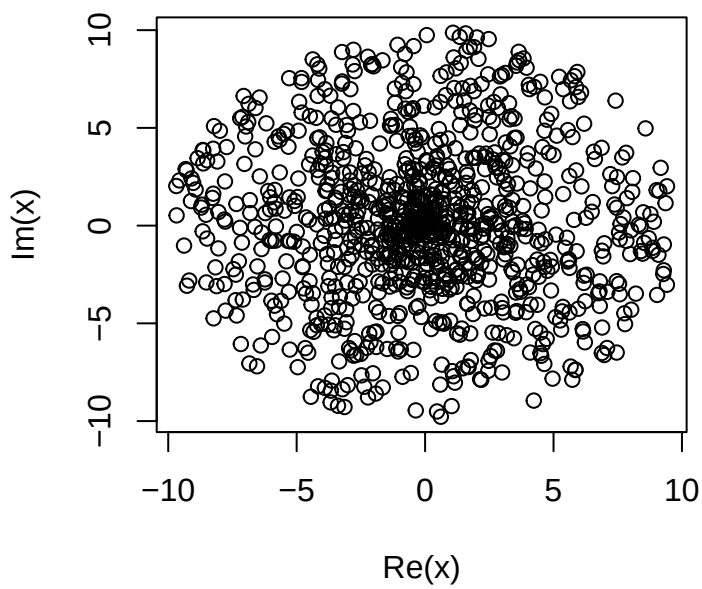

Figure 1.7: C.r.v. generated from a uniform distribuiton.

The example in Figure 1.7 demonstrates how a c.r.v. with uncorrelated real and imaginary parts looks on the complex plane. If there is a relation between the parts then the circle transforms into ellipse. So, it is only logical to consider the

distributions that rely on a circle or an ellipse for the c.r.v. instead of any other geometric shapes.

When it comes to distribution functions, in case of c.r.v. in general we should consider the joint bivariate distribution, thus all the probability density and cumulative distribution functions will be plotted in 3D with the density/probability in z-axis and the complex variable in x and y axes.

1.3.1 Complex Normal distribution

One of the most popular distributions in statistics is the Normal distribution. There exists a complex counterpart of that distribution, which has the following probability density function (**?**):

$$f(\underline{x}) = \frac{1}{\pi^2\sqrt{\sigma^4 - \varsigma^2\tilde{\varsigma}^2}} \exp\left(-\frac{1}{2}\begin{pmatrix}\tilde{\underline{x}} - \tilde{\underline{\mu}} & \underline{x} - \underline{\mu}\end{pmatrix}\begin{pmatrix}\sigma^2 & \varsigma^2 \\ \tilde{\varsigma}^2 & \sigma^2\end{pmatrix}^{-1}\begin{pmatrix}\underline{x} - \underline{\mu} \\ \tilde{\underline{x}} - \tilde{\underline{\mu}}\end{pmatrix}\right), \tag{1.48}$$

where $\underline{\mu}$ is the first moment of the c.r.v. $\underline{x}$ and σ^2 and ς^2 are the conjugate and direct variances as defined in Section 1.2. The variable that follows this distribution can be denoted as $\underline{x} \sim \mathcal{CN}(\underline{\mu}, \sigma^2, \varsigma^2)$ Note that both σ^2 and ς^2 are important in the PDF (1.48) to define the shape of the distribution. For the real-valued variables, the conventional standard deviation σ characterises the dispersion of the random variable, and in case of the normal distribution shows how many z-scores away the random variable can lie from its centre. For the c.r.v., σ plays a similar role, but not in terms of units along the x axis, but rather in terms of radii from the centre. The direct variance ς^2, on the other hand characterises the shape of the distribution itself: its real part shows whether the ellipse of the normal distribution should be wide (positive numbers) or narrow (negative numbers), while the imaginary part shows how close the points should be to the line drawn through them. These moments are shown visually in Figure 1.8.

The overall conjugate standard deviation σ_x represents the radius of the unit circle in Figure 1.8, while the imaginary part of ς_x^2 (i.e. $2\sigma_{x_r,x_i}$) is shown with a green line. The real part of ς_x^2 is not visualised, but for this plot it will be negative, showing that the distribution is narrower along the x-axis than along the y-axis.

Figure 1.9 shows the PDF of $x \sim \mathcal{CN}(100 + 100i, 100, -20 + i30)$, while Figure 1.10 shows the heatmap of the same PDF. The brighter areas in both Figures correspond to the higher values of PDF. For each specific value of density, there is a multitude of values of x, forming an ellipse of a specific length.

One special case of a Complex Normal distribution is a so called "Circular Symmetric" Complex Normal distribution. It is a distribution for which the direct variance equals to zero, which implies that the variances of the real and the imaginary parts are equal and that the covariance between them is equal to

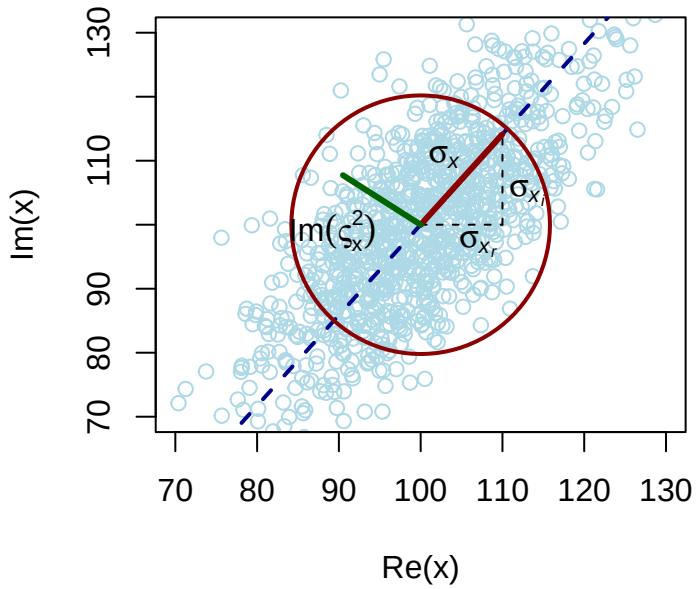

Figure 1.8: Visual representation of variances for a c.r.v. that follows Complex Normal distribution.

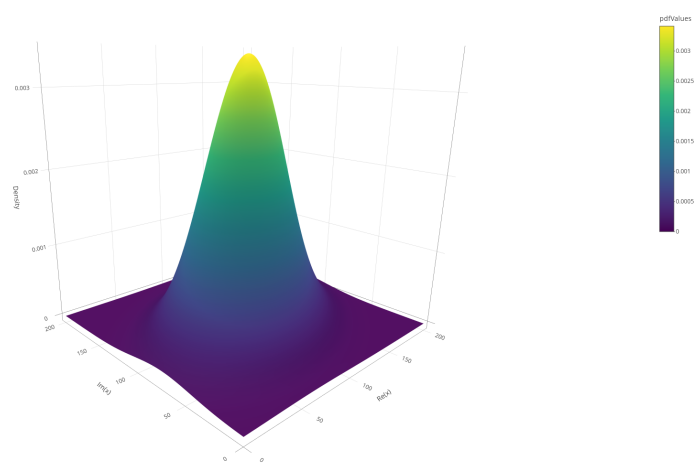

Figure 1.9: Plot of the probability density function of complex normal distribution.

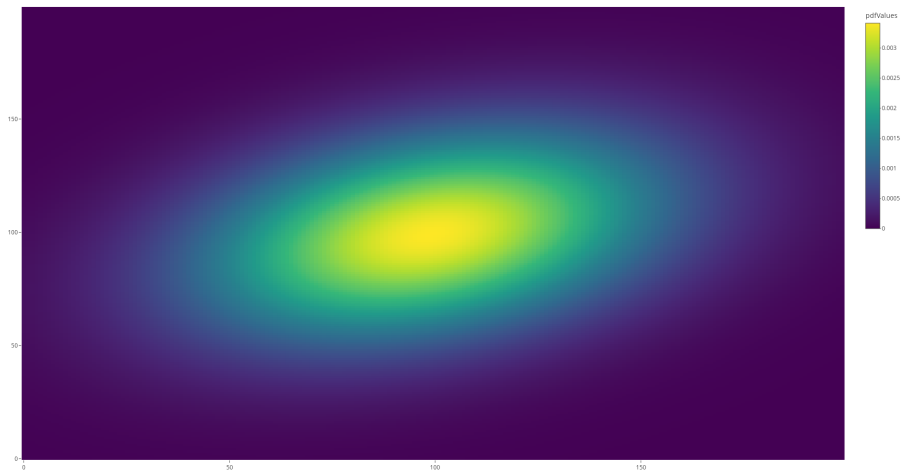

Figure 1.10: Heatmap of density of Complex Normal distribution.

zero. This specific distribution is used extensively in signal processing literature (**?**).

The functions of the Complex Normal distribution are implemented as `dcnorm()`, `pcnorm()`, `qmcnorm()` and `rmcnorm()` in the `complex` package.

The Complex Normal distribution can be used for the joint estimation of parameters of a complex-valued model and for generation of prediction intervals from it. However, in some cases it might be more convenient to consider a c.r.v. model as a bivariate vector model based on (1.22) and thus to revert to the multivariate normal distribution.

1.3.2 Multivariate Normal distribution

The same normal distribution for a complex variable can be parametrised via a covariance matrix $\mathbf{\Sigma}$ from (1.41) instead of the conjugate and direct variances σ^2 and ς^2. In general, the Probability Density Function for a vector of n variables can be written as:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}\right)^{\top}\mathbf{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right). \tag{1.49}$$

For the bivariate case of a c.r.v. represented in a form of a vector, $n = 2$ and the covariance matrix $\mathbf{\Sigma}$ has dimensionality of 2×2. In that case the PDF will be the same as for the Complex Normal distribution. In fact, it is possible to recreate the covariance matrix based on the values of conjugate and direct variances:

$$\mathbf{\Sigma} = \frac{1}{2}\begin{pmatrix} \sigma^2 + \mathcal{R}(\varsigma^2) & \mathcal{I}(\varsigma^2) \\ \mathcal{I}(\varsigma^2) & \sigma^2 - \mathcal{R}(\varsigma^2) \end{pmatrix}. \tag{1.50}$$

When it comes to constructing confidence regions from the multivariate normal distribution then the following inequality is used to determine values of the vector of quantiles $\mathbf{q}$:

$$(\mathbf{x}-\mathbf{q})^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{q}) \leq \chi^2(n,p), \tag{1.51}$$

where p is the confidence level. For the case of c.r.v., when $n=2$, the $\chi^2(n,p)$ transforms into the exponential distribution $\mathcal{E}(0.5)$.

Note however that working with confidence regions is challenging, because there is an infinite number of values of the vector $\mathbf{q}$ that give the same critical value $\chi^2(n,p)$. So, in order to make it practical, unconditional prediction intervals can be constructed for each of the parts, dropping the other one. In this case, the classical formula for the intervals can be used, for example for the real part of the c.r.v.:

$$x_r \in \left(\mu_{x_r} + \sigma_{x_r} z\left(\frac{\alpha}{2}\right), \mu_{x_r} + \sigma_{x_r} z\left(\frac{1+\alpha}{2}\right)\right), \tag{1.52}$$

where $z\left(\frac{\alpha}{2}\right)$ and $z\left(\frac{1+\alpha}{2}\right)$ are the respective lower and upper quantiles of the standard normal distribution.

1.3.3 Hotelling's T-squared distribution

When it comes to distribution of complex variables statistics, one of the most important ones is a distribution of a mean of a complex variable. In the case of real variables, it is well known that if the Central Limit Theorem (CLT) holds then the sample mean will follow Normal distribution, and confidence interval for the mean can be constructed using Student's t distribution (when the population variance is unknown, i.e. in reality). For c.r.v., the situation is similar: when CLT holds, the c.r.v. follows Complex Normal distribution, but instead of Student's t distribution, we need to use its multivariate counterpart, which is called "Hotelling's T-squared" distribution. It is more convenient to consider the statistics for a c.r.v in a vector form (i.e. treat it as bivariate). If we estimate the sample mean $\hat{\boldsymbol{\mu}}_x$ and a sample covariance matrix of the mean $\hat{\mathbf{\Sigma}}_{\mu_x}$ then the following holds (if CLT holds):

$$(\hat{\boldsymbol{\mu}}_x - \boldsymbol{\mu}_x)' \hat{\mathbf{\Sigma}}_{\mu_x}^{-1} (\hat{\boldsymbol{\mu}}_x - \boldsymbol{\mu}_x) \sim T^2(2, n-1) = \frac{2(n-1)}{n-2} F(2, n-2), \tag{1.53}$$

where $T^2(2,n-1)$ is a T-squared Hotelling's statistics with 2 and $n-1$ degrees of freedom, and $F(2,n-2)$ is the Fisher's distribution statistics with 2 and $n-2$ degrees of freedom. Equation (1.53) can be used to construct confidence interval for a multivariate mean or for testing statistical hypotheses. One of the possible hypotheses that would make sense in case of c.r.v. and could be tested based on (1.53) is the following:

$$\begin{aligned} \mathrm{H}_0 &: \boldsymbol{\mu}_x = 0 \\ \mathrm{H}_1 &: \boldsymbol{\mu}_x \neq 0 \end{aligned} \tag{1.54}$$

or equivalently:

$$\begin{aligned} &\mathrm{H}_0 : \mu_{x_r} = \mu_{x_i} = 0 \\ &\mathrm{H}_1 : \mu_{x_r} \neq 0 \vee \mu_{x_i} \neq 0 \end{aligned} \tag{1.55}$$

However, from practical point of view, testing the joint hypothesis (1.55) might not be very useful. Instead, testing the classical hypotheses from the univariate context based on Student's t distribution for each part of c.r.v. might provide more detailed information, i.e. whether the specific real or imaginary part of the variable equals to zero or not. In contrast, in case of the hypothesis (1.55), rejecting the null hypothesis implies that something is not equal to zero, but it is not clear, what specifically.

Chapter 2

Simple Complex Linear Regression

We start with the simplest Complex Linear Regression (CLR), which we call "simple", because it captures the relation between two variables. The main difference from the conventional Simple Linear Regression is that each of the variables is complex. This makes the model more complicated than the one in real numbers domain.

2.1 Model formulation

The simple Complex Linear Regression can be written as:

$$\underline{y_j} = \underline{\beta_0} + \underline{\beta_1}\underline{x_j} + \underline{\epsilon_j}, \tag{2.1}$$

where j is the index for an observation and every parameter and variable is a complex number, i.e. $\underline{y_j} = y_{r,j} + iy_{i,j}$ is the complex response variable, $\underline{x_j} = x_{r,j} + ix_{i,j}$ is the complex explanatory variable, $\underline{\beta_l} = \beta_{l,r} + i\beta_{l,i}$ is the l-th parameter and $\underline{\epsilon_j} = \epsilon_{r,j} + i\epsilon_{i,j}$ is the error term. For now, we do not make any specific assumptions about the distribution of the complex error term, we will come back to this later in this chapter. Inserting these values in (2.1) leads to:

$$y_{r,j} + iy_{i,j} = (\beta_{0,r} + i\beta_{0,i}) + (\beta_{1,r} + i\beta_{1,i})(x_{r,j} + ix_{i,j}) + (\epsilon_{r,j} + i\epsilon_{i,j}), \tag{2.2}$$

which is in fact a multivariate model, capturing how the change of values in pair of variables x_r, x_i leads to the change in the pair of variables y_r and y_i. Given that any complex equation can be represented as a system of two equations, the model (2.2) can be represented as a system of two linear regressions:

$$\begin{aligned} y_{r,j} =& \beta_{0,r} + \beta_{1,r}x_{r,j} - \beta_{1,i}x_{i,j} + \epsilon_{r,j} \\ y_{i,j} =& \beta_{0,i} + \beta_{1,r}x_{i,j} + \beta_{1,i}x_{r,j} + \epsilon_{i,j} \end{aligned} \tag{2.3}$$

This model captures a very specific dynamics between the real and imaginary parts, given that they share the same set of parameters for the slope. But most importantly it shows how each complex variable y relates to variable x in a four dimensional space. Note that the real and imaginary parts of y can be exchanged without a serious impact on the model: in that case the values of parameters would change, but the relation between x and y would stay the same. Similarly, changing x_r with x_i would only lead to different estimates of parameters, the general relation will hold.

Given that we deal with a sample of values, the model (2.1) can be represented in a vector form for all observations j from 1 to n (based on discussion in Subsection 1.1.2):

$$\underline{\mathbf{y}} = \underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\epsilon}}, \tag{2.4}$$

where $\underline{\mathbf{y}} = \begin{pmatrix} \underline{y_1} \\ \underline{y_2} \\ \vdots \\ \underline{y_n} \end{pmatrix}$, $\underline{\mathbf{X}} = \begin{pmatrix} 1 & \underline{x_1} \\ 1 & \underline{x_2} \\ \vdots & \\ 1 & \underline{x_n} \end{pmatrix}$, $\underline{\boldsymbol{\beta}} = \begin{pmatrix} \underline{\beta_0} \\ \underline{\beta_1} \end{pmatrix}$ and $\underline{\boldsymbol{\epsilon}} = \begin{pmatrix} \underline{\epsilon_1} \\ \underline{\epsilon_2} \\ \vdots \\ \underline{\epsilon_n} \end{pmatrix}$, where each element of the objects above is a complex number.

Going even further, using the vector and matrix representations of complex variables, the same system of equations (2.3) can be rewritten as:

$$\begin{pmatrix} y_{r,j} \\ y_{i,j} \end{pmatrix} = \begin{pmatrix} \beta_{0,r} \\ \beta_{0,i} \end{pmatrix} + \begin{pmatrix} x_{r,j} & -x_{i,j} \\ x_{i,j} & x_{r,j} \end{pmatrix} \begin{pmatrix} \beta_{1,r} \\ \beta_{1,i} \end{pmatrix} + \begin{pmatrix} \epsilon_{r,j} \\ \epsilon_{i,j} \end{pmatrix}, \tag{2.5}$$

or uniting all the parameters in one vector:

$$\begin{pmatrix} y_{r,j} \\ y_{i,j} \end{pmatrix} = \begin{pmatrix} 1 & 0 & x_{r,j} & -x_{i,j} \\ 0 & 1 & x_{i,j} & x_{r,j} \end{pmatrix} \begin{pmatrix} \beta_{0,r} \\ \beta_{0,i} \\ \beta_{1,r} \\ \beta_{1,i} \end{pmatrix} + \begin{pmatrix} \epsilon_{r,j} \\ \epsilon_{i,j} \end{pmatrix}. \tag{2.6}$$

This form can then be represented in a classical matrix notations:

$$\mathbf{y}_j = \underset{\sim}{\mathbf{X}}_j \boldsymbol{\beta} + \boldsymbol{\epsilon}_j. \tag{2.7}$$

And if we stack each of the vectors and matrices for each $j = 1 \dots n$ we get even more compact form:

$$\mathbf{Y} = \underset{\sim}{\mathbf{X}}\boldsymbol{\beta} + \mathbf{E}, \tag{2.8}$$

where $\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}$, $\underset{\sim}{\mathbf{X}} = \begin{pmatrix} \underset{\sim}{\mathbf{X}}_1 \\ \underset{\sim}{\mathbf{X}}_2 \\ \vdots \\ \underset{\sim}{\mathbf{X}}_n \end{pmatrix}$ and $\mathbf{E} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_n \end{pmatrix}$. The form (2.8) shows the connection between the complex regression and the conventional real valued one.

As it can be seen, they both can be represented in matrix forms, which means that some of the approaches used for the latter can be transferred to the former under some conditions.

2.2 Estimation

Whenever we estimate a model, we substitute the "true" parameters by their sample estimates. So, for the CLR, the model applied to the data should be written as:

$$\underline{y_j} = \underline{b_0} + \underline{b_1}\underline{x_j} + \underline{e_j}, \tag{2.9}$$

or

$$y_{r,j} + iy_{i,j} = (b_{0,r} + ib_{0,i}) + (b_{1,r} + ib_{1,i})(x_{r,j} + ix_{i,j}) + (e_{r,j} + ie_{i,j}), \tag{2.10}$$

or

$$\begin{aligned} y_{r,j} =& b_{0,r} + b_{1,r}x_{r,j} - b_{1,i}x_{i,j} + e_{r,j} \\ y_{i,j} =& b_{0,i} + b_{1,r}x_{i,j} + b_{1,i}x_{r,j} + e_{i,j} \end{aligned} \tag{2.11}$$

or equivalently in matrix notations:

$$\mathbf{y}_j = \underset{\sim}{\mathbf{X}_j}\boldsymbol{b} + \boldsymbol{e}_j, \tag{2.12}$$

where $\underline{b_l}$, $b_{l,r} + ib_{l,i}$ and $\boldsymbol{b}$ are the estimates of respective $\underline{\beta_l}$, $\beta_{l,r} + i\beta_{l,i}$ and $\boldsymbol{\beta}$ and $\underline{e_j}$, $e_{r,j} + ie_{i,j}$ and $\boldsymbol{e}_j$ are the residuals of the model. Now in order to get estimates of parameters, we can use one of several approaches.

2.2.1 Ordinary Least Squares

We start with the conventional Ordinary Least Squares (OLS), which relies on the multiplication by conjugate number. For the simple CLR, it comes to minimising the following loss based on the residuals:

$$\sum_{j=1}^{n}(\underline{e_j}\underline{\tilde{e}_j}) = \sum_{j=1}^{n}(e_{r,j}^2 + e_{i,j}^2). \tag{2.13}$$

The residuals can be substituted as: $\underline{e_j} = \underline{y_j} - \underline{b_0} - \underline{b_1}\underline{x_j}$ and $\underline{\tilde{e}_j} = \underline{\tilde{y}_j} - \underline{\tilde{b}_0} - \underline{\tilde{b}_1}\underline{\tilde{x}_j}$ to get:

$$\sum_{j=1}^{n}(\underline{y_j} - \underline{b_0} - \underline{b_1}\underline{x_j})(\underline{\tilde{y}_j} - \underline{\tilde{b}_0} - \underline{\tilde{b}_1}\underline{\tilde{x}_j}), \tag{2.14}$$

which after opening the brackets becomes:

$$\sum_{j=1}^{n}\left(\underline{y_j}\underline{\tilde{y}_j} - \underline{y_j}\underline{\tilde{b}_0} - \underline{y_j}\underline{\tilde{b}_1}\underline{\tilde{x}_j} - \underline{b_0}\underline{\tilde{y}_j} + \underline{b_0}\underline{\tilde{b}_0} + \underline{b_0}\underline{\tilde{b}_1}\underline{\tilde{x}_j} - \underline{b_1}\underline{x_j}\underline{\tilde{y}_j} + \underline{b_1}\underline{x_j}\underline{\tilde{b}_0} + \underline{b_1}\underline{x_j}\underline{\tilde{b}_1}\underline{\tilde{x}_j}\right). \tag{2.15}$$

In order to minimise the sum of squared errors (2.13), we need to take derivative of (2.15) with respect to each of the parameters $b_{0,r}$, $b_{0,i}$, $b_{1,r}$ and $b_{1,i}$ and equate each of the resulting equations to zero to find the extrema:

$$\begin{aligned}
\frac{d\sum_{j=1}^{n}(\underline{e_j}\underline{\tilde{e}_j})}{db_{0,r}} &= 0 \\
\frac{d\sum_{j=1}^{n}(\underline{e_j}\underline{\tilde{e}_j})}{db_{0,i}} &= 0 \\
\frac{d\sum_{j=1}^{n}(\underline{e_j}\underline{\tilde{e}_j})}{db_{1,r}} &= 0 \\
\frac{d\sum_{j=1}^{n}(\underline{e_j}\underline{\tilde{e}_j})}{db_{1,i}} &= 0
\end{aligned} \tag{2.16}$$

to get:

$$\begin{aligned}
&-\sum_{j=1}^{n}\underline{y_j} - \sum_{j=1}^{n}\underline{\tilde{y}_j} + 2nb_{0,r} + \underline{\tilde{b}_1}\sum_{j=1}^{n}\underline{\tilde{x}_j} + \underline{b_1}\sum_{j=1}^{n}x_j = 0 \\
&i\sum_{j=1}^{n}\underline{y_j} - i\sum_{j=1}^{n}\underline{\tilde{y}_j} + 2nb_{0,i} + i\underline{\tilde{b}_1}\sum_{j=1}^{n}\underline{\tilde{x}_j} - i\underline{b_1}\sum_{j=1}^{n}x_j = 0 \\
&-\sum_{j=1}^{n}\underline{y_j\tilde{x}_j} + \underline{b_0}\sum_{j=1}^{n}\underline{\tilde{x}_j} - \sum_{j=1}^{n}\underline{x_j\tilde{y}_j} + \underline{\tilde{b}_0}\sum_{j=1}^{n}x_j + 2b_{1,r}\sum_{j=1}^{n}\underline{x_j\tilde{x}_j} = 0 \\
&i\sum_{j=1}^{n}\underline{y_j\tilde{x}_j} - i\underline{b_0}\sum_{j=1}^{n}\underline{\tilde{x}_j} - i\sum_{j=1}^{n}\underline{x_j\tilde{y}_j} + i\underline{\tilde{b}_0}\sum_{j=1}^{n}x_j + 2b_{1,i}\sum_{j=1}^{n}\underline{x_j\tilde{x}_j} = 0.
\end{aligned} \tag{2.17}$$

As shown by **?**, solving the system of equations (2.17) gives the following formulae for the parameters of the model:

$$\begin{aligned}
\underline{b_1} &= \frac{\sum_{j=1}^{n}(\underline{y_j} - \underline{\hat{\mu}_y})(\underline{\tilde{x}_j} - \underline{\hat{\tilde{\mu}}_x})}{\sum_{j=1}^{n}(\underline{x_j} - \underline{\hat{\mu}_x})(\underline{\tilde{x}_j} - \underline{\hat{\tilde{\mu}}_x})} \\
\underline{b_0} &= \frac{1}{n}\sum_{j=1}^{n}\underline{y_j} - \underline{b_1}\sum_{j=1}^{n}\underline{x_j},
\end{aligned} \tag{2.18}$$

which can also be written in terms of conjugate moments as:

$$\begin{aligned}
\underline{b_1} &= \frac{\hat{\sigma}_{x,y}}{\hat{\sigma}_x^2} \\
\underline{b_0} &= \underline{\hat{\mu}_y} - \underline{b_1}\underline{\hat{\mu}_x}.
\end{aligned} \tag{2.19}$$

or after expanding the moments (based on formulae from Section 1.2.2) as:

$$\underline{b_1} = \frac{\hat{\sigma}_{x_r,y_r} + \hat{\sigma}_{x_i,y_i} + i(\hat{\sigma}_{x_r,y_i} - \hat{\sigma}_{x_i,y_r})}{\hat{\sigma}_{x_r}^2 + \hat{\sigma}_{x_i}^2} \tag{2.20}$$

As we can see, the formula (2.19) is similar to the one that is typically used for the conventional simple linear regression with the only difference that the parameters in (2.19) are complex and that each of the moments in (2.19) is a moment for respective complex variable.

It is apparent what the estimates of parameters (2.19) correspond to: they minimise the loss (2.13), thus in the case, when $\mathrm{E}(\underline{e_j}) = 0$ they minimise variances of real and imaginary parts of the complex residuals. A thing to note is that they ignore the potential covariance between them, which in some cases might be a desirable property, but in the others might cause issues.

In terms of properties of the estimator, it can be shown that $\underline{b_1}$ equals to:

$$\underline{b_1} = \frac{\mathrm{cov}(\underline{\tilde{x}}, \underline{y})}{\mathrm{cov}(\underline{\tilde{x}}, \underline{x})} = \frac{\mathrm{cov}(\underline{\tilde{x}}, \underline{\beta_0} + \underline{\beta_1}\underline{x} + \underline{\epsilon})}{\mathrm{cov}(\underline{\tilde{x}}, \underline{x})} = \frac{\mathrm{cov}(\underline{\tilde{x}}, \underline{\beta_0}) + \mathrm{cov}(\underline{\tilde{x}}, \underline{\beta_1}\underline{x}) + \mathrm{cov}(\underline{\tilde{x}}, \underline{\epsilon})}{\mathrm{cov}(\underline{\tilde{x}}, \underline{x})}, \tag{2.21}$$

which then simplifies to:

$$\underline{b_1} = \underline{\beta_1} + \frac{\mathrm{cov}(\underline{\tilde{x}}, \underline{\epsilon})}{\mathrm{cov}(\underline{\tilde{x}}, \underline{x})} = \underline{\beta_1} + \frac{\hat{\sigma}_{x,\epsilon}}{\hat{\sigma}_x^2}, \tag{2.22}$$

or:

$$\underline{b_1} = \underline{\beta_1} + \frac{\hat{\sigma}_{x_r,\epsilon_r} + \hat{\sigma}_{x_i,\epsilon_i} + i(\hat{\sigma}_{x_r,\epsilon_i} - \hat{\sigma}_{x_i,\epsilon_r})}{\hat{\sigma}_{x_r}^2 + \hat{\sigma}_{x_i}^2}. \tag{2.23}$$

On small samples, the covariance between the true error and the available $\underline{x}$ can be not equal to zero, implying that the value of the estimated parameter would differ from the true one. If the basic regression assumptions hold, in the population estimated covariances will converge to their true values and then $\sigma_{x_r,\epsilon_r} = \sigma_{x_i,\epsilon_i} = \sigma_{x_r,\epsilon_i} = \sigma_{x_i,\epsilon_r} = 0$, which implies that the expectation of $\underline{b_1}$ is:

$$\mathrm{E}(\underline{b_1}) = \underline{\beta_1}. \tag{2.24}$$

This implies that OLS gives unbiased estimates of the slope parameter. The same property can be shown to hold for the intercept $\underline{b_0}$.

The same formula (2.23) can be rewritten separately for the real and imaginary parts of the parameter:

$$\begin{aligned} b_{1,r} =& \beta_{1,r} + \frac{\hat{\sigma}_{x_r,\epsilon_r} + \hat{\sigma}_{x_i,\epsilon_i}}{\hat{\sigma}_{x_r}^2 + \hat{\sigma}_{x_i}^2} \\ b_{1,i} =& \beta_{1,i} + \frac{\hat{\sigma}_{x_r,\epsilon_i} - \hat{\sigma}_{x_i,\epsilon_r}}{\hat{\sigma}_{x_r}^2 + \hat{\sigma}_{x_i}^2}, \end{aligned} \tag{2.25}$$

which shows what specific covariances impact different parts of the slope parameter.

As for the variance of $\underline{b_1}$, for the real part of the parameter, it can be shown to be equal to:

$$\mathrm{V}(b_{1,r}) = \mathrm{V}\left(\frac{\hat{\sigma}_{x_r,\epsilon_r} + \hat{\sigma}_{x_i,\epsilon_i}}{\hat{\sigma}_{x_r}^2 + \hat{\sigma}_{x_i}^2}\right). \tag{2.26}$$

The variance for the imaginary part will be similar. We do not expand it further, because the formula (2.26) is sufficient for the comparison of OLS with other estimators.

2.2.2 Complex Least Squares

An alternative estimation technique involves the minimisation of a rather exotic loss function:

$$\sum_{j=1}^{n}(\underline{e_j e_j}) = \sum_{j=1}^{n}(e_{r,j}^2 - e_{i,j}^2 + i2e_{r,j}e_{i,j}), \tag{2.27}$$

for which in case of $\mathrm{E}(\underline{e_j}) = 0$, the real part corresponds to the difference between the variances of the real and imaginary parts of the residuals, while the imaginary parts corresponds to the covariance between them. This loss be considered exotic, because its value is a complex number. It is difficult to explain how one can minimise a complex number, but from what follows, we show that the estimation technique has some meaning, works and gives adequate estimates of parameters.

The logic for the derivation of CLS is similar to OLS. We expand (2.27) to:

$$\sum_{j=1}^{n}(\underline{e_j e_j}) = \sum_{j=1}^{n}\underline{y_j}^2 + n\underline{b_0}^2 + \underline{b_1}^2\sum_{j=1}^{n}\underline{x_j}^2 - 2\underline{b_0}\sum_{j=1}^{n}\underline{y_j} - 2\underline{b_1}\sum_{j=1}^{n}\underline{x_j y_j} + 2\underline{b_0}\underline{b_1}\sum_{j=1}^{n}\underline{x_j} \tag{2.28}$$

and then take derivatives of (2.28) with respect to parameter $\underline{b_0}$ and $\underline{b_1}$ and then equate the resulting equations to zero:

$$\begin{aligned} \frac{d\sum_{j=1}^{n}(\underline{e_j}^2)}{d\underline{b_0}} &= 0 \\ \frac{d\sum_{j=1}^{n}(\underline{e_j}^2)}{d\underline{b_1}} &= 0. \end{aligned} \tag{2.29}$$

The derivatives (2.29) give the following system of complex equations:

$$\begin{aligned} 2n\underline{b_0} - 2\sum_{j=1}^{n}\underline{y_j} + 2\underline{b_1}\sum_{j=1}^{n}\underline{x_j} &= 0 \\ 2\underline{b_1}\sum_{j=1}^{n}\underline{x_j}^2 - 2\sum_{j=1}^{n}\underline{x_j y_j} + 2\underline{b_0}\sum_{j=1}^{n}\underline{x_j} &= 0. \end{aligned} \tag{2.30}$$

The solution for this system of equations, as it was shown by **?**, is:

$$\begin{aligned} \underline{b_1} &= \frac{\sum_{j=1}^{n}(\underline{y_j} - \underline{\hat{\mu}_y})(\underline{x_j} - \underline{\hat{\mu}_x})}{\sum_{j=1}^{n}(\underline{x_j} - \underline{\hat{\mu}_x})^2} \\ \underline{b_0} &= \frac{1}{n}\sum_{j=1}^{n}\underline{y_j} - \underline{b_1}\sum_{j=1}^{n}\underline{x_j}, \end{aligned} \tag{2.31}$$

or in terms of direct moments for complex random variables:

$$\begin{aligned}\underline{b_1} &= \frac{\hat{\varsigma}_{x,y}}{\hat{\varsigma}_x^2} \\ \underline{b_0} &= \underline{\hat{\mu}_y} - \underline{b_1}\underline{\hat{\mu}_x},\end{aligned} \tag{2.32}$$

or after inserting the values for direct variance and covariance (from Section 1.2.2):

$$\begin{aligned}\underline{b_1} &= \frac{\hat{\sigma}_{x_r,y_r} - \hat{\sigma}_{x_i,y_i} + i(\hat{\sigma}_{x_i,y_r} + \hat{\sigma}_{x_r,y_i})}{\hat{\sigma}_{x_r}^2 - \hat{\sigma}_{x_i}^2 + i2\hat{\sigma}_{x_r,x_i}} \\ \underline{b_0} &= \underline{\hat{\mu}_y} - \underline{b_1}\underline{\hat{\mu}_x}.\end{aligned} \tag{2.33}$$

Similarly to how it was done with OLS estimate of the slope, we can consider the estimate of $\underline{b_1}$, expanding the covariance in the numerator of (2.32):

$$\underline{b_1} = \frac{\mathrm{cov}(\underline{x}, \underline{y})}{V(\underline{x})} = \frac{\mathrm{cov}(\underline{x}, \underline{\beta_0} + \underline{\beta_1}\underline{x} + \underline{\epsilon})}{V(\underline{x})}, \tag{2.34}$$

which after some simplifications becomes:

$$\underline{b_1} = \underline{\beta_1} + \frac{\mathrm{cov}(\underline{x}, \underline{\epsilon})}{V(\underline{x})}, \tag{2.35}$$

which can be expanded to:

$$\underline{b_1} = \underline{\beta_1} + \frac{\mathrm{cov}(x_r, \epsilon_r) - \mathrm{cov}(x_i, \epsilon_i) + i(\mathrm{cov}(x_r, \epsilon_i) + \mathrm{cov}(x_i, \epsilon_r))}{V(x_r) - V(x_i) + 2i\mathrm{cov}(x_r, x_i)}. \tag{2.36}$$

Note that both numerator and denominator of (2.36) are complex numbers. In order to have a proper split into real and imaginary parts we need to multiply the fraction by the complex number conjugate to the denominator:

$$\underline{b_1} = \underline{\beta_1} + \frac{(\mathrm{cov}(x_r,\epsilon_r)-\mathrm{cov}(x_i,\epsilon_i)+i(\mathrm{cov}(x_r,\epsilon_i)+\mathrm{cov}(x_i,\epsilon_r)))(V(x_r)-V(x_i)-2i\mathrm{cov}(x_r,x_i))}{(V(x_r)-V(x_i))^2+4\mathrm{cov}(x_r,x_i)^2}. \tag{2.37}$$

This then can be split into two parts:

$$\begin{aligned}b_{1,r} =& \beta_{1,r} + \frac{(\mathrm{cov}(x_r, \epsilon_r) - \mathrm{cov}(x_i, \epsilon_i))\,(V(x_r) - V(x_i)) - 2\mathrm{cov}(x_r, x_i)\,(\mathrm{cov}(x_r, \epsilon_i) + \mathrm{cov}(x_i, \epsilon_r))}{(V(x_r) - V(x_i))^2 + 4\mathrm{cov}(x_r, x_i)^2} \\ b_{1,i} =& \beta_{1,i} + \frac{(\mathrm{cov}(x_r, \epsilon_i) + \mathrm{cov}(x_i, \epsilon_r))\,(V(x_r) - V(x_i)) + 2\mathrm{cov}(x_r, x_i)\,(\mathrm{cov}(x_i, \epsilon_i) - \mathrm{cov}(x_r, \epsilon_r))}{(V(x_r) - V(x_i))^2 + 4\mathrm{cov}(x_r, x_i)^2}\end{aligned} \tag{2.38}$$

or

$$\begin{aligned}b_{1,r} =& \beta_{1,r} + \frac{(\hat{\sigma}_{x_r,\epsilon_r} - \hat{\sigma}_{x_i,\epsilon_i})\left(\hat{\sigma}_{x_r}^2 - \hat{\sigma}_{x_i}^2\right) - 2\hat{\sigma}_{x_r,x_i}\left(\hat{\sigma}_{x_r,\epsilon_i} + \hat{\sigma}_{x_i,\epsilon_r}\right)}{\left(\hat{\sigma}_{x_r}^2 - \hat{\sigma}_{x_i}^2\right)^2 + 4\hat{\sigma}_{x_r,x_i}^2} \\ b_{1,i} =& \beta_{1,i} + \frac{(\hat{\sigma}_{x_r,\epsilon_i} + \hat{\sigma}_{x_i,\epsilon_r})\left(\hat{\sigma}_{x_r}^2 - \hat{\sigma}_{x_i}^2\right) + 2\hat{\sigma}_{x_r,x_i}\left(\hat{\sigma}_{x_i,\epsilon_i} - \hat{\sigma}_{x_r,\epsilon_r}\right)}{\left(\hat{\sigma}_{x_r}^2 - \hat{\sigma}_{x_i}^2\right)^2 + 4\hat{\sigma}_{x_r,x_i}^2}\end{aligned} \tag{2.39}$$

If we now consider the properties of $b_{1,r}$, we can see that it is an asymptotically unbiased estimate of $\beta_{1,r}$ when the basic regression assumptions hold (i.e. $\sigma_{x_r,\epsilon_r} = \sigma_{x_i,\epsilon_i} = \sigma_{x_r,\epsilon_i} = \sigma_{x_i,\epsilon_r} = 0$). Similar holds for the $b_{1,i}$.

The variance of $b_{1,r}$ can be written as:

$$\mathrm{V}(b_{1,r}) = \mathrm{V}\left(\frac{\left(\hat{\sigma}_{x_r,\epsilon_r} - \hat{\sigma}_{x_i,\epsilon_i}\right)\left(\hat{\sigma}^2_{x_r} - \hat{\sigma}^2_{x_i}\right) - 2\hat{\sigma}_{x_r,x_i}\left(\hat{\sigma}_{x_r,\epsilon_i} + \hat{\sigma}_{x_i,\epsilon_r}\right)}{\left(\hat{\sigma}^2_{x_r} - \hat{\sigma}^2_{x_i}\right)^2 + 4\hat{\sigma}^2_{x_r,x_i}}\right). \quad (2.40)$$

We will come back to this variance, when comparing the efficiency of OLS and CLS estimators in Section 2.3.

In order to better understand what is specifically minimised, when the formulae (2.32) are used for the estimation of parameters, we conduct a small experiment in R with the following code:

```
# Create real part of a c.r.v. x
xr <- rnorm(1000,0,10)
# Create a c.r.v. x
x <- complex(real=xr, imaginary=1.5*xr+rnorm(1000,0,10))
# Create a c.r.v. y
y <- (1.5 + 1.2i) * x +
    complex(real=rnorm(1000,0,10), imaginary=rnorm(1000,0,10))

# Define number of iterations and the matrix with the values
nsim <- 10000
clsValues <- matrix(NA, nsim, 4,
                    dimnames=list(NULL,
                                  c("b1r","b1i","CLSr","CLSi")))

# CLS loss function
clsLoss <- function(y, yHat){
    return(sum((y - yHat)^2))
}

# Loop for values of b1r from 0.52 to 2.5 and
# for b1i from 0.22 to 2.2
l <- 1
for(i in 1:(nsim/100)){
    for(j in 1:(nsim/100)){
        b <- complex(real=i/100*2+0.5, imaginary=j/100*2+0.2)
        yHat <- (10 + 15i) + b * x
        clsResult <- clsLoss(y, yHat)
        clsValues[l,1] <- Re(b);
        clsValues[l,2] <- Im(b);
        clsValues[l,3] <- Re(clsResult);
        clsValues[l,4] <- Im(clsResult);
```

```
        l <- l+1;
    }
}

# Estimate b1 using CLS
bOptimal <- ccov(x, y, method="direct") /
            cvar(x, method="direct")
# Produce fitted values
yHat <- bOptimal * x
# Calculate the loss
clsResult <- clsLoss(y, yHat)
```

In this experiment we generate complex variables x and y and then calculate values of the complex loss function based on a combination of values of the complex slope $\underline{b_1}$ (which we call just `b` in the code). After running the code above we end up with 10,000 values of the loss function and one additional, which corresponds to the optimal point according to (2.32). We can then produce several plots to better understand what is happening.

```
plot(clsValues[,3:4], type="p",
     xlab="Re(CLS loss)", ylab="Im(CLS loss)")
abline(h=0, col="grey", lty=2, lwd=2)
abline(v=0, col="grey", lty=2, lwd=2)
points(Re(clsResult), Im(clsResult), col="red", pch=19)
```

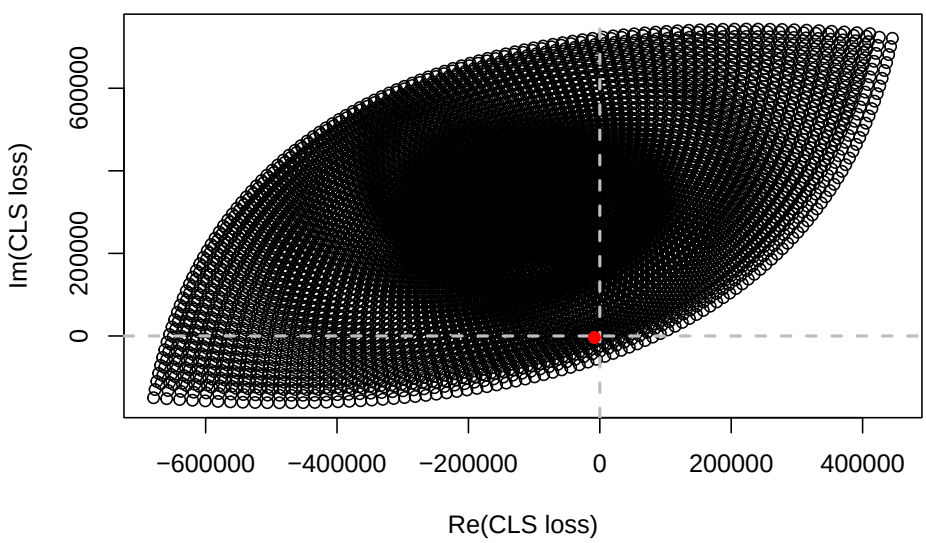

Figure 2.1: Variety of CLS loss function values for different values of $\underline{b_1}$.

Figure 2.1 shows a scatterplot of values of the complex loss (2.27). The red point

close to the origin corresponds to the estimate obtained using (2.32). As we can see, it is close to the origin, which implies that the minimisation of the loss (2.27) is equivalent to making both real and imaginary parts of it close to zero. This means that CLS makes variances of real and imaginary residuals similar and shrinks the covariance between them to zero.

Another way of looking at how CLS works is to reduce its dimensionality via the Multidimensional Scaling (**?**) and then visualise it. The R code below is relatively slow but produces a solution for this task.

```
# Calculate distance matrix for all losses
# (including the CLS point)
clsMDSDistance <- dist(rbind(clsValues[,3:4],
                             c(Re(clsResult),Im(clsResult))))

# Do Multidimensional scaling
clsMDS <- cmdscale(clsMDSDistance, k=1)

# Create a data frame with the coordinates
clsMDSValues <- data.frame(z=clsMDS,
                           x=c(clsValues[,1],Re(bOptimal)),
                           y=c(clsValues[,2],Im(bOptimal)))

# Create a matrix with loss values for 3d plotting
clsMDSValuesZ <- matrix(clsMDSValues$z[-10001], 100, 100)
```

Then, to produce the 3d surface on the plane of "loss - $b_{1,r}$ - $b_{1,i}$, we can use functions from the `plotly` package in R as shown in the code below:

```
plot_ly(z=clsMDSValuesZ,
        x=unique(clsMDSValues$x[-10001]),
        y=unique(clsMDSValues$y[-10001])) |>
    plotly::layout(scene=list(xaxis = list(title = "Re(b1)"),
                      yaxis = list(title = "Im(b1)"),
                      zaxis = list(title = "MDS of CLS loss"))) |>
    add_surface() |>
    add_trace(data = tail(clsMDSValues,1),
              x = ~x,
              y = ~y,
              z = ~z,
              mode = "markers",
              type = "scatter3d",
              marker = list(size = 10))
```

Figure 2.2 demonstrates how the projection of the two-dimensional loss behaves for a variety of parameters of the regression. The orange point in the middle corresponds to the one obtained via CLS. We can see that it corresponds to the loss being close to zero, and represents a 2-dimensional inflection point, where

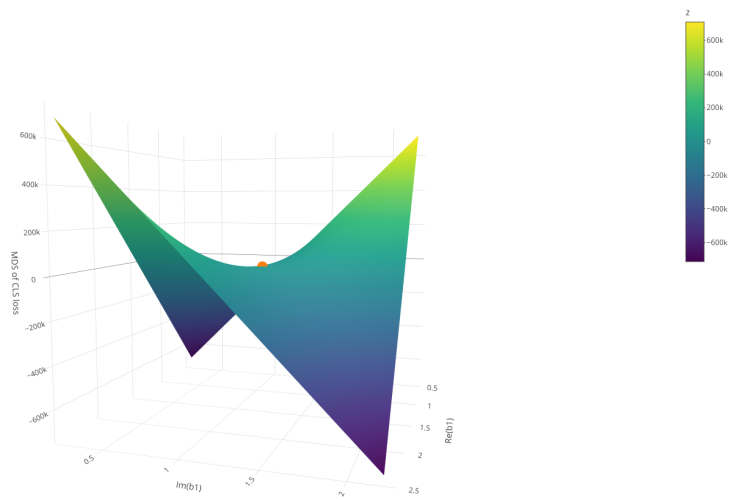

Figure 2.2: Plot of the 3d surface for the MDS of the CLS loss.

the surface bends in different directions.

So, we can see that despite being counter-intuitive, the CLS produces the estimates of parameters that lead to the residuals being close to spherical (normal if we assume normality), because the variances of real and imaginary parts of the residuals become closer to each other, while the covariance between them becomes close to zero.

2.2.3 Likelihood

Finally, another way of estimating the simple CLR is by assuming a distribution of the residuals and maximising the respective likelihood. Given the connection between the linear and the vector forms of the complex regression, we can assume that the error term follows a multivariate normal distribution with a covariance matrix:

$$\boldsymbol{\Sigma}_\epsilon = \begin{pmatrix} \sigma^2_{\epsilon_r} & \sigma_{\epsilon_r,\epsilon_i} \\ \sigma_{\epsilon_r,\epsilon_i} & \sigma^2_{\epsilon_i} \end{pmatrix}. \tag{2.41}$$

The log-likelihood in this case can be written as:

$$\ell(\boldsymbol{\theta}, \boldsymbol{\Sigma}_\epsilon|\mathbf{Y}) = -\frac{n}{2}\left(2\log(2\pi) + \log|\boldsymbol{\Sigma}_\epsilon|\right) - \frac{1}{2}\sum_{j=1}^{n}\left(\boldsymbol{\epsilon}_j'\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{\epsilon}_j\right), \tag{2.42}$$

where $\boldsymbol{\theta}$ is the vector of estimated parameters and $\boldsymbol{\epsilon}_j = \begin{pmatrix} \epsilon_{r,j} \\ \epsilon_{i,j} \end{pmatrix}$ is the two dimensional error term. This likelihood is maximised when the covariance matrix

(2.41) is estimated via:

$$\hat{\boldsymbol{\Sigma}}_\epsilon = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{e}_j \boldsymbol{e}_j'. \tag{2.43}$$

It can be shown that if the (2.43) is inserted in (2.42) then the following concentrated log-likelihood can be obtained (see, for example, **?**):

$$\ell^*(\boldsymbol{\theta}, \hat{\boldsymbol{\Sigma}}_\epsilon|\mathbf{Y}) = -\frac{n}{2}\left(2\log(2\pi e) + \log|\hat{\boldsymbol{\Sigma}}_\epsilon|\right). \tag{2.44}$$

It is obvious that the maximisation of the log-likelihood (2.44) is equivalent to minimising the generalised variance of the complex error term (as discussed in Subsection 1.2.2):

$$\text{GV} = |\hat{\boldsymbol{\Sigma}}_\epsilon| = \sigma^2_{\epsilon_r}\sigma^2_{\epsilon_i} - \sigma^2_{\epsilon_r,\epsilon_i}. \tag{2.45}$$

The minimisation of GV in its turn implies the joint minimisation of variances of the real and imaginary part of the residuals and a maximisation of the square of the covariance between them, thus making the residuals more linearly related. This loss function might be especially useful if we indeed can assume that the real and imaginary part of the residuals are linearly related and we want to use this information in the estimation.

The main difference between the Likelihood and the other two estimators discussed in this Section is that the former can only be maximised via a numeric optimisation - there is no analytical solution for estimates of parameters via likelihood.

2.3 Comparing different estimators for SCLR

Arguably, all three estimators discussed in this Section give adequate estimates of parameters, but inevitably they will have different efficiency and would be appropriate in different situations. In this subsection, we will explore the performance of estimators with the increase of sample size.

First, comparing variances of OLS and CLS estimates of $b_{1,r}$ (2.40) with (2.26):

$$\begin{aligned}
\text{V}(b^{\text{OLS}}_{1,r}) =& \text{V}\left(\frac{\hat{\sigma}_{x_r,\epsilon_r} + \hat{\sigma}_{x_i,\epsilon_i}}{\hat{\sigma}^2_{x_r} + \hat{\sigma}^2_{x_i}}\right) \\
\text{V}(b^{\text{CLS}}_{1,r}) =& \text{V}\left(\frac{\left(\hat{\sigma}_{x_r,\epsilon_r} - \hat{\sigma}_{x_i,\epsilon_i}\right)\left(\hat{\sigma}^2_{x_r} - \hat{\sigma}^2_{x_i}\right) - 2\hat{\sigma}_{x_r,x_i}\left(\hat{\sigma}_{x_r,\epsilon_i} + \hat{\sigma}_{x_i,\epsilon_r}\right)}{\left(\hat{\sigma}^2_{x_r} - \hat{\sigma}^2_{x_i}\right)^2 + 4\hat{\sigma}^2_{x_r,x_i}}\right)
\end{aligned}$$

we can conclude that there are some situations, when the estimates of parameters via CLS are less efficient than the estimates via OLS. There are some cases, when the variance (2.40) is much higher than (2.26). For example, when the real and imaginary parts of x are not correlated (i.e. $\hat{\sigma}_{x_r,x_i} = 0$) and when the variances of the real and imaginary parts of x are equal, the variance of $b_{1,r}$ explodes. However, there are also some cases, when CLS estimates are more efficient than

Table 2.1: Several scenarios for the comparison of estimators.

	cor(x_r, x_i)	std.dev. of x_r and x_i	cor(ϵ_r, ϵ_i)	std.dev. of ϵ_r and ϵ_i
Scenario 1	0	$\sigma_{x_r} = 10$, $\sigma_{x_i} = 20$	0	$\sigma_{\epsilon_r} = \sigma_{\epsilon_i} = 1.5$
Scenario 2	1	$\sigma_{x_r} = 10$, while $\sigma_{x_i} = 15$	0	$\sigma_{\epsilon_r} = \sigma_{\epsilon_i} = 1.5$
Scenario 3	0	$\sigma_{x_r} = \sigma_{x_i} = 20$	0	$\sigma_{\epsilon_r} = \sigma_{\epsilon_i} = 1.5$
Scenario 4	medium	$\sigma_{x_r} = 10$, $\sigma_{x_i} = 15$	NA	$\sigma_{\epsilon_r} = \sigma_{\epsilon_i} = 0$
Scenario 5	medium	$\sigma_{x_r} = 10$, $\sigma_{x_i} = 15$	medium	$\sigma_{\epsilon_r} = 10$, $\sigma_{\epsilon_i} = 8$
Scenario 6	medium	$\sigma_{x_r} = 10$, $\sigma_{x_i} = 15$	0	$\sigma_{\epsilon_r} = \sigma_{\epsilon_i} = 100$

the OLS ones. For example, when $\hat{\sigma}_{x_r,\epsilon_i} = \hat{\sigma}_{x_i,\epsilon_i}$ and $\hat{\sigma}_{x_r,x_i} = 0$, the variance (2.26) will be equal to zero, while the variance (2.40) will be greater than zero. So, in general, we cannot conclude which of the estimators will be more efficient, but we know that in different circumstances, they will have different properties.

We cannot compare the efficiency of the likelihood estimator directly with the OLS and CLS, but we know from the statistics literature (**?**) that likelihood gives consistent and asymptotically efficient estimates of parameters.

In terms of consistency, in one specific situation, when $\hat{\sigma}^2_{x_r} = \hat{\sigma}^2_{x_i}$ and $\hat{\sigma}_{x_r,x_i} = 0$, CLS might produce non-consistent estimates of parameters. This means that if we deal with explanatory variables with these properties, we should use either OLS or Likelihood.

Similar analysis can be done for the coefficient $b_{1,i}$, with the conclusions similar to the above, so we skip this discussion.

In order to do a more thorough comparison of the three estimators, we set up an experiment based on the following simple CLR:

$$y_r + iy_i = 10 + 15i + (2 - 1.5i)(x_r + ix_i) + (\epsilon_r + i\epsilon_i)$$

with several scenarios, parameters for which are shown in Table 2.1. They covered several important situations: when x_r and x_i are not correlated, have medium correlation and perfectly correlated, when their variances are similar or different and then when the real and imaginary parts of the error term are not correlated, have medium correlation, are equal to zero and when their variances have high or low values.

These six scenarios in Table 2.1 cover different theoretically possible situations and show how the three estimators behave in these conditions. The sample size was first set to 20 observations and then was increased iteratively by one observation until reaching 10,000. This should give us an understanding of how the estimators behave both on small samples and asymptotically. While we recorded both values of estimated intercept and slope, we are mainly interested in the latter, and the plots shown below focus on how the complex $\underline{b_1}$ is estimated. The experiments were done using `clm()` function from the `complex` package with `method` parameter equal to "OLS", "CLS" and "likelihood" for each of the

respective estimators. A sample of the script used in the experiments is shown below:

```
obs <- 10000
x0 <- rnorm(obs,10,10)
x <- complex(real=x0,imaginary=rnorm(obs,0,10))

b0 <- 10 + 15i
b1 <- 2-1.5i
y <- b0 + b1 * x + 1.5*complex(real=rnorm(obs,0,1),
                               imaginary=rnorm(obs,0,1))

complexData <- cbind(y,x)

nsim <- 9980
parametersValues <-
    array(NA, c(nsim,3,2),
          dimnames=list(NULL,
                        c("CLS","OLS","Likelihood"),
                        c("b0","b1")))

for(i in 1:nsim){
    test <- clm(y~x, complexData, loss="CLS",
                subset=sample(c(1:obs),20+i))
    parametersValues[i,1,] <- coef(test)
    test <- clm(y~x, complexData, loss="OLS",
                subset=sample(c(1:obs),20+i))
    parametersValues[i,2,] <- coef(test)
    test <- clm(y~x, complexData, loss="likelihood",
                subset=sample(c(1:obs),20+i))
    parametersValues[i,3,] <- coef(test)
}
```

Figure 2.3 demonstrates how estimates of parameters change with the increase of sample size in Scenario 1.

Apparently, all three estimators produce very similar estimates of parameters in this case and converge to the true values quite fast. A similar behaviour is observed for the Scenario 2, shown in Figure 2.4. The difference between the three estimators does not look substantial.

In both Scenarios 1 and 2 the error term has a small variance, in Scenario 2 x_r and x_i are perfectly correlated, transforming the original model into: $y_r + iy_i = 10 + 15i + (2 - 1.5i)(1 + 1.5i)x_r + (\epsilon_r + i\epsilon_i)$.

The Scenario 3 demonstrates an exotic case, when the variances of x_r and x_i are the same (as shown in Figure 2.5).

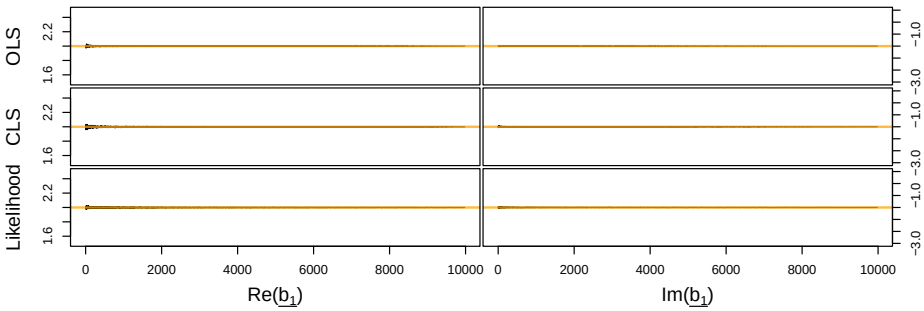

Figure 2.3: Estimation of parameters using OLS, CLS and likelihood and their convergence to the true value of $\underline{\beta_1} = 2 - 1.5i$ (orange horizontal lines on the plot). Scenario 1.

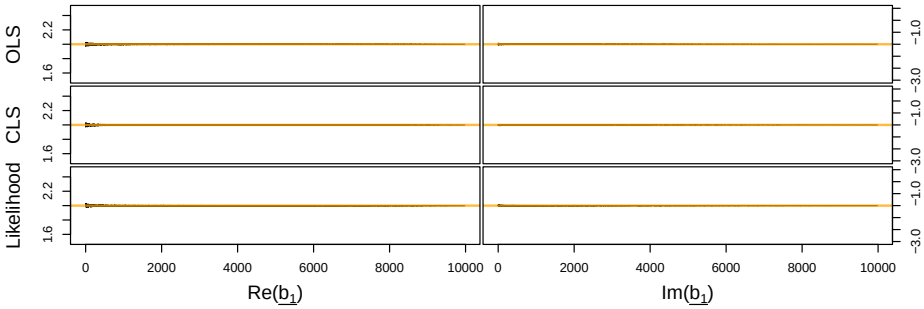

Figure 2.4: Estimation of parameters using OLS, CLS and likelihood and their convergence to the true value of $\underline{\beta_1} = 2 - 1.5i$ (orange horizontal lines on the plot). Scenario 2.

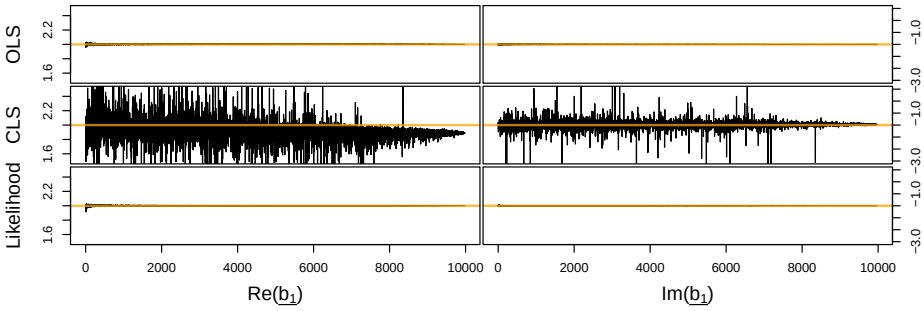

Figure 2.5: Estimation of parameters using OLS, CLS and likelihood and their convergence to the true value of $\underline{\beta_1} = 2 - 1.5i$ (orange horizontal lines on the plot). Scenario 3.

In this scenario, OLS and Likelihood produce efficient, unbiased and consistent estimates of parameters, which cannot be said about the CLS. The behaviour of CLS is explainable because for this specific scenario the direct variance of the complex variable $x_r + ix_i$ becomes close to zero. As a result, the direct covariance in (2.33) is divided by zero, and the estimate of the parameter becomes unstable. Scenarios 1 and 3 could be considered as two special cases of the spectrum of values, showing that the closer the variances of the real and imaginary parts of x are, the less consistent, efficient and unbiased estimates of parameters are produced by CLS. Scenario 2 is complimentary, because it shows that if the real and imaginary parts are correlated, the CLS estimates of parameters become as good (in statistical terms) as the estimates of OLS and/or Likelihood. So, the CLS becomes unreliable in a special case, when $\sigma^2_{x_r} = \sigma^2_{x_i}$ and $\sigma_{x_r,x_i} = 0$.

Next, the three estimators give the same estimates of parameters in Scenario 4 of functional relation between x and y, which is shown in Figure 2.6.

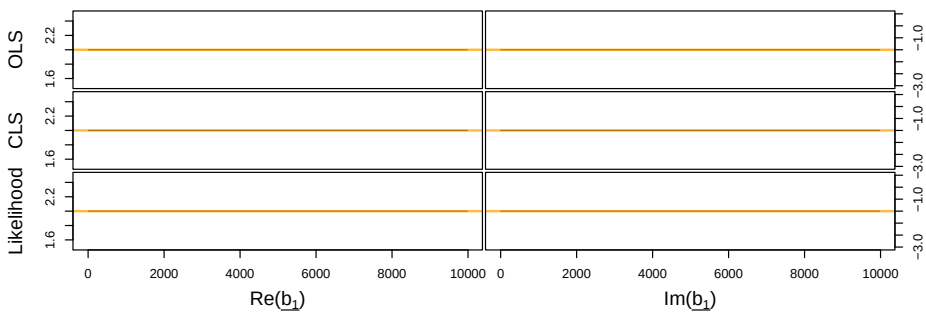

Figure 2.6: Estimation of parameters using OLS, CLS and likelihood and their convergence to the true value of $\underline{\beta_1} = 2 - 1.5i$ (orange horizontal lines on the plot). Scenario 4.

The main difficulties for the estimators appear, when the variance of the error term increases. Figure 2.7 shows how the three perform in case of a higher variance of the error term in Scenario 5.

It becomes apparent that OLS and Likelihood produce more efficient estimates of parameters than CLS on small samples. This is because the variability of estimates of parameter is higher for CLS than for the other two on small samples.

The situations worsens for the three estimators when we consider Scenario 6, when the variances of the error term become much higher than before, which is shown in Figure 2.8.

In addition to being less efficient, the real parts of estimates of parameters exhibit a bias, which is diminished with the increase of sample size. Comparing the three estimators, it appears that CLS produces less efficient and more biased estimates of parameters than OLS or Likelihood in this Scenario.

These six scenarios show a variety of situations in which the three estimators

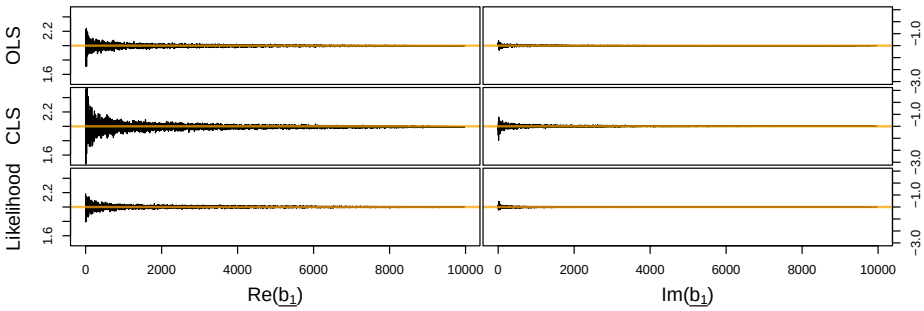

Figure 2.7: Estimation of parameters using OLS, CLS and likelihood and their convergence to the true value of $\underline{\beta_1} = 2 - 1.5i$ (orange horizontal lines on the plot). Scenario 5.

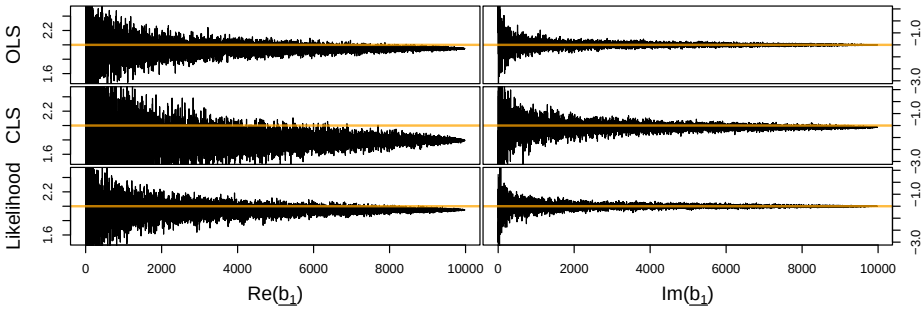

Figure 2.8: Estimation of parameters using OLS, CLS and likelihood and their convergence to the true value of $\underline{\beta_1} = 2 - 1.5i$ (orange horizontal lines on the plot). Scenario 6.

produce different estimates of parameters. Arguably, OLS and Likelihood could be considered as robust alternatives, but all the three estimators seem to perform well in several sensible situations.

Chapter 3

Correlation analysis of complex random variables

When it comes to measuring associations between variables, most frequently analysts use coefficient of correlation. While it is straightforward for real-valued variables, for complex variables this become challenging, because each c.r.v. has two parts, so the correlation needs to take them both into account.

For modelling purposes, it might be useful to have the information about all possible correlations involved relation of two c.r.v. This comes to analysing the following covariances for variables $\underline{x}$ and $\underline{y}$:

1. σ_{x_r,x_i},
2. σ_{y_r,y_i},
3. σ_{x_r,y_r},
4. σ_{x_i,y_i},
5. σ_{x_r,y_i},
6. σ_{y_r,x_i}.

3.1 Visualisation of relations

In order to better understand what correlations between c.r.v. imply, we need to understand how to produce scatterplots for them. While in case of two real variables it is straightforward (a variable per axes), in our situation, this becomes challenging. We propose considering a set of scatterplots shown in Figure 3.1 for two generated complex random variables $\underline{x}$ and $\underline{y}$, created using `cplot()` function from `complex` package in R.

```
# Create real part of a c.r.v. x
xr <- rnorm(1000,0,10)
# Create a c.r.v. x
```

```
x <- complex(real=xr, imaginary=1.5*xr+rnorm(1000,0,10))
# Create a c.r.v. y
y <- (10 + 15i) + (1.5 + 1.2i) * x +
    complex(real=rnorm(1000,0,10), imaginary=rnorm(1000,0,10))
# Produce the plot
cplot(x, y, which=1)
```

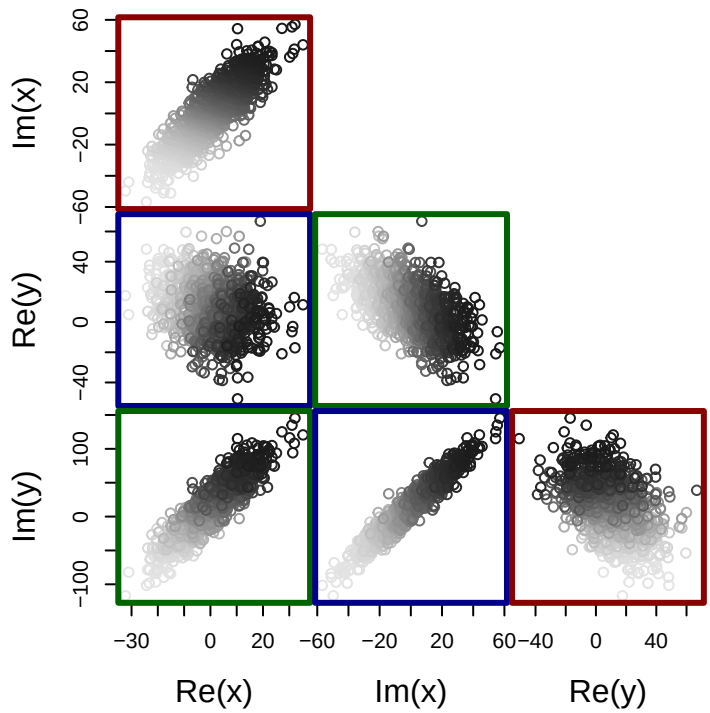

Figure 3.1: Visualisation of relations between two complex variables

This scatterplot has several important elements in it:

- It shows relations between real and imaginary parts of each variable (e.g. the two scatterplots in the bold dark red frame),
- It shows cross-relations between parts of one variable and parts of the other one (e.g. the rest four plots),
- The colour shows ordering of the original variable $\underline{x}$ with dark values corresponding to point with higher magnitude and light ones being closer to zero. This way, we can see what the original points in $\underline{x}$ correspond to in $\underline{y}$.

The plots are positioned to satisfy two rules:

1. When a scatterplot for a c.r.v. is produced, the real part should be in

x-axis, while the imaginary should be in the y-axis.

2. When parts of variables $\underline{x}$ and $\underline{y}$ are compared, the part for $\underline{x}$ should be in x-axis, while the part for $\underline{y}$ should be in y-axis, which should the reflect the idea that $\underline{x}$ could be an explanatory variable for $\underline{y}$.

While a simple scatterplot matrix could have been constructed instead of Figure 3.1, we argue that the latter has a logical grouping and should be preferred for analysis of complex variables. For example, based on the plots in Figure 3.1 we can conclude that:

- There is a positive relation between the real and imaginary parts of $\underline{x}$,
- There is a negative relation between the real and imaginary parts of $\underline{y}$,
- Real parts of $\underline{x}$ and $\underline{y}$ do not exhibit a strong linear relation,
- But the respective imaginary parts of $\underline{x}$ and $\underline{y}$ have the positive relation between them,
- Finally, we see that with the increase of real and imaginary parts of $\underline{x}$, the real part of $\underline{y}$ decreases, while the imaginary one increases. This sort of behaviour implies positive complex slope in the potential linear regression (discussed in Section 2).

We think that this visualisation is useful when analysing relations between two complex random variables. But it also shows how complicated it is to capture the relation between them and how many aspects need to be considered.

As an alternative to the plot above, it is also possible to use some dimensionality reduction techniques to plot complex variables on a two dimensional plot. For example, we can use Multidimensional Scaling for this (MDS, **?**) to create a project of one complex variable on x-axis and another one on the y-axis. In R, we can use the `cmdscale()` function from `stats` package for this (in the example below, we use euclidean distance for the dissimilarities matrix via `dist()` function from `stats`, and we use the `complex2vec()` function from `complex` package to transform complex variable to a collection of vectors):

```
xScaled <- cmdscale(dist(complex2vec(x)), k=1)
yScaled <- cmdscale(dist(complex2vec(y)), k=1)
plot(xScaled,yScaled)
```

The same code is implemented in `cplot()` function from `complex` package in R:

```
cplot(x, y, which=2, main="")
```

The plot in Figure 3.2 is much easier to read than the collection of scatterplots in Figure 3.1, and in our example, we can conclude that the two complex variables exhibit strong linear relation.

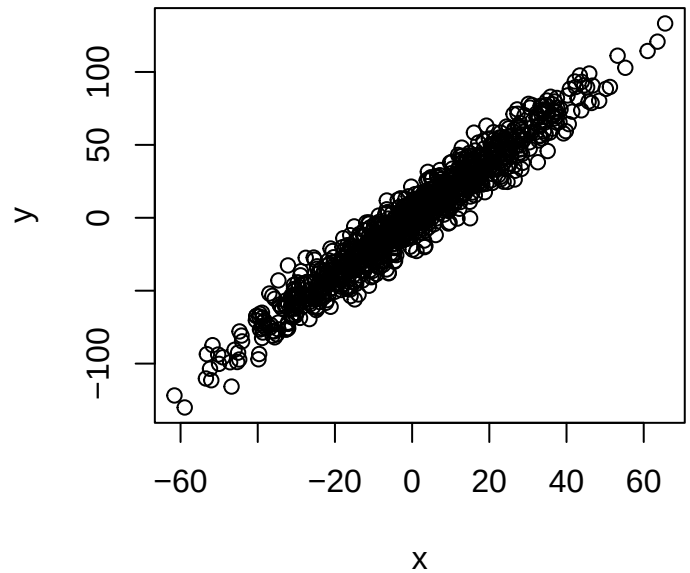

Figure 3.2: Scatterplot of MDS of two complex variables.

3.2 Types of correlation coefficients

The literature knows two correlation coefficients for complex variables (**?**): the conjugate and the direct correlation (the latter is known in the literature as "pseudo-correlation"). Their formula are based on the respective covariances and variances (conjugate and direct discussed in Subsection 1.2.2):

1. Conjugate correlation
$$\rho_{x,y} = \frac{\sqrt{\sigma_{x,y}\sigma_{y,x}}}{\sigma_x \sigma_y}, \tag{3.1}$$

2. Direct correlation
$$\varrho_{x,y} = \frac{\varsigma_{x,y}}{\varsigma_x \varsigma_y}. \tag{3.2}$$

Note that the conjugate correlation has the geometric mean of standard deviations in the numerator. This is needed because of the issue with the conjugate covariance (its value changes with the change of conjugate number). If we use only one of covariances (as done, for example, by **?**) then the value of correlation coefficient will be ambiguous, implying that the correlation between $\underline{x}$ and $\underline{y}$ differs from the correlation between $\underline{y}$ and $\underline{x}$. Furthermore, such correlation coefficient would not work as intended. For example, if we have a positive functional linear relation between $\underline{x}$ and $\underline{y}$, the coefficient should be equal to one. But as an R example below demonstrates this value is obtained only if we have the geometric mean of covariances.

```
x <- complex(real=rnorm(100,0,10), imaginary=rnorm(100,0,10))
y <- (10 + 15i) + (1 + 1i) * x
# Variant 1
ccov(y,x,method="conj") /
    sqrt(cvar(x,method="conj")*cvar(y,method="conj"))
```

```
## [1] 0.7071068-0.7071068i
```

```
# Variant 2
ccov(x,y,method="conj") /
    sqrt(cvar(x,method="conj")*cvar(y,method="conj"))
```

```
## [1] 0.7071068+0.7071068i
```

```
# Variant 3 (correct conjugate correlation)
sqrt(ccov(y,x,method="conj")*ccov(x,y,method="conj")) /
    sqrt((cvar(x,method="conj")*cvar(y,method="conj")))
```

```
## [1] 1+0i
```

The formulae (3.1) and (3.2) are derived based on the original definition of Pearson's correlation coefficient (**?**). It follows from the idea that the correlation coefficient equals to the geometric mean of slopes of two regression lines:

$$\begin{aligned} \underline{y} &= \underline{\beta_0} + \underline{\beta_1}\underline{x} + \underline{\epsilon} \\ \underline{x} &= \underline{\alpha_0} + \underline{\alpha_1}\underline{y} + \underline{\upsilon}, \end{aligned} \tag{3.3}$$

where $\underline{\alpha_0}$ and $\underline{\beta_0}$ are the intercepts, $\underline{\alpha_1}$ and $\underline{\beta_1}$ are the slopes of the regression lines and $\underline{\epsilon}$ and $\underline{\upsilon}$ are the residuals of the models. Note that all of these variables and parameters in our case are complex. As discussed in Section 2, the parameters of slope can be estimated using either Ordinary Least Squares, or the Complex Least Squares (discussed in Subsection 2.2). For the OLS, the formulae for the slopes are:

$$\begin{aligned} \underline{b_1} &= \frac{\hat{\sigma}_{x,y}}{\hat{\sigma}_x} \\ \underline{a_1} &= \frac{\hat{\sigma}_{y,x}}{\hat{\sigma}_y}. \end{aligned} \tag{3.4}$$

Taking their geometric means leads to:

$$\hat{\rho}_{x,y} = \sqrt{\underline{a_1}\underline{b_1}}, \tag{3.5}$$

which then leads to the formula (3.1). Similarly, for CLS estimated regressions, we have:

$$\begin{aligned} \underline{b_1} &= \frac{\hat{\varsigma}_{x,y}}{\hat{\varsigma}_x} \\ \underline{a_1} &= \frac{\hat{\varsigma}_{x,y}}{\hat{\varsigma}_y}, \end{aligned} \tag{3.6}$$

which after taking the same geometric means leads to (3.2). Note that the estimates of the slope parameters will differ between the OLS and the CLS and thus the direct and conjugate correlations will differ as well.

3.3 Conjugate correlation

When it comes to the interpretation of the coefficients, the conjugate one is a real number. It can be written as:

$$\rho_{x,y} = \frac{\sqrt{(\sigma_{x_r,y_r} + \sigma_{x_i,y_i})^2 + (\sigma_{x_i,y_r} - \sigma_{x_r,y_i})^2}}{\sqrt{(\sigma^2_{x_r} + \sigma^2_{x_i})(\sigma^2_{y_r} + \sigma^2_{y_i})}}. \tag{3.7}$$

As can be seen from the formula (3.7), the coefficient is a real number, showing the average strength of linear relation between two complex variables $\underline{x}$ and $\underline{y}$. The coefficient will be equal to zero, when there are no linear relations between the respective real and imaginary parts of complex variables $\underline{x}$ and $\underline{y}$ and when the cross-covariances are equal (i.e. $\sigma_{x_r,y_r} = \sigma_{x_i,y_i} = 0$ and $\sigma_{x_i,y_r} = \sigma_{x_r,y_i}$). A special case of this condition is when all the covariances are equal to zero. On the other hand, it will be close to one by absolute value if the complex relation between variables $\underline{y}$ and $\underline{x}$ is close to the linear, i.e. $\underline{a_1} = \frac{1}{\underline{b_1}}$.

In order to better understand what the conjugate correlation means, we expand the formula (3.5) by substituting $\underline{b_1} = b_{1,r} + ib_{1,i}$ and $\underline{a_1} = a_{1,r} + ia_{1,i}$:

$$\begin{aligned}\hat{\rho}_{x,y} =& \sqrt{(a_{1,r} + ia_{1,i})(b_{1,r} + ib_{1,i})} = \\ & \sqrt{a_{1,r}b_{1,r} - a_{1,i}b_{1,i} + i(a_{1,r}b_{1,i} + a_{1,i}b_{1,r})},\end{aligned} \tag{3.8}$$

or in the exponential form:

$$\hat{\rho}_{x,y} = R^{\frac{1}{2}} e^{i\frac{1}{2}\phi}, \tag{3.9}$$

where

$$\begin{aligned} R &= \sqrt{(a_{1,r}b_{1,r} - a_{1,i}b_{1,i})^2 + (a_{1,r}b_{1,i} + a_{1,i}b_{1,r})^2} \\ \phi &= \arctan\left(\frac{a_{1,r}b_{1,i} + a_{1,i}b_{1,r}}{a_{1,r}b_{1,r} - a_{1,i}b_{1,i}}\right), \end{aligned} \tag{3.10}$$

As can be seen from (3.10), there is a multitude of combinations of parameters of the model that can give the unity magnitude R. For example, if $a_{1,r} = 0.5$, $b_{1,r} = 1$, $a_{1,i} = -0.5$ and $b_{1,i} = 1$, R would be equal to one. Similarly, there is a multitude of values of parameters that would give angles of 0 and π, leading to positive or negative real numbers. In fact, for the example above, $\phi = 0$, implying that the correlation coefficient will be equal to one, implying that the variables $\underline{x}$ and $\underline{y}$ exhibit a strong linear relation.

An R example of a conjugate correlation (via `ccor()` function from `complex` package) with the aforementioned values of parameters is shown below and in Figure 3.3.

```
# Create a c.r.v. x
x <- complex(real=rnorm(100,0,10), imaginary=rnorm(100,0,10))
# Create a c.r.v. y
y <- (10 + 15i) + (1 + 1i) * x +
    complex(real=rnorm(100,0,1), imaginary=rnorm(100,0,1))
# Produce the plot
cplot(x, y, main="")
```

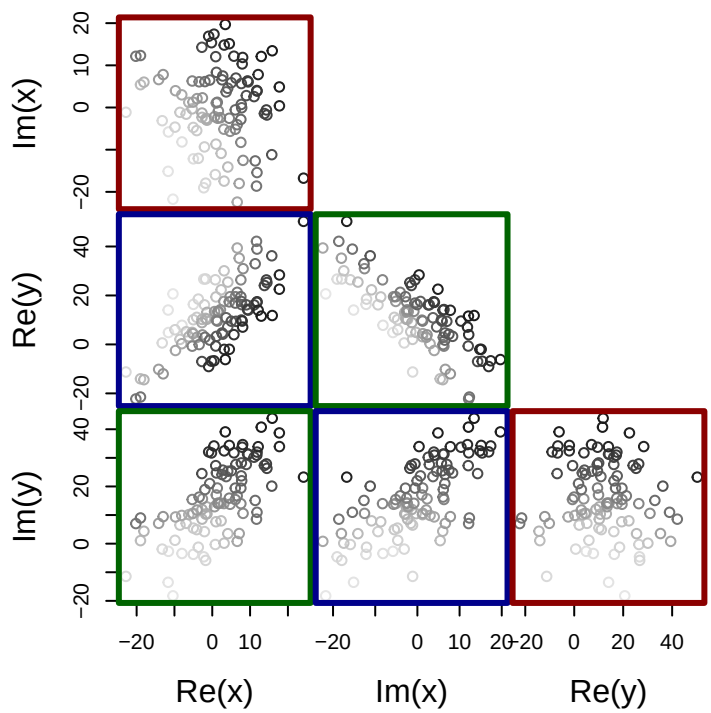

Figure 3.3: Two complex variables with conjugate correlation being close to one.

```
# Conjugate correlation
ccor(x, y, method="conjugate")
```

```
## [1] 0.9973417+0i
```

As can be seen from the Figure 3.3 and the value of the conjugate correlation, there is a relation between the two complex variables $\underline{x}$ and $\underline{y}$. The conjugate correlation says that this is a strong linear relation, and one of ways how we can check this is by analysing the MDS of the two variables.

```
cplot(x, y, which=2, main="")
```

As can be seen from the plot in Figure 3.4, it seems that the variables indeed exhibit a strong linear relation. So, the conjugate correlation has provided an

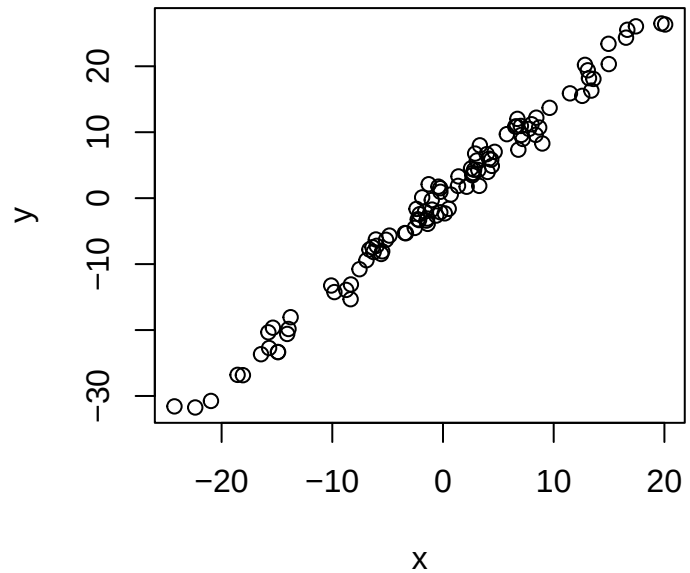

Figure 3.4: Visualisation of relations between two complex variables

adequate information about it.

3.4 Direct correlation

The direct correlation can be expanded to:

$$\varrho_{x,y} = \frac{\sigma_{x_r,y_r} - \sigma_{x_i,y_i} + i(\sigma_{x_i,y_r} + \sigma_{x_r,y_i})}{\sqrt{(\sigma^2_{x_r} - \sigma^2_{x_i} + i2\sigma_{x_r,x_i})(\sigma^2_{y_r} - \sigma^2_{y_i} + i2\sigma_{y_r,y_i})}}. \tag{3.11}$$

Given that it contains a complex number in the denominator, it is more complicated than the conjugate one. But it has several apparent properties:

1. The magnitude of the coefficient will be equal to zero (and thus the coefficient will be equal to zero) only when $\sigma_{x_i,y_r} = \sigma_{x_r,y_i} = 0$ and $\sigma_{x_r,y_r} = \sigma_{x_i,y_i}$. One of the special cases of this is when all cross-covariances between the real and imaginary parts of $\underline{x}$ and $\underline{y}$ are equal to zero.

2. When the complex variables $\underline{x}$ and $\underline{y}$ have functional relation between them, so that $\underline{a_1} = \frac{1}{\underline{b_1}}$, the coefficient will be equal to one.

Note however that due to the division by a complex number, the value of 1 can also be obtained due to the values of direct variance of variables $\underline{x}$ and/or $\underline{y}$.

Coefficient equal to -1 and to i.

3.5 Pearson's correlation

We can also use MDS to analyse the projections of complex variables on x and y axes, similar to how we have done that in Subsection 3.1. In that case, we can calculate Pearson's correlation coefficient:

```
# Create projections of two complex variables onto axes
xScaled <- cmdscale(dist(complex2vec(x)), k=1)
yScaled <- cmdscale(dist(complex2vec(y)), k=1)
# Calculate the correlation coefficient
cor(xScaled,yScaled)
```

```
##           [,1]
## [1,] 0.9932741
```

Or using the `ccor()` function with `method="pearson"`, which does exactly the same thing in one line of code:

```
ccor(x, y, method="pearson")
```

The interpretation of the coefficient is straightforward and follows the conventional interpretation taught in any Statistics module. Note however that in general in the optimisation phase of MDS, it can reach the local optimum and not being able to get to the global one. As a result, the sign of the correlation might not represent the real relation between the two complex variables and in general should be ignored. Furthermore, inevitably when we move from four dimensions to two, we loose some information, so this approach is prone to possible mistakes and should be used with care. Finally, MDS is computationally expensive, especially on the long series. This means that in some cases it might take plenty of time before we get the Pearson's correlation value. Nonetheless, this is an additional way of analysing relations between complex variables.

3.6 Correlation matrix

Finally, as discussed in Subsection 1.2.2, it is possible to calculate the covariance matrix between two c.r.v. Based on that matrix, we can calculate the correlation matrix, which will summarise all the relations between the real and imaginary parts of $\underline{x}$ and $\underline{y}$. This is done by dividing each element of the covariance matrix by geometric means of variances of the variables under consideration. In R, this can be done using `covar()` function from the `complex` package and `cov2cor()` function from the `stats` package:

```
cov2cor(covar(cbind(x,y)))
```

```
##             x_r         x_i         y_r         y_i
## x_r  1.00000000 -0.06718548  0.70576891  0.66047520
## x_i -0.06718548  1.00000000 -0.75071852  0.70117264
## y_r  0.70576891 -0.75071852  1.00000000 -0.06095876
```

```
## y_i  0.66047520  0.70117264 -0.06095876  1.00000000
```

This matrix will not tell us how the variables $\underline{x}$ and $\underline{y}$ are related overall, but it will contain information about each of their individual elements.

Chapter 4

Multiple Complex Linear Regression

We now move to the discussion of the multiple CLR, the model that captures relations between one complex random variable, $y_r + iy_i$ and a set of explanatory complex random variables.

4.1 Model formulation

Similarly to how the multiple linear regression is formulated for real valued variables, the multiple complex linear regression can be written as:

$$\underline{y_j} = \underline{\beta_0} + \underline{\beta_1}\underline{x_{1,j}} + \underline{\beta_2}\underline{x_{2,j}} + \cdots + \underline{\beta_k}\underline{x_{k,j}} + \underline{\epsilon_j}, \tag{4.1}$$

where k is the number of complex random variables. Similarly to how it was done with SCLR in (2.3), we can expand the formula (4.1) as a system of two equations, taking that every parameter and every variable in (4.1) is complex:

$$\begin{aligned} y_{r,j} =& \beta_{0,r} + \beta_{1,r}x_{1,r,j} - \beta_{1,i}x_{1,i,j} + \cdots + \beta_{k,r}x_{k,r,j} - \beta_{k,i}x_{k,i,j} + \epsilon_{r,j} \\ y_{i,j} =& \beta_{0,i} + \beta_{1,r}x_{1,i,j} + \beta_{1,i}x_{1,r,j} + \cdots + \beta_{k,r}x_{k,i,j} + \beta_{k,i}x_{k,r,j} + \epsilon_{i,j}. \end{aligned} \tag{4.2}$$

As can be seen from (4.2), the multiple CLR captures more complex dynamics than the conventional multiple linear regression. Both parts of the system use the same set of parameters and explanatory variables, but in different combinations, resulting in a versatile modelling framework.

This system can be represented in a more compact form, similarly to (2.4):

$$\underline{\mathbf{y}} = \underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\epsilon}}, \tag{4.3}$$

where now $\underline{\mathbf{X}} = \begin{pmatrix} 1 & \underline{x_{1,1}} & \dots & \underline{x_{k,1}} \\ 1 & \underline{x_{1,2}} & \dots & \underline{x_{k,2}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \underline{x_{1,n}} & \dots & \underline{x_{k,n}} \end{pmatrix}$ and $\underline{\boldsymbol{\beta}} = \begin{pmatrix} \underline{\beta_0} \\ \underline{\beta_1} \\ \vdots \\ \underline{\beta_k} \end{pmatrix}$, where each of the elements in the matrices and vectors above is a complex number.

Furthermore, the system (4.2) can also be used to represent the multiple CLR in a simple form using vector and matrix notations, avoiding complex numbers:

$$\mathbf{y}_j = \underset{\sim}{\mathbf{X}}_j \boldsymbol{\beta} + \boldsymbol{\epsilon}_j, \tag{4.4}$$

where $\mathbf{y}_j = \begin{pmatrix} y_{r,j} \\ y_{i,j} \end{pmatrix}$, $\underset{\sim}{\mathbf{X}}_j = \begin{pmatrix} 1 & 0 & x_{1,r,j} & -x_{1,i,j} & \dots & x_{k,r,j} & -x_{k,i,j} \\ 0 & 1 & x_{1,i,j} & x_{1,r,j} & \dots & x_{k,i,j} & x_{k,r,j} \end{pmatrix}$, $\boldsymbol{\beta}' = \begin{pmatrix} \beta_{0,r} & \beta_{0,i} & \beta_{1,r} & \beta_{1,i} & \dots & \beta_{1,k} & \beta_{1,k} \end{pmatrix}$ and $\epsilon_j = \begin{pmatrix} \epsilon_{r,j} \\ \epsilon_{i,j} \end{pmatrix}$. This can be then represented in the even more compact form, using the same principles as discussed in Section 2.1 in formula (2.8):

$$\mathbf{Y} = \underset{\sim}{\mathbf{X}} \boldsymbol{\beta} + \mathbf{E} \tag{4.5}$$

with where $\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}$, $\underset{\sim}{\mathbf{X}} = \begin{pmatrix} \underset{\sim}{\mathbf{X}}_1 \\ \underset{\sim}{\mathbf{X}}_2 \\ \vdots \\ \underset{\sim}{\mathbf{X}}_n \end{pmatrix}$ and $\mathbf{E} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_n \end{pmatrix}$. Formula (4.5) becomes especially useful for multiple CLR for the model estimation via OLS, CLS or Likelihood in the matrix form. The form (4.5) sidesteps complex numbers all together, representing the set of equations in matrices and vectors, containing real numbers only. This is convenient for many purposes and in inference.

The main difference between the form (4.3) and (4.5) is that the former contains complex numbers inside each of the matrices and vectors.

4.2 Estimation

In order to estimate the parameters of the model (4.3), we can use the same methods as in the Chapter 2: OLS, CLS and Likelihood. We will write the estimated model as:

$$\underline{\mathbf{y}} = \underline{\mathbf{X}\boldsymbol{b}} + \underline{\mathbf{e}}, \tag{4.6}$$

where $\underline{\boldsymbol{b}}$ is the estimate of $\underline{\boldsymbol{\beta}}$ and $\underline{\mathbf{e}}$ is the estimate of $\underline{\epsilon}$. And in case of matrix notations, instead of (4.5) we will have:

$$\mathbf{Y} = \underset{\sim}{\mathbf{X}} \boldsymbol{b} + \hat{\mathbf{E}} \tag{4.7}$$

4.2.1 Ordinary Least Squares

The criterion of OLS for multiple CLR can be written as:

$$\min S^{\mathrm{OLS}}(\underline{\boldsymbol{b}}) = \min\left(\underline{\mathbf{e}}'\underline{\mathbf{e}}\right), \tag{4.8}$$

which can be expanded to:

$$\begin{aligned} S^{\mathrm{OLS}}(\underline{\boldsymbol{b}}) = \left(\underline{\mathbf{y}} - \underline{\mathbf{X}\boldsymbol{b}}\right)' \left(\underline{\mathbf{y}} - \underline{\mathbf{X}\boldsymbol{b}}\right) = \\ \underline{\mathbf{y}}'\underline{\mathbf{y}} - \underline{\mathbf{y}}'\underline{\mathbf{X}\boldsymbol{b}} - \underline{\boldsymbol{b}}'\underline{\mathbf{X}}'\underline{\mathbf{y}} + \underline{\boldsymbol{b}}'\underline{\mathbf{X}}'\underline{\mathbf{X}\boldsymbol{b}} \end{aligned}. \tag{4.9}$$

Taking derivative of (4.9) with respect to $\underline{\boldsymbol{b}}$ and equating it to zero, results in the following system of normal equations:

$$\underline{\mathbf{X}}'\underline{\mathbf{X}\boldsymbol{b}} = \underline{\mathbf{X}}'\underline{\mathbf{y}}, \tag{4.10}$$

which gives the classical formula for the estimation of parameters of the model (4.5):

$$\underline{\boldsymbol{b}} = \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\mathbf{y}} \tag{4.11}$$

Given that (4.11) corresponds to the classical OLS, it will maintain all of its conventional properties, i.e. its estimates being unbiased, efficient and consistent. Note that, as discussed Subsection 1.1.1, the operator $'$ denotes conjugate transposition, which means that for the special case of simple CLR, the formula (4.11) will become (2.19).

Finally, using the same principles, we can show that the estimates of parameters can be obtained if we use the form (4.7) instead of the vectors of complex variables:

$$\boldsymbol{b} = \left(\underset{\sim}{\tilde{\mathbf{X}}}'\underset{\sim}{\mathbf{X}}\right)^{-1}\underset{\sim}{\tilde{\mathbf{X}}}'\mathbf{Y}. \tag{4.12}$$

The form (4.12) becomes especially useful for futher inference.

4.2.2 Complex Least Squares

As discussed in Section 2.2, there is also an alternative approach to estimation of CLR, the Complex Least Squares. In order to get the estimates based on it, we need to apply the same principles as with OLS, but directly to the form (4.3), i.e. minimise the criterion (which is the same as the one discussed in Subsection 2.2.2):

$$\min S^{\mathrm{CLS}}(\underline{\boldsymbol{b}}) = \min\left(\underline{\mathbf{e}}^{\top}\underline{\mathbf{e}}\right). \tag{4.13}$$

Using the same logic as with OLS, it can be shown that the minimisation of this criterion implies the solution of the following system of normal equations:

$$\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}\boldsymbol{b}} = \underline{\mathbf{X}}^{\top}\underline{\mathbf{y}}, \tag{4.14}$$

which then results in the following formula for the CLS estimate of parameters:

$$\underline{\boldsymbol{b}} = \left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^{\top}\underline{\mathbf{y}}. \tag{4.15}$$

For the simple CLR, the formula (4.15) becomes equivalent to (2.31). The main difference between (4.15) and (4.12), as we can see, is that in the CLS, the transposition is done without the conjugation.

Finally, based on the form (4.7), it can be shown that the same estimates (but in a form of real-valued vector) can be obtained via:

$$\boldsymbol{b} = \left(\underset{\sim}{\mathbf{X}}^\top \underset{\sim}{\mathbf{X}}\right)^{-1} \underset{\sim}{\mathbf{X}}^\top \mathbf{Y}. \tag{4.16}$$

The two formulae (4.15) and (4.16) result in exactly the same values of parameters, but will be useful for inference in the following sections.

4.2.3 Issues with OLS and CLS

Note that both OLS and CLS imply that the individual contributions of the real and imaginary parts of the error term are lost, and that the estimates of parameters are obtained for an overall variance of the complex error. In case of the OLS, this can be seen from the criterion (4.8), the minimisation of which is equivalent to the minimisation of the sum of variances:

$$\min S^{\mathrm{OLS}}(\underline{\boldsymbol{b}}) = \min\left(\underline{\mathbf{e}}'\underline{\mathbf{e}}\right) \iff \min\left(\hat{\sigma}^2_{e_r} + \hat{\sigma}^2_{e_i}\right), \tag{4.17}$$

where $\hat{\sigma}^2_{e_r}$ and $\hat{\sigma}^2_{e_i}$ are the variances of the real and imaginary parts of the error term respectively. The connection becomes apparent if we recall tha the main assumption of a regression model is that the expectation of the error term equals to zero. Because of that, the estimates of OLS lead to averaged out performance, ignoring the individual contribution of e_r and e_i and the covariance between the parts.

When it comes to CLS, the criterion (4.13) implies that:

$$\min S^{\mathrm{CLS}}(\underline{\boldsymbol{b}}) = \min\left(\underline{\mathbf{e}}^\top\underline{\mathbf{e}}\right) \iff \min\left(\hat{\sigma}^2_{e_r} - \hat{\sigma}^2_{e_i} + 2i\hat{\sigma}_{e_r,e_i}\right), \tag{4.18}$$

which now takes the covariance into account but ignores the sizes of the individual variances of the real and imaginary parts of the complex residuals.

In order to take the individual variances and the covariance into account, we need to use a different criterion and, as a result, a different estimator. One of such estimators is the Maximum Likelihood Estimator (MLE).

4.2.4 Likelihood

Similarly to how it was done in Subsection 2.2.3, we can make an assumption about the distribution of the error term of the CLR. The conventional one is that it follows a normal distribution. If we formulate the model in the vector form (4.4) then after being estimated it becomes:

$$\mathbf{y}_j = \underset{\sim}{\mathbf{X}}_j \boldsymbol{b} + \boldsymbol{e}_j. \tag{4.19}$$

The concentrated log-likelihood for the complex regression model (4.19) will be exactly the same as for the simple CLR:

$$\ell^*(\boldsymbol{\theta}, \hat{\boldsymbol{\Sigma}}_\epsilon|\mathbf{Y}) = -\frac{n}{2}\left(2\log(2\pi e) + \log|\hat{\boldsymbol{\Sigma}}_\epsilon|\right),$$

where (as a reminder)

$$\hat{\boldsymbol{\Sigma}}_\epsilon = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{e}_j \boldsymbol{e}_j'.$$

Maximising this likelihood, as discussed in Subsection 2.2.3, implies minimising Generalised Variance and guarantees that the estimates of parameters are efficient and consistent.

4.3 Inference

It is possible to calculate the variance of estimates of parameters, using the same approach as used in conventional OLS for the real-valued regression, substituting the formula for parameters with either (4.11) or (4.15). In fact, in case of c.r.v., it is possible to calculate both direct and conjugate covariance matrices of parameters. These variance can then be used in hypothesis testing or confidence interval construction. In order to do that, we need to replace the actual value $\underline{\mathbf{y}}$ with $\underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\epsilon}}$ in the OLS formula (4.11):

$$\begin{aligned}\underline{\boldsymbol{b}}^{\mathrm{OLS}} = &\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\mathbf{y}} = \\ &\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'(\underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\epsilon}}) = \\ &\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}} = \\ &\underline{\boldsymbol{\beta}} + \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}}.\end{aligned} \tag{4.20}$$

One thing that becomes apparent from this expansion is that the OLS estimates of parameters for a multiple complex linear regression will be unbiased as long as the expectation of the error term $\underline{\boldsymbol{\epsilon}}$ is zero and it is not correlated with the explanatory variables. It is easy to show that the same holds for the CLS estimates as well based on the following expanded formula:

$$\underline{\boldsymbol{b}}^{\mathrm{CLS}} = \underline{\boldsymbol{\beta}} + \left(\underline{\mathbf{X}}^\top\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^\top\underline{\boldsymbol{\epsilon}}. \tag{4.21}$$

This is a standard result from the real-valued regression analysis, but it is useful to know that it holds for the both estimation methods in complex-valued regression as well.

Based on (4.20), we can calculate the conjugate variance to get conjugate covariance matrix of parameters:

$$\mathrm{V}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) = \mathrm{V}\left(\underline{\boldsymbol{\beta}} + \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}}\right). \tag{4.22}$$

In the formula (4.22), the true estimates of parameters $\underline{\boldsymbol{\beta}}$ will be independent of the explanatory variables and the error term, so the formula can be represented as a sum of variances. Furthermore, the true parameters do not have any uncertainty, i.e. $\mathrm{V}\left(\underline{\boldsymbol{\beta}}\right) = 0$, meaning that the formula (4.22) can be transformed into:

$$\begin{aligned}\mathrm{V}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) =& \mathrm{V}\left(\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}}\right) = \\ & \mathrm{E}\left(\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}}\widetilde{\left(\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}}\right)}\right).\end{aligned} \tag{4.23}$$

We switch from the variance to the expectation in (4.23), dropping the expectation of the term in the brackets from the formula because it will be equal to zero as long as the explanatory variables and the error term are uncorrelated (one of the classical assumptions of a regression model). The tilde over the second term in (4.23) shows that this is the conjugate of the original complex variable. Recalling distributive properties of the conjugation (1.13), the same formula can be rewritten as:

$$\mathrm{V}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) = \mathrm{E}\left(\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}}\left(\underline{\tilde{\mathbf{X}}}'\underline{\tilde{\mathbf{X}}}\right)^{-1}\underline{\tilde{\mathbf{X}}}'\underline{\tilde{\boldsymbol{\epsilon}}}\right), \tag{4.24}$$

The expectation of the product in (4.24) can be rewritten as a product of expectations plus a covariance between the terms:

$$\begin{aligned}\mathrm{V}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) =& \mathrm{E}\left(\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}}\right)\mathrm{E}\left(\left(\underline{\tilde{\mathbf{X}}}'\underline{\tilde{\mathbf{X}}}\right)^{-1}\underline{\tilde{\mathbf{X}}}'\underline{\tilde{\boldsymbol{\epsilon}}}\right) + \\ & \mathrm{cov}\left(\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}}, \left(\underline{\tilde{\mathbf{X}}}'\underline{\tilde{\mathbf{X}}}\right)^{-1}\underline{\tilde{\mathbf{X}}}'\underline{\tilde{\boldsymbol{\epsilon}}}\right) = \\ & \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\tilde{\mathbf{X}}}\left(\underline{\tilde{\mathbf{X}}}'\underline{\tilde{\mathbf{X}}}\right)^{-1\prime}\mathrm{cov}\left(\underline{\boldsymbol{\epsilon}}, \underline{\tilde{\boldsymbol{\epsilon}}}\right) = \\ & \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\tilde{\mathbf{X}}}\left(\underline{\tilde{\mathbf{X}}}'\underline{\tilde{\mathbf{X}}}\right)^{-1\prime}\sigma^2_{\underline{\epsilon}},\end{aligned} \tag{4.25}$$

where the expectations of each term are equal to zero as long as the classical regression assumptions hold and $\sigma^2_{\underline{\epsilon}}$ is the conjugate variance of the error term. Using the same logic, it can be shown that the direct variance of parameters can be calculated as:

$$\mathcal{V}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) = \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\varsigma^2_{\underline{\epsilon}}. \tag{4.26}$$

Finally, in a similar fashion, conjugate and direct covariance matrices of parameters can be calculated for the CLS:

$$\mathrm{V}\left(\underline{\boldsymbol{b}}^{\mathrm{CLS}}\right) = \left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^{\top}\underline{\tilde{\mathbf{X}}}\left(\underline{\tilde{\mathbf{X}}}^{\top}\underline{\tilde{\mathbf{X}}}\right)^{-1\top}\sigma^2_{\underline{\epsilon}}. \tag{4.27}$$

and

$$\mathcal{V}\left(\underline{\boldsymbol{b}}^{\mathrm{CLS}}\right) = \left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\varsigma^2_{\underline{\epsilon}}. \tag{4.28}$$

for the CLS. Having both direct and conjugate variances for OLS and CLS not only allows analysing how the estimators will perform and which of them is more efficient, but also allows creating confidence intervals, testing hypotheses about the values of parameters and correctly capturing the uncertainty when constructing prediction interval from the estimated model. To construct the intervals or test hypotheses, we can use both direct and conjugate variances to calculate the covariance matrix $\hat{\mathbf{\Sigma}}$ and then use it in calculation of either Student's t-statistics or Hotelling's T^2 statistics.

4.4 Diagnostics

4.5 Forecasting

4.6 Examples of application (Production functions)

Bibliography

Svetunkov, I., 2022. complex: Time Series Analysis and Forecasting Using Complex Variables. R package version 0.0.2.41013.
URL https://github.com/config-i1/complex