Complex-valued Econometrics with Examples in R

Sergey Svetunkov and Ivan Svetunkov

2024-03-02

Contents

Introduction **5**
Who is this monograph for? 7

1 Introduction to theory of complex variables **9**
1.1 Theory of complex numbers 9
1.1.1 Complex variables 18
1.1.2 Vectors of complex variables 21
1.2 Complex random variable 22
1.2.1 First moment 22
1.2.2 Second moment 23
1.3 Parametric Distributions of c.r.v. 28
1.3.1 Complex Normal distribution 30
1.3.2 Multivariate Normal distribution 32
1.3.3 Hotelling's T-squared distribution 33
1.4 Complex-valued Modelling 34

2 Simple Complex Linear Regression **37**
2.1 Model formulation 37
2.2 Estimation 39
2.2.1 Ordinary Least Squares 39
2.2.2 Complex Least Squares 42
2.2.3 Likelihood 47
2.3 Comparing different estimators for SCLR 48

3 Correlation analysis of complex random variables **55**
3.1 Visualisation of relations 55
3.2 Types of correlation coefficients 58
3.3 Conjugate correlation 60
3.4 Direct correlation 63
3.5 Pearson's correlation 65
3.6 Correlation matrix 66

4 Multiple Complex Linear Regression **67**

4.1 Model formulation . . . 67
4.2 Estimation . . . 68
4.2.1 Ordinary Least Squares . . . 69
4.2.2 Complex Least Squares . . . 69
4.2.3 Issues with OLS and CLS . . . 70
4.2.4 Likelihood . . . 70
4.3 Inference . . . 71
4.3.1 Demonstration in R . . . 73
4.4 Capturing uncertainty in the multiple cLR . . . 76
4.4.1 Demonstration in R . . . 78
4.5 Dummy variables . . . 79
4.5.1 Demonstration in R . . . 80

5 Assumptions of Complex Linear Models 83
5.1 Model is correctly specified . . . 84
5.1.1 Model does not omit any important variables . . . 84
5.1.2 Model does not have redundant variables . . . 86
5.1.3 Variables transformations . . . 87
5.1.4 No outliers . . . 93
5.2 Residuals are i.i.d. . . . 93
5.2.1 Homoscedastic residuals . . . 93
5.2.2 No autocorrelation in residuals . . . 94
5.2.3 Expectation of residuals is zero . . . 101
5.2.4 Residuals follow an assumed distribution . . . 102
5.2.5 Residuals follow a circular distribution . . . 102
5.3 Explanatory variables are not correlated with anything but the response variable . . . 103
5.3.1 Demonstration in R . . . 106

6 Complex Dynamic Models 117
6.1 Complex model with trend . . . 117
6.1.1 Example in R . . . 119
6.2 Complex AR . . . 121
6.2.1 Example in R . . . 125
6.3 Complex MA . . . 129
6.3.1 Example in R . . . 132
6.4 Complex ARMA and ARIMA . . . 137

7 Examples of application 141
7.1 cARIMA applied to Box-Jenkins Sales data . . . 141
7.2 Seatbels law in Great Britain . . . 147

Conclusions 153

Introduction

The theory of complex variables functions is actively used in a variety of disciplines, including modern physics, engineering sciences (Mandic et al., 2007; Jia and Yang, 2016), information theory (Eriksson and Koivunen, 2006), signal processing (Arens, 1957; Brown and Crane, 1969; Gardner, 1993; Chargé et al., 2001; Gerstacker et al., 2003; Delmas, 2004; Chevalier and Pipon, 2006; Buzzi et al., 2006; Percival, 2006; Tauböck, 2007; Rubin-Delanchy and Walden, 2007; Adali et al., 2008) and others (e.g. Kociuba and Rowe, 2016). 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, such as linear regression, ARIMA, ETS etc. For example, there are only few papers in which methods and models of the theory of complex variables functions would be used in economics, and they are typically used as instruments for diagnostics or statistical tests (for example, for the unit root test, such as ADF from Dickey and Fuller, 1979) rather than in 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) summarised the main principles of using complex variables in economics and business, and discussed how to estimate some of those models. Before him, the monograph more focused on economics was written by Tamari (1997). The author first introduced Wealth as a complex variable consisting of Output (real part) and Money (imaginary part) and showed how modelling can be done efficiently in the area of economics even with such a simple representation. Unfortunately, this work has gone unnoticed in scientific world. We became aware of this work only in 2016, when Ben Tamari himself kindly sent Sergey Svetunkov his monograph.

Since 2012, there has been some developments in the area of modelling using complex variables in business and economics, notably a paper by Svetunkov et al. (2022) on Complex Exponential Smoothing and Kourentzes et al. (2021) 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 engineering and signal processing. The latter group has been using complex autoregressions for a couple of decades, modelling and predicting signals. 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, poor development of theory of complex-valued modelling 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 main principles in statistics and econometrics of complex variables. We hope that the theory explained in this monograph will help statisticians, econometricians and forecasters in using complex-valued models for a wider variety of problems.

Chapter 1 discusses the theory of complex numbers, random complex variables, conventional and complex-valued statistics. Chapter 2 moves to the discussion of the simple complex-valued linear regression and how it can be estimated in sample. We then move to Chapter 3 to discuss correlation analysis of complex variables and then extend the material from Chapter 2 to multiple complex linear regression in Chapter 4. After that, we discuss assumptions of complex-valued models and what issues can their violation cause. We finally move to the dynamic complex-valued models in Chapter 6, showing how complex ARIMA can be estimated and what properties it has. Last but not least, we provide two examples of application of complex-valued models, on two datasets, focusing on dynamic and on the regression complex-valued models in Chapter 7. All of this material is supported by examples in R, which will rely heavily on the package `complex`, developed specifically to support this monograph. This means that anyone can then use the proposed models for purposes of time series analysis and 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 some of 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 uses the `complex` package for R. Many of the examples in R will use functions from this package, so make sure to install it from CRAN before running any R code (Svetunkov, 2024):

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

We will also need several other packages for our analysis, including the following:

```
library(mvtnorm)
library(tseries)
library(greybox)
```

Who is this monograph for?

This monograph will be useful for PhD students and academics working with multivariate models, such as vector autoregressive models and/or multivariate regression. The models discussed here can in general be applied to many real life problems where a pair of response variables needs to be or can be modelled and/or predicted.

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 (Bombelli, 1572). 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. 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:

$$\ln \underline{z} = \ln\left(e^{\ln r + i\phi}\right) = \ln r + i\phi. \tag{1.12}$$

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

$$\underline{z} \times \tilde{\underline{z}} = (x + iy)(x - iy) = x^2 + y^2, \tag{1.13}$$

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

$$\begin{aligned} \widetilde{x + y} &= \tilde{x} + \tilde{y} \\ \widetilde{x - y} &= \tilde{x} - \tilde{y} \\ \widetilde{x \times y} &= \tilde{x} \times \tilde{y} \\ \widetilde{\left(\frac{x}{y}\right)} &= \frac{\tilde{x}}{\tilde{y}}. \end{aligned} \tag{1.14}$$

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

$$\boldsymbol{z} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{1.15}$$

or like a matrix (Halliwell, 2015, provides some examples based on these presentations):

$$\underset{\sim}{\boldsymbol{z}} = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}, \tag{1.16}$$

where the symbol $\sim$ 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 (1.13). For the vector:

$$\boldsymbol{z}\boldsymbol{z}' = \begin{pmatrix} x & y \end{pmatrix} \times \begin{pmatrix} x \\ y \end{pmatrix} = \left(x^2 + y^2\right) \tag{1.17}$$

and for the matrix:

$$\underset{\sim}{\boldsymbol{z}}\underset{\sim}{\boldsymbol{z}}' = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \times \begin{pmatrix} x & y \\ -y & x \end{pmatrix} = \begin{pmatrix} x^2 + y^2 & 0 \\ 0 & x^2 + y^2 \end{pmatrix}. \tag{1.18}$$

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:

$$\underset{\sim}{\boldsymbol{z}}\underset{\sim}{\boldsymbol{z}}^{\top} = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \times \begin{pmatrix} x & -y \\ y & x \end{pmatrix} = \begin{pmatrix} x^2 - y^2 & -2xy \\ 2xy & x^2 - y^2 \end{pmatrix}. \tag{1.19}$$

In this monograph, we will call the multiplications (1.18) and (1.19) "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:

$$\begin{aligned} y_r =& 3 \\ y_i =& 2 \end{aligned}, \tag{1.20}$$

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

$$y_r + iy_i = f(x_r + ix_i) \tag{1.21}$$

we have:

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

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 (1.22) in the linear form (1.21) 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 (1.22) becomes:

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

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 (1.22) can be written as:

$$\begin{pmatrix} y_r \\ y_i \end{pmatrix} = \begin{pmatrix} \mathcal{R}(f(x_r + ix_i)) \\ \mathcal{I}(f(x_r + ix_i)) \end{pmatrix}. \tag{1.23}$$

or:

$$\begin{pmatrix} \mathcal{R}(f(x_r + ix_i)) & -\mathcal{I}(f(x_r + ix_i)) \\ \mathcal{I}(f(x_r + ix_i)) & \mathcal{R}(f(x_r + ix_i)) \end{pmatrix}. \tag{1.24}$$

The representations (1.21), (1.22), (1.23) and (1.24) are equivalent and can be used interchangeably depending on the task at hand.

Finally, in terms of definitions, the function (1.21) 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:

$$y_r + iy_i = (a_r + ia_i)(x_r + ix_i),$$

while for the latter, an example is

$$y_r + iy_i = (a_r + ia_i)(x_r + ix_i)^2,$$

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.

Another important function of complex variables that will be useful in what follows is the power function of the form:

$$y_r + iy_i = (x_r + ix_i)^{a_r + ia_i}. \tag{1.25}$$

In this function, the complex variable $x_{1,r} + ix_{1,i}$ is transformed non-linearly by taking the complex power of it. To better understand how the response variable is connected with the input one, the function (1.25) can be linearised:

$$\ln(y_r + iy_i) = (a_r + ia_i)\ln(x_r + ix_i),$$

which can then be represented in logarithms of exponential form:

$$\frac{1}{2}\ln\left(y_r^2 + y_i^2\right) + i\text{Arg}(y_r + iy_i)) = (a_r + ia_i)\left(\frac{1}{2}\ln\left(x_r^2 + x_i^2\right) + i\text{Arg}(x_r + ix_i))\right).$$

Opening the brackets on the right hand side, we finally get:

$$\begin{aligned}&\frac{1}{2}\ln\left(y_r^2 + y_i^2\right) + i\text{Arg}(y_r + iy_i) = \frac{a_r}{2}\ln\left(x_r^2 + x_i^2\right) - a_i\text{Arg}(x_r + ix_i) + \\ &i\left(\frac{a_i}{2}\ln\left(x_r^2 + x_i^2\right) + a_r\text{Arg}(x_r + ix_i)\right)\end{aligned}.$$

This equation can be represented as a set of two equations, showing how the magnitude and how the argument of the response complex variable are formed:

$$\begin{aligned}\left(y_r^2 + y_i^2\right)^{\frac{1}{2}} &= \left(x_r^2 + x_i^2\right)^{\frac{a_r}{2}} e^{-a_i\text{Arg}(x_r + ix_i)} \\ \text{Arg}(y_r + iy_i) &= \frac{a_i}{2}\ln\left(x_r^2 + x_i^2\right) + a_r\text{Arg}(x_r + ix_i)\end{aligned}. \tag{1.26}$$

In the set of equations (1.26), we see that the parameter a_r connects directly the magnitude and the argument of $\underline{x}$ with the magnitude and the argument of $\underline{y}$: the magnitude is increased in the power of a_r, while the argument is rotated by the same value. On the other hand, a_i captures the cross-connections between the variables: the higher it is, the lower the impact of the argument of $\underline{x}$ on the magnitude of $\underline{y}$ and the higher is the impact of the magnitude of $\underline{x}$ on the argument of $\underline{y}$. The two parts of the complex parameter $a_r + ia_i$ have some similarities with the coefficient of elasticity and with marginal effect, where:

1. a_r shows the percentage change of the magnitude of $\underline{y}$ with the increase of the magnitude of $\underline{x}$ by 1%;
2. a_r at the same time shows by how many radians the argument of $\underline{y}$ changes with the increase of the argument of $\underline{x}$ by one radiant, i.e. it reflects how fast the rotation of the vector happens (the angle increases);
3. The magnitude of $\underline{y}$ changes roughly by $a_i \times 100$ percent with the increase of the argument of $\underline{x}$ by one radiant (this interpretation implies that a_i is close to zero);
4. The argument of $\underline{y}$ will change approximately by $\frac{a_i}{100}$ radians with the increase of the magnitude of $\underline{x}$ by 1% (this interpretation also relies on the assumption that a_i is close to zero).

If the explanatory variable $\underline{x}$ was a real positive rather than complex, the set of equations (1.26) would simplify to:

$$\begin{aligned}\left(y_r^2 + y_i^2\right)^{\frac{1}{2}} &= (x)^{\frac{a_r}{2}} \\ \text{Arg}(y_r + iy_i) &= \frac{a_i}{2}\ln(x)\end{aligned}. \tag{1.27}$$

In this case, the a_r has only the interpretation (1) (substituting the magnitude increase by the percentage increase of x), while the a_i maintains the interpretation (4), which simplifies the whole interpretation substantially.

While there are many other potential non-linear transformations of complex variables, we only discuss the ones above which are typically important in econometrics.

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:

$$\underline{\mathbf{x}} = \begin{pmatrix} x_{r,1} + ix_{i,1} \\ x_{r,2} + ix_{i,2} \\ x_{r,3} + ix_{i,3} \end{pmatrix}, \tag{1.28}$$

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 (1.15), the same complex vector can be written as a matrix:

$$\mathbf{x} = \begin{pmatrix} x_{r,1} & x_{i,1} \\ x_{r,2} & x_{i,2} \\ x_{r,3} & x_{i,3} \end{pmatrix}, \tag{1.29}$$

Finally, given the connection of complex numbers and matrices (1.16), the same complex vector x can be written as a matrix:

$$\underset{\sim}{\mathbf{x}} = \begin{pmatrix} x_{r,1} & -x_{i,1} \\ x_{i,1} & x_{r,1} \\ x_{r,2} & -x_{i,2} \\ x_{i,2} & x_{r,2} \\ x_{r,3} & -x_{i,3} \\ x_{i,3} & x_{r,3} \end{pmatrix}. \tag{1.30}$$

We use the symbol $\sim$ below a character to denote the complex variable in the matrix form (1.30). Each of the forms (1.28), (1.29) and (1.30) 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 (1.28) will not have the same effect - it will only switch rows and columns of the complex vector:

$$\underline{\mathbf{x}}^\top = \begin{pmatrix} x_{r,1} + ix_{i,1} & x_{r,2} + ix_{i,2} & x_{r,3} + ix_{i,3} \end{pmatrix}, \tag{1.31}$$

If we then switch to either the form (1.15) or (1.16), 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:

$$\mathbf{x}^\top = \begin{pmatrix} x_{r,1} & x_{r,2} & x_{r,3} \\ x_{i,1} & x_{i,2} & x_{i,3} \end{pmatrix} \tag{1.32}$$

and

$$\underset{\sim}{\mathbf{x}}^\top = \begin{pmatrix} x_{r,1} & -x_{i,1} & x_{r,2} & -x_{i,2} & x_{r,3} & -x_{i,3} \\ x_{i,1} & x_{r,1} & x_{i,2} & x_{r,2} & x_{i,3} & x_{r,3} \end{pmatrix}. \tag{1.33}$$

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:

$$\underline{\mathbf{x}}' = \begin{pmatrix} x_{r,1} - ix_{i,1} & x_{r,2} - ix_{i,2} & x_{r,3} - ix_{i,3} \end{pmatrix}, \tag{1.34}$$

$$\mathbf{x}' = \begin{pmatrix} x_{r,1} & x_{r,2} & x_{r,3} \\ -x_{i,1} & -x_{i,2} & -x_{i,3} \end{pmatrix} \tag{1.35}$$

and

$$\underset{\sim}{\mathbf{x}}' = \begin{pmatrix} x_{r,1} & x_{i,1} & x_{r,2} & x_{i,2} & x_{r,3} & x_{i,3} \\ -x_{i,1} & x_{r,1} & -x_{i,2} & x_{r,2} & -x_{i,3} & x_{r,3} \end{pmatrix}. \tag{1.36}$$

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 (see, for example, Reed, 1962). 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:

$$\underline{\mu} = \mathrm{E}(\underline{x}) = \mathrm{E}(x_r + ix_i) = \mathrm{E}(x_r) + i\mathrm{E}(x_i) = \mu_r + i\mu_i, \tag{1.37}$$

due to the law of total expectation and because i is a constant. In (1.37), 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:

$$\underline{\hat{\mu}} = \hat{\mu}_r + i\hat{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} x_{r,j} + i\frac{1}{n}\sum_{j=1}^{n} x_{i,j}, \tag{1.38}$$

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):

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_r \\ \mu_i \end{pmatrix} \tag{1.39}$$

or

$$\underset{\sim}{\boldsymbol{\mu}} = \begin{pmatrix} \mu_r & -\mu_i \\ \mu_i & \mu_r \end{pmatrix}. \tag{1.40}$$

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

```
# Random seed for reproducibility
set.seed(41)
# 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] 51.90313+99.91763i
```

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 (Panchev, 1971). So for the c.r.v. $x_r + ix_i$ the variance is:

$$\begin{aligned} \mathrm{V}(\underline{x}) = \sigma_x^2 =& \mathrm{E}((\underline{x} - \underline{\mu})(\tilde{\underline{x}} - \tilde{\underline{\mu}})) = \\ & \mathrm{E}\left(((x_r - \mu_r) + i(x_i - \mu_i))((x_r - \mu_r) - i(x_i - \mu_i))\right) = \\ & \mathrm{E}((x_r - \mu_r)^2) + \mathrm{E}((x_i - \mu_i)^2) \end{aligned} \tag{1.41}$$

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

$$\sigma_x^2 = \sigma_{x_r}^2 + \sigma_{x_i}^2. \tag{1.42}$$

Multiplication by conjugate in (1.41) 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: Visual representation of variance of a complex random variable.

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 (1.42).

Picinbono and Bondon (1997) note that the moment (1.42) is not sufficient to entirely describe the second order statistics of a c.r.v. This is because it ignores the potential covariance 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" (in many papers in signal processing, for example, in Javidi et al., 2011), which can be obtained by applying the conventional formula of variance directly to the c.r.v. without the multiplication by conjugate:

$$\begin{aligned}\mathcal{V}(\underline{x}) = \varsigma_x^2 =& \mathrm{E}((\underline{x}-\underline{\mu})^2) = \mathrm{E}\left(((x_r-\mu_r)+i(x_i-\mu_i))^2\right) = \\ & \mathrm{E}((x_r-\mu_r)^2) - \mathrm{E}((x_i-\mu_i)^2) + i2\mathrm{E}((x_r-\mu_r)(x_i-\mu_i))\end{aligned} \tag{1.43}$$

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

$$\varsigma_x^2 = \sigma_{x_r}^2 - \sigma_{x_i}^2 + i2\sigma_{x_r,x_i}. \tag{1.44}$$

In the literature, the moment (1.44) is called pseudo-variance (e.g., Trampitsch, 2013), 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 (1.44) 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 (1.42) implies minimising both variances of the real and imaginary parts of x (ignoring the covariance between them), minimising (1.44) 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".

Having calculated the two variances, we can then extract variances of real and imaginary parts and the covariance between them via formulae (which follows directly from (1.42) and (1.44)):

$$\begin{aligned}
\sigma^2_{x_r} &= \frac{\mathcal{R}(\sigma^2_x) + \mathcal{R}(\varsigma^2_x)}{2} \\
\sigma^2_{x_i} &= \frac{\mathcal{R}(\sigma^2_x) - \mathcal{R}(\varsigma^2_x)}{2} \\
\sigma_{x_r,x_i} &= \frac{\mathcal{I}(\varsigma^2_x)}{2}
\end{aligned} . \tag{1.45}$$

These individual real-valued moments might be useful in case of hypothesis testing or constructing confidence intervals for each of the parts of c.r.v. independently.

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

```
# Random seed for reproducibility
set.seed(41)
# 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
## 113.0543+0i
```

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

```
    setNames("Direct variance")
```

```
##    Direct variance
## -60.19156-5.69425i
```

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:

$$\mathbf{\Sigma}_x = \begin{pmatrix} \sigma^2_{x_r} & \sigma_{x_r,x_i} \\ \sigma_{x_r,x_i} & \sigma^2_{x_i} \end{pmatrix}. \tag{1.46}$$

The minimisation of the matrix (1.46) does not make sense, but instead it is possible to minimise the determinant of that matrix, which is called "Generalised Variance" (GV):

$$\text{GV} = |\mathbf{\Sigma}_x| = \sigma^2_{x_r}\sigma^2_{x_i} - \sigma^2_{x_r,x_i}. \tag{1.47}$$

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

```
covar(x)
```

```
##           x_r       x_i
## x_r 26.431373 -2.847127
## x_i -2.847127 86.622930
```

The minimisation of GV, as can be seen from the formula (1.47), 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] 2281.457
```

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 (Svetunkov, 2022):

$$\begin{aligned}\hat{\sigma}^2_{x_r} =& \frac{1}{n-k}\sum_{j=1}^{n}(x_{r,j}-\bar{x}_{r,j})^2\\ \hat{\sigma}^2_{x_i} =& \frac{1}{n-k}\sum_{j=1}^{n}(x_{i,j}-\bar{x}_{i,j})^2\\ \hat{\sigma}_{x_r,x_i} =& \frac{1}{n-k}\sum_{j=1}^{n}(x_{r,j}-\bar{x}_{r,j})(x_{i,j}-\bar{x}_{i,j})^2,\end{aligned} \tag{1.48}$$

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 $y = y_r + iy_i$ (Picinbono and Bondon, 1997). The "conjugate" covariance is calculated similarly to (1.41) via the multiplication by conjugate (Neeser and Massey, 1993):

$$\begin{aligned}\sigma_{x,y} =& \mathrm{E}((\tilde{x}-\tilde{\mu}_x)(y-\mu_y)) =\\ & \mathrm{E}\left(((x_r-\mu_{x,r})-i(x_i-\mu_{x,i}))((y_r-\mu_{y,r})+i(y_i-\mu_{y,i}))\right) =\\ & \mathrm{E}((x_r-\mu_{x,r})(y_r-\mu_{y,r}))+\mathrm{E}((x_i-\mu_{x,i})(y_i-\mu_{y,i}))+\\ & i\left(\mathrm{E}((x_r-\mu_{x,r})(y_i-\mu_{y,i}))-\mathrm{E}((x_i-\mu_{x,i})(y_r-\mu_{y,r}))\right)\end{aligned} \tag{1.49}$$

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:

$$\sigma_{x,y} = \sigma_{x_r,y_r} + \sigma_{x_i,y_i} + i(\sigma_{x_r,y_i} - \sigma_{x_i,y_r}), \tag{1.50}$$

which similarly to the variance (1.42) 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 (1.50) 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, which is also known in the signal processing literature as pseudo-covariance (Neeser and Massey, 1993, Gao et al. (2019)), which can be shown to be equal to:

$$\varsigma_{x,y} = \sigma_{x_r,y_r} - \sigma_{x_i,y_i} + i(\sigma_{x_i,y_r} + \sigma_{x_r,y_i}). \tag{1.51}$$

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:

$$\mathbf{\Sigma}_{x,y} = \begin{pmatrix} \sigma^2_{x_r} & \sigma_{x_r,x_i} & \sigma_{x_r,y_r} & \sigma_{x_r,y_i}\\ \sigma_{x_r,x_i} & \sigma^2_{x_i} & \sigma_{x_i,y_r} & \sigma_{x_i,y_i}\\ \sigma_{x_r,y_r} & \sigma_{x_i,y_r} & \sigma^2_{y_r} & \sigma_{y_r,y_i}\\ \sigma_{x_r,y_i} & \sigma_{x_i,y_i} & \sigma_{y_r,y_i} & \sigma^2_{y_i}\end{pmatrix}. \tag{1.52}$$

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

```
# Random seed for reproducibility
set.seed(41)
# 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
## 70.4003-101.2452i
```

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

```
##  Direct covariance
## -51.28506+50.8207i
```

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 -60.19156- 5.69425i -51.28506+50.82070i
## y -51.28506+50.82070i  12.71148+48.05904i
```

While there exist higher order moments for complex random variables (see, for example, Javidi et al., 2011), 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):

```
# Random seed for reproducibility
set.seed(41)
# 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 (Wooding, 1956):

$$f(\underline{x}) = \frac{1}{\pi^2\sqrt{\sigma^4 - \varsigma^2\tilde{\varsigma}^2}} \exp\left(-\frac{1}{2}\begin{pmatrix}\tilde{\underline{x}} - \tilde{\underline{\mu}} & \underline{x} - \underline{\mu}\end{pmatrix}\begin{pmatrix}\sigma^2 & \varsigma^2 \\ \tilde{\varsigma}^2 & \sigma^2\end{pmatrix}^{-1}\begin{pmatrix}\underline{x} - \underline{\mu} \\ \tilde{\underline{x}} - \tilde{\underline{\mu}}\end{pmatrix}\right), \tag{1.53}$$

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 (1.53) 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.

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

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.

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

One special case of a Complex Normal distribution is a so called "Circular Symmetric" Complex Normal distribution (Neeser and Massey, 1993). 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 zero (Amblard et al., 1996, studied its properties in more detail). This specific distribution is used extensively in signal processing literature and also relates to the so-called "propriety" property, meaning that the direct covariance of the complex variables is zero (Walden and Rubin-Delanchy, 2009; Adali et al., 2011).

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 (1.23) and thus to revert to the

Figure 1.10: Heatmap of density of Complex Normal distribution.

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 (1.46) 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:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}\right)^{\top}\mathbf{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right). \tag{1.54}$$

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:

$$\mathbf{\Sigma} = \frac{1}{2}\begin{pmatrix} \mathcal{R}(\sigma^2) + \mathcal{R}(\varsigma^2) & \mathcal{I}(\varsigma^2) \\ \mathcal{I}(\varsigma^2) & \mathcal{R}(\sigma^2) - \mathcal{R}(\varsigma^2) \end{pmatrix}. \tag{1.55}$$

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 $\mathbf{q}$:

$$(\mathbf{x}-\mathbf{q})^{\top}\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{q}) \leq \chi^2(n,p), \tag{1.56}$$

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.:

$$x_r \in \left(\mu_{x_r} + \sigma_{x_r} z\left(\frac{\alpha}{2}\right), \mu_{x_r} + \sigma_{x_r} z\left(\frac{1+\alpha}{2}\right)\right), \tag{1.57}$$

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.

Another distribution related to the multivariate normal one is the multivariate complex normal. It was proposed and studied by Goodman (1963) and then extended by van den Bos (1995). We do not use it in what comes in the monograph and mention it only for completeness.

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 $\hat{\boldsymbol{\Sigma}}_{\mu_x}$ then the following holds (if CLT holds):

$$(\hat{\boldsymbol{\mu}}_x - \boldsymbol{\mu}_x)'\hat{\boldsymbol{\Sigma}}_{\mu_x}^{-1}(\hat{\boldsymbol{\mu}}_x - \boldsymbol{\mu}_x) \sim T^2(2, n-1) = \frac{2(n-1)}{n-2}F(2, n-2), \tag{1.58}$$

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 (1.58) 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 (1.58) is the following:

$$\begin{aligned} \mathrm{H}_0 &: \boldsymbol{\mu}_x = 0 \\ \mathrm{H}_1 &: \boldsymbol{\mu}_x \neq 0 \end{aligned} \tag{1.59}$$

or equivalently:

$$\begin{aligned} \mathrm{H}_0 &: \mu_{x_r} = \mu_{x_i} = 0 \\ \mathrm{H}_1 &: \mu_{x_r} \neq 0 \vee \mu_{x_i} \neq 0 \end{aligned} \tag{1.60}$$

However, from practical point of view, testing the joint hypothesis (1.60) 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 (1.60), rejecting the null hypothesis implies that something is not equal to zero, but it is not clear, what specifically.

1.4 Complex-valued Modelling

Now that we have discussed basics of complex variables theory, we can discuss the modelling aspects: it might not look straightforward, what should be included in the real and the imaginary part of a complex variable. There are two philosophical views on complex-valued modelling:

1. You need to find processes, where complex variables arise naturally;
2. You can model anything with complex variables as long as you follow some rules.

If you follow the former view you need to find some variables that form two parts of the whole that work together. An example of this case is the wind power and direction forecasting (Goh et al., 2004; Javidi et al., 2008; Knight and Nunes, 2019). The two variables in this case will be naturally represented in the exponential form:

$$\underline{y} = r\mathrm{e}^{i\phi},$$

where r is the power of wind and ϕ is its direction. In this formulation, we can then model, for example, how the two characteristics of wind on some observation t depend on the previous values and some external information:

$$\underline{y}_t = f(\underline{y}_{t-1}, \underline{y}_{t-2}, \dots, \underline{y}_{t-p}, \underline{\mathbf{x}}_t, \underline{\epsilon}_t),$$

where t is the index of the observation, p is the order of the model, $\underline{\mathbf{x}}_t$ is the vector of all explanatory complex variables under consideration and $\underline{\epsilon}_t$ is the error term. Some aspects of such a model are discussed in the Chapter 4 of this book, while the others are covered in Chapter 6.

However, it might be hard to find examples like the one with the wind above in economics and business domains. Trying to find something analogous (power + i direction) might lead us into a dead end. This is why the second philosophical view exists. Svetunkov (2012) argues that any two variables can be united in a complex variable (in the arithmetic form) as long as:

1. They reflect two sides of a process/phenomenon;
2. They have the same scale.

The second condition implies that some sort of scaling is required if the variables have different scales. If this condition is not satisfied then the operations with

complex variables might lead to meaningless results. For example, if we consider a variable of sales quantity and product price, $q+ip$ and multiply it by a complex parameter $a_r + ia_i$, we get a complex variable that has a strange combination in it:

$$\underline{z} = (q + ip)(a_r + ia_i) = qa_r - pa_i + i(qa_i + pa_r). \tag{1.61}$$

In the equation (1.61), the resulting real part includes the multiplication of the number of units sold times the parameter a_r, while the imaginary part has the price (e.g. in U.S. dollars) times the same parameter a_r. If the both parts of the variable were scaled the parameter a_r would act as a transformer from one value to another. But if the both parts are not scaled then in one case it acts as a scaler of units, while the in the other one as a scaler of U.S. dollars, which does not make much sense. This example shows why the condition (2) above is important in complex-valued modelling.

The scaling itself can be done using one of the conventional ways:

a. Normalisation: subtract the minimum value from all the values and then divide by the range. This way the normalised variable will lie between 0 and 1;
b. Standardisation: subtract the mean and divide the variable by its standard deviation. This way we centre the variable and get rid of scale;
c. Any other form of scaling, where the division is done by the value of the variable, thus making it unitless.

If we deal with time series data, an additional operation might be important - taking differences to make the data stationary (this is discussed in Section 6.4). If we do not do that, the new values of the variable might lie outside of the range defined by the scaling.

Incidentally, the example with the wind power above abides to these two conditions, because the variable itself is represented in the exponential form, thus the condition will be satisfied automatically when we use formula (1.6) to move to the arithmetic one. This variable does not require any additional transformations and can be treated as is.

Furthermore, based on the discussion in Subsection 1.1.2, we can note that the complex-valued models share some structural similarities with vector models, such as Vector Autoregression, when they are applied to two variables. From this angle, the complex-valued models can be considered as special cases of the vector ones, when some restrictions are imposed on the matrix of parameters. This important property gives the complex-valued models their power: they need fewer parameters to estimate in comparison with their vector counterparts.

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. There are many approaches for estimation of parameters of a complex-valued regression. In this monograph, we focus only on the three of them: Ordinary Least Squares (OLS), Complex Least Squares (CLS), and Likelihood.

2.2.1 Ordinary Least Squares

We start with the conventional Ordinary Least Squares (OLS), which relies on the multiplication by conjugate number. To our knowledge, van den Bos (1994) was the first paper that discussed it in the complex-valued domain. For the simple cLR, it comes to minimising the following loss based on the residuals:

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

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

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

which after opening the brackets becomes:

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

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:

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

to get:

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

As shown by Svetunkov (2012), solving the system of equations (2.17) gives the following formulae for the parameters of the model:

$$\begin{aligned}
\underline{b}_1 &= \frac{\sum_{j=1}^{n}(\underline{y}_j - \underline{\hat{\mu}}_y)(\tilde{\underline{x}}_j - \hat{\tilde{\mu}}_x)}{\sum_{j=1}^{n}(\underline{x}_j - \underline{\hat{\mu}}_x)(\tilde{\underline{x}}_j - \hat{\tilde{\mu}}_x)} \\
\underline{b}_0 &= \frac{1}{n}\sum_{j=1}^{n}\underline{y}_j - \underline{b}_1\sum_{j=1}^{n}\underline{x}_j,
\end{aligned} \tag{2.18}$$

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. Coming back to figure 1.6 from Subsection 1.2.2, this also implies that they minimise the hypotenuse on the figure, thus moving the residuals closer to the middle line. So, if the real and imaginary parts of the residuals are correlated in the "true" model, the OLS will automatically take care of this situation. However, it is worth noting that the OLS ignores the potential covariance between the real and the imaginary parts, 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:

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

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:

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

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:

$$\sum_{j=1}^{n}(\underline{e}_j\underline{e}_j) = \sum_{j=1}^{n}\underline{y}^2_j + n\underline{b}^2_0 + \underline{b}^2_1\sum_{j=1}^{n}\underline{x}^2_j - 2\underline{b}_0\sum_{j=1}^{n}\underline{y}_j - 2\underline{b}_1\sum_{j=1}^{n}\underline{x}_j\underline{y}_j + 2\underline{b}_0\underline{b}_1\sum_{j=1}^{n}\underline{x}_j \tag{2.28}$$

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:

$$\begin{aligned} \frac{d\sum_{j=1}^{n}(\underline{e}^2_j)}{d\underline{b}_0} &= 0 \\ \frac{d\sum_{j=1}^{n}(\underline{e}^2_j)}{d\underline{b}_1} &= 0. \end{aligned} \tag{2.29}$$

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

$$\begin{aligned} 2n\underline{b}_0 - 2\sum_{j=1}^{n}\underline{y}_j + 2\underline{b}_1\sum_{j=1}^{n}\underline{x}_j &= 0 \\ 2\underline{b}_1\sum_{j=1}^{n}\underline{x}^2_j - 2\sum_{j=1}^{n}\underline{x}_j\underline{y}_j + 2\underline{b}_0\sum_{j=1}^{n}\underline{x}_j &= 0. \end{aligned} \tag{2.30}$$

The solution for this system of equations, as it was shown by Svetunkov (2012),

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

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

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:

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

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:

```
# Random seed for reproducibility
set.seed(41)
# 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 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)
```

Figure 2.1: Variety of CLS loss function values for different values of $\underline{b}_1$.

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

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 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 (van den Bos, 1994). 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):

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

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 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 $\hat{\sigma}_{x_r}^2 = \hat{\sigma}_{x_i}^2$ 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

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	$\sigma_{\epsilon_r} = \sigma_{\epsilon_i} = 100$

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 (Scenario 1):

```
# Random seed for reproducibility
set.seed(41)
# Number of observations
obs <- 10000
x0 <- rnorm(obs,10,10)
x <- complex(real=x0,imaginary=rnorm(obs,0,20))

# Parameters of the model
b0 <- 10 + 15i
b1 <- 2-1.5i
# Response variable
y <- b0 + b1 * x + 1.5*complex(real=rnorm(obs,0,1),
                               imaginary=rnorm(obs,0,1))

# Form a matrix of variables
complexData <- cbind(y,x)

# Number of iterations and an array for parameters
nsim <- 9980
parametersValues <-
    array(NA, c(nsim,3,2),
          dimnames=list(NULL,
                        c("CLS","OLS","Likelihood"),
                        c("b0","b1")))

# The loop with cLR estimated with CLS, OLS and Likelihood
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.

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.

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.

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.

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 2.5: 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.

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.

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.

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.

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 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.

```
# Random seed for reproducibility
set.seed(41)
# 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):

```
complex2vec(x) |> dist() |> cmdscale(k=1) -> xScaled
complex2vec(y) |> dist() |> cmdscale(k=1) -> yScaled
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 main correlation coefficients for complex variables (Schreier and Scharf, 2010): the conjugate and the direct correlation (the latter is known in the literature as "pseudo-correlation", see, for example, Chung, 2020). 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 Panchev, 1971) 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.

```r
# Random seed for reproducibility
set.seed(41)
# Generate the explanatory variable
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
```

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

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

```r
# 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
```

```r
# Same thing using the ccor() function
ccor(x,y,method="conjugate")
```

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

This also means that the conjugate correlation coefficient is always positive, only showing the average strength of the relation between variables, but not its direction. We note that Schreier and Scharf (2010) acknowledged the issue, discussing the properties of the coefficient and its derivatives.

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. There are other ways to estimate correlation coefficients (e.g., Miyabe et al., 2015) and there are other formulae of coefficients (Schreier and Scharf, 2010), but they lie outside of the scope of the monograph.

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. Another condition leading to zero conjugate correlation coefficient is $\sigma_{x_r,y_r} = -\sigma_{x_i,y_i}$ and $\sigma_{x_i,y_r} = \sigma_{x_r,y_i}$, which means that if the coefficient equals to zero, this does not mean that the there is no linear relation between two complex variables.

On the other hand, the coefficient will be close to one if the complex relation between variables $\underline{y}$ and $\underline{x}$ is close to the linear, i.e. $\underline{a}_1 = \frac{1}{\underline{b}_1}$. However, it does not show the direction of the relation, because it cannot be negative due to the calculation of the magnitude in the numerator of (3.7).

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.

```
# Set seed for reproducibility
set.seed(41)
# 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="")
```

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

```
## [1] 0.997059
```

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.

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

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

Figure 3.4: Visualisation of relations between two complex variables

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 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 in theory also be obtained due to the values of direct variance of variables $\underline{x}$ and/or $\underline{y}$.

To get other insights about the direct correlation coefficient, we express its denominator in the exponential form:

$$\sqrt{(\sigma_{x_r}^2 - \sigma_{x_i}^2 + i2\sigma_{x_r,x_i})(\sigma_{y_r}^2 - \sigma_{y_i}^2 + i2\sigma_{y_r,y_i})} = R_1^{\frac{1}{2}} e^{i\frac{\phi_1}{2}}, \tag{3.12}$$

where

$$\begin{aligned} R_1 =& \sqrt{\left((\sigma_{x_r}^2 - \sigma_{x_i}^2)(\sigma_{y_r}^2 - \sigma_{y_i}^2) - 4\sigma_{x_r,x_i}\sigma_{y_r,y_i}\right)^2 + 4\left((\sigma_{x_r}^2 - \sigma_{x_i}^2)\sigma_{y_r,y_i} + (\sigma_{y_r}^2 - \sigma_{y_i}^2)\sigma_{x_r,x_i}\right)^2} \\ \phi_1 =& \arctan\left(\frac{2\left((\sigma_{x_r}^2 - \sigma_{x_i}^2)\sigma_{y_r,y_i} + (\sigma_{y_r}^2 - \sigma_{y_i}^2)\sigma_{x_r,x_i}\right)}{(\sigma_{x_r}^2 - \sigma_{x_i}^2)(\sigma_{y_r}^2 - \sigma_{y_i}^2) - 4\sigma_{x_r,x_i}\sigma_{y_r,y_i}}\right), \end{aligned} \tag{3.13}$$

which are obtained by opening the brackets inside the square root of (3.12). If we now insert (3.12) in (3.11) and multiply both numerator and denominator by conjugate number to the (3.12) we will get:

$$\varrho_{x,y} = \frac{\left(\sigma_{x_r,y_r} - \sigma_{x_i,y_i} + i(\sigma_{x_i,y_r} + \sigma_{x_r,y_i})\right) e^{-i\frac{\phi_1}{2}}}{R_1^{\frac{1}{2}}}. \tag{3.14}$$

Representing the $\sigma_{x_r,y_r} - \sigma_{x_i,y_i} + i(\sigma_{x_i,y_r} + \sigma_{x_r,y_i})$ as $R_2 e^{i\phi_2}$, where $R_2 = \sqrt{(\sigma_{x_r,y_r} - \sigma_{x_i,y_i})^2 + (\sigma_{x_i,y_r} + \sigma_{x_r,y_i})^2}$ and $\phi_2 = \arctan\left(\frac{\sigma_{x_i,y_r} + \sigma_{x_r,y_i}}{\sigma_{x_r,y_r} - \sigma_{x_i,y_i}}\right)$, and inserting these values in (3.14) we get:

$$\varrho_{x,y} = \frac{R_2}{\sqrt{R_1}} e^{i\left(\phi_2 - \frac{\phi_1}{2}\right)}. \tag{3.15}$$

or in an even shorter exponential form $\varrho_{x,y} = |\varrho_{x,y}| e^{i\arg(\varrho_{x,y})}$, where:

$$|\varrho_{x,y}| = \frac{R_2}{\sqrt{R_1}} = \sqrt{\frac{(\sigma_{x_r,y_r} - \sigma_{x_i,y_i})^2 + (\sigma_{x_i,y_r} + \sigma_{x_r,y_i})^2}{\sqrt{\left((\sigma_{x_r}^2 - \sigma_{x_i}^2)(\sigma_{y_r}^2 - \sigma_{y_i}^2) - 4\sigma_{x_r,x_i}\sigma_{y_r,y_i}\right)^2 + 4\left((\sigma_{x_r}^2 - \sigma_{x_i}^2)\sigma_{y_r,y_i} + (\sigma_{y_r}^2 - \sigma_{y_i}^2)\sigma_{x_r,x_i}\right)^2}}} \tag{3.16}$$

and

$$\begin{aligned} \arg(\varrho_{x,y}) =& \phi_2 - \frac{1}{2}\phi_1 = \arctan\left(\frac{\sigma_{x_i,y_r} + \sigma_{x_r,y_i}}{\sigma_{x_r,y_r} - \sigma_{x_i,y_i}}\right) - \\ & \frac{1}{2}\arctan\left(\frac{2\left((\sigma_{x_r}^2 - \sigma_{x_i}^2)\sigma_{y_r,y_i} + (\sigma_{y_r}^2 - \sigma_{y_i}^2)\sigma_{x_r,x_i}\right)}{(\sigma_{x_r}^2 - \sigma_{x_i}^2)(\sigma_{y_r}^2 - \sigma_{y_i}^2) - 4\sigma_{x_r,x_i}\sigma_{y_r,y_i}}\right). \end{aligned} \tag{3.17}$$

Analysing (3.16) (3.17), we can identify several conditions that lead to specific values of the direct correlation coefficient:

1. It is pure real valued, when all covariances are zero, i.e. there is no linear relation between parts of variables.
2. Another situation with real valued direct correlation is when cross-covariances are zero and variances of respective real and imaginary parts are equal. This is a more exotic case than (1);

3. In general, the number will be real when $\phi_2 = \frac{\phi_1}{2}$, which assumes that there is a multitude of combinations of covariances and variances that will satisfy the condition. Analysing all of them becomes close to impossible due to the non-linearity in the condition (3.17).

In the case (1), the magnitude of the coefficient, $\frac{R_2}{\sqrt{R_1}}$ will be equal to zero as well, which agrees with what we discussed earlier. In addition, the analysis of the magnitude (3.16) shows that in an exotic case of $\sigma_{x_r,y_r} = \sigma_{x_i,y_i}$ and $\sigma_{x_i,y_r} = -\sigma_{x_r,y_i}$, the value of the coefficient of the direct correlation will also be equal to zero. All of this means that in general, making solid conclusions just based on the direct correlation coefficient is not possible - we need to calculate both the direct and the conjugate correlations to get a full picture about the relations between two complex variables.

For the same example as above, the direct correlation coefficient is:

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

```
## [1] 1.004573+0.000395i
```

The fact that the coefficient is close to one, means that the relation between the two complex variables is close to linear. On the other hand, the imaginary part being close to zero implies that the real and imaginary parts of the variables $\underline{x}$ and $\underline{y}$ are close to each other (this is how we generated the data). As we see, the direct correlation gives additional information that the conjugate one does not provide. This demonstrates that they should be used in the analysis of relations jointly.

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
complex2vec(x) |> dist() |> cmdscale(k=1) -> xScaled
complex2vec(y) |> dist() |> cmdscale(k=1) -> yScaled
# Calculate the correlation coefficient
cor(xScaled,yScaled)
```

```
##           [,1]
## [1,] 0.9969845
```

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

```
ccor(x, y, method="pearson") |> round(3)
```

The interpretation of the coefficient is straightforward and follows the conven-

tional 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:

```
cbind(x,y) |> covar() |> cov2cor() |> round(3)
```

```
##        x_r    x_i    y_r   y_i
## x_r  1.000 -0.060  0.761 0.712
## x_i -0.060  1.000 -0.689 0.654
## y_r  0.761 -0.689  1.000 0.091
## y_i  0.712  0.654  0.091 1.000
```

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:

$$\underline{y}_j = \underline{\beta}_0 + \underline{\beta}_1 \underline{x}_{1,j} + \underline{\beta}_2 \underline{x}_{2,j} + \cdots + \underline{\beta}_{k-1} \underline{x}_{k-1,j} + \underline{\epsilon}_j, \tag{4.1}$$

where $k-1$ 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:

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

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 $\underline{\mathbf{X}} = \begin{pmatrix} 1 & \underline{x}_{1,1} & \cdots & \underline{x}_{k-1,1} \\ 1 & \underline{x}_{1,2} & \cdots & \underline{x}_{k-1,2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \underline{x}_{1,n} & \cdots & \underline{x}_{k-1,n} \end{pmatrix}$ and $\underline{\boldsymbol{\beta}} = \begin{pmatrix} \underline{\beta}_0 \\ \underline{\beta}_1 \\ \vdots \\ \underline{\beta}_{k-1} \end{pmatrix}$, 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 $\mathbf{y}_j = \begin{pmatrix} y_{r,j} \\ y_{i,j} \end{pmatrix}$, $\underset{\sim}{\mathbf{X}}_j = \begin{pmatrix} 1 & 0 & x_{1,r,j} & -x_{1,i,j} & \cdots & x_{k-1,r,j} & -x_{k-1,i,j} \\ 0 & 1 & x_{1,i,j} & x_{1,r,j} & \cdots & x_{k-1,i,j} & x_{k-1,r,j} \end{pmatrix}$, $\boldsymbol{\beta}' = \begin{pmatrix} \beta_{0,r} & \beta_{0,i} & \beta_{1,r} & \beta_{1,i} & \cdots & \beta_{1,k-1} & \beta_{1,k-1} \end{pmatrix}$ 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):

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

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:

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

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.

However, **?** shows for multivariate models that the maximum of the likelihood gives the same estimates of parameters as the OLS as long as the Multivariate Normal distribution with zero covariance is assumed for the error term. The results might differ for the other distributions or for the case, when the covariance in $\hat{\boldsymbol{\Sigma}}_\epsilon$ is not zero.

4.3 Inference

It is possible to calculate the variance of estimates of parameters, using the same approach as used in the 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 variances can then be used in hypothesis testing or confidence interval construction (their distribution was studied by Tavares and Tavares, 2007). 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):

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

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:

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

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 (1.14), the same formula can be rewritten as:

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

which uses the property: $\tilde{\underline{\mathbf{X}}}' = \underline{\mathbf{X}}^{\top}$ The expectation of the product in (4.24) can be rewritten as a product of expectations plus a covariance between the terms:

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

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:

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

and

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

for the CLS. Having both direct and conjugate variances for OLS and CLS, we can calculate individual variances for each of the parameters and covariances between them to form a classical covariance matrix. The general formula for this would be:

$$\begin{aligned}
\mathrm{V}_{\boldsymbol{b}_r} &= \frac{\mathcal{R}\left(\mathrm{V}\left(\underline{\boldsymbol{b}}\right)\right) + \mathcal{R}\left(\mathcal{V}\left(\underline{\boldsymbol{b}}\right)\right)}{2} \\
\mathrm{V}_{\boldsymbol{b}_i} &= \frac{\mathcal{R}\left(\mathrm{V}\left(\underline{\boldsymbol{b}}\right)\right) - \mathcal{R}\left(\mathcal{V}\left(\underline{\boldsymbol{b}}\right)\right)}{2}, \\
\mathrm{V}_{\boldsymbol{b}_{r,i}} &= \frac{\mathcal{I}\left(\mathcal{V}\left(\underline{\boldsymbol{b}}\right)\right)}{2}
\end{aligned}$$

where $\mathrm{V}_{\boldsymbol{b}_r}$ is the part of the covariance matrix of real values of parameters, $\mathrm{V}_{\boldsymbol{b}_i}$ is the part of the covariance matrix of imaginary values of parameters and $\mathrm{V}_{\boldsymbol{b}_{r,i}}$ is the part of the covariance matrix between the real and imaginary parts of parameters. The final covariance matrix can be formed as:

$$\hat{\boldsymbol{\Sigma}}_{\beta} = \begin{pmatrix} \mathrm{V}_{\boldsymbol{b}_r} & \mathrm{V}_{\boldsymbol{b}_{r,i}} \\ \mathrm{V}_{\boldsymbol{b}_{r,i}} & \mathrm{V}_{\boldsymbol{b}_i} \end{pmatrix}. \tag{4.29}$$

Finally, in case of the likelihood estimation, the covariance matrix of parameters cannot be calculated analytically, but can be obtained numerically via the Hessian calculation. Given the formulation in this case, there is no need to do any additional transformations, the matrix will contain the variances and covariances of each individual parameter of the model.

4.3.1 Demonstration in R

In R, the estimation of the cLR can be done using the `clm()` function from the `complex` package. The specific estimation method can be selected using the `loss` parameter. For demonstration purposes, we will create an artificial data and see how the estimators work.

```
set.seed(41)
# Sample size
obs <- 100
# Generate parameters
b0 <- 100 - 150i
b1 <- 2.5 + 1.5i
b2 <- 1.5 - 0.75i
# Create the explanatory variables
x1 <- rcnorm(obs, mu=100+150i, sigma2=200, varsigma2=100)
# Create the explanatory variables
x2 <- rcnorm(obs, mu=150+100i, sigma2=200, varsigma2=150)
# Generate error term from the complex normal distribution
```

```
e <- rcnorm(obs, mu=0, sigma2=200, varsigma2=80+150i)
# Generate the response variable
y <- b0 + b1 * x1 + b2 * x2 + e
# Form a data frame with the variables
dataArtificial <- data.frame(y=y, x1=x1, x2=x2)
```

The data generated using the code above will have the same variances of the real and imaginary parts of `x1` and the different ones for the `x2`. Furthermore, it will have residuals with correlated real and imaginary parts. We expect that in this situation the likelihood would be more appropriate than either OLS, or CLS.

We start by estimating the model using OLS:

```
clm(y~., dataArtificial, loss="OLS") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: OLS
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r  103.7956    16.7547    70.5421    137.0491 *
## (Intercept)_i -151.3386    12.1603  -175.4735   -127.2037 *
## x1_r             2.4161     0.0724     2.2724      2.5597 *
## x1_i             1.4877     0.0525     1.3834      1.5919 *
## x2_r             1.5625     0.0632     1.4371      1.6879 *
## x2_i            -0.6895     0.0459    -0.7805     -0.5985 *
##
## Error covariance matrix:
##         e_r     e_i
## e_r 96.1432 52.0283
## e_i 52.0283 50.6449
##
## Sample size: 100
## Number of estimated parameters: 3
## Number of degrees of freedom: 97
```

As we see from the output above, the estimates of parameters are reasonable and close to the ones used in the data generation, while the 95% confidence intervals include the true values of parameters. While it does not mean that this will always be the case universally, this demonstration shows how the OLS works and how the confidence intervals of parameters (based on the standard errors) can be obtained using the `clm()` function.

Keeping in mind that the MLE should give similar (if not the same) estimates of parameters, we should expect that the command below should produce a very similar output to the one above:

```
clm(y~., dataArtificial, loss="likelihood") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: likelihood
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r  103.7631    16.8860    70.2424    137.2838 *
## (Intercept)_i -151.3386    12.2552  -175.6666   -127.0107 *
## x1_r             2.4169     0.0729     2.2721      2.5616 *
## x1_i             1.4877     0.0529     1.3826      1.5927 *
## x2_r             1.5625     0.0637     1.4361      1.6889 *
## x2_i            -0.6895     0.0462    -0.7812     -0.5977 *
##
## Error covariance matrix:
##         e_r     e_i
## e_r 97.6559 52.8591
## e_i 52.8591 51.4379
##
## Sample size: 100
## Number of estimated parameters: 4.5
## Number of degrees of freedom: 95.5
## Information criteria:
##      AIC     AICc      BIC     BICc
## 1338.345 1338.869 1350.069 1351.275
```

While the estimates of parameters are not exactly the same, they are very similar. The difference arises from the estimation routine in the two methods: the OLS has analytical formulae for the estimates of parameters, while the likelihood has to rely on a numerical optimisation. Note that the error covariance matrix estimated using likelihood is slightly different than the one from OLS. One of the reasons for this is because the maximisation of the likelihood implies in addition to the minimisation of variances also the maximisation of the covariance between the real and imaginary parts of the complex residuals. Note however that the difference is not substantial, both methods produced very similar results.

The CLS, on the other hand, will give other estimates of parameters:

```
clm(y~., dataArtificial, loss="CLS") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: CLS
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
```

```
## (Intercept)_r   74.8585    22.6298    29.9447    119.7723 *
## (Intercept)_i -131.4084    31.7400  -194.4036    -68.4132 *
## x1_r             2.3853     0.1567     2.0743      2.6964 *
## x1_i             1.2779     0.1384     1.0031      1.5527 *
## x2_r             1.5833     0.0823     1.4198      1.7467 *
## x2_i            -0.6633     0.0613    -0.7850     -0.5416 *
##
## Error covariance matrix:
##          e_r     e_i
## e_r 102.1875 51.3523
## e_i  51.3523 53.2803
##
## Sample size: 100
## Number of estimated parameters: 3
## Number of degrees of freedom: 97
```

The estimates of parameters in the output above are less efficient than the ones from OLS/likelihood (as expected), but they are still not far from the true values of parameters. Notably, the CLS produced the lowest covariance between the real and imaginary parts of the complex residuals, which makes sense given how the CLS loss is formulated.

This demonstration shows how simple it is to estimate a cLR on data and how its outputs can be read. As one can see, the outputs do not differ substantially from the ones of the real values models.

4.4 Capturing uncertainty in the multiple cLR

Having both direct and conjugate variances for OLS and CLS, it is possible to do several things:

1. Construct confidence intervals for parameters;
2. Construct a confidence ellipse for each complex parameter;
3. Test hypothesis about each of parameters (univariate distribution);
4. Test hypothesis for each complex parameter (considering joint distribution);
5. Produce confidence intervals for the conditional expectation from the model (fitted values and point forecasts);
6. Produce prediction intervals for the actual values.

The elements (1) and (3) can be done assuming that the Central Limit Theorem works and thus the estimates of parameters follow Normal distribution. This can be done using Student's t-statistics as discussed in Section 6.4 and 8.1 of Svetunkov (2022). The elements (2) and (4) rely on a different distribution and can be done using Hotelling's T^2 statistics, discussed in Subsection 1.3.3. When it comes to elements (5) and (6), we need to impose some assumptions on the complex error term of the model $\underline{\epsilon}_j$. The most popular assumption in statistics is that $\underline{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\epsilon)$, where $\mathbf{0}$ is the two-dimensional vector of zeroes and $\mathbf{\Sigma}_\epsilon$ is

the square two-by-two covariance matrix. The estimate of this matrix is typically obtained via the formula:

$$\hat{\boldsymbol{\Sigma}}_{\epsilon} = \frac{1}{n - \frac{k}{2}} \sum_{j=1}^{n} \boldsymbol{e}_j \boldsymbol{e}_j',$$

where k is the number of estimated parameters and $n - \frac{k}{2}$ is the number of degrees of freedom per part of the model, division by which reduces the bias in the estimation of the covariance matrix. The division of k by 2 is required to reflect the fact that the complex response variable consists of two parts and the whole complex-valued equation can be represented in a form of a system of two equations, where the same parameters are used in both parts. To make this clearer, consider the following model:

$$y_{r,1} + iy_{i,1} = (a_{r,1} + ia_{i,1})(x_{r,1} + ix_{i,1}) + \epsilon_{r,1} + i\epsilon_{i,1}.$$

It can be estimated on one observation and in that case will have zero degrees of freedom. Another way to look at this problem is to calculate the number of degrees of freedom per series, which in case of complex-valued model comes to two series resulting in the same $n - \frac{k}{2}$.

To calculate prediction intervals, we need to calculate the variance of the conditional mean of a complex linear model. In case of complex variables, we need both direct and conjugate variances conditional on the values of the available explanatory variables to get the full information about the distribution:

$$\begin{aligned} \mathrm{V}\left(\underline{y}_j | \underline{x}_{1,j}, \ldots, \underline{x}_{k-1,j}\right) =& \mathrm{V}\left(\underline{b}_0 + \underline{b}_1 \underline{x}_{1,j} + \underline{b}_2 \underline{x}_{2,j} + \cdots + \underline{b}_{k-1} \underline{x}_{k-1,j} + \underline{\epsilon}_j\right) \\ & \sum_{i=0}^{k-1} \underline{x}_{i,j} \underline{\tilde{x}}_{i,j} \mathrm{V}\left(\underline{b}_i\right) + 2 \sum_{i \neq l} \underline{x}_{i,j} \underline{\tilde{x}}_{l,j} \mathrm{cov}\left(\underline{b}_i, \underline{b}_l\right) + \sigma^2_{\underline{\epsilon}}, \end{aligned} \tag{4.30}$$

where $\mathrm{cov}(\cdot)$ is the conjugate covariance between the variables and $\underline{x}_{0,j} = 1 + i$. Similarly, we can calculate the direct conditional variance:

$$\mathcal{V}\left(\underline{y}_j | \underline{x}_{1,j}, \ldots, \underline{x}_{k-1,j}\right) = \sum_{i=0}^{k-1} \underline{x}_{i,j} \underline{\tilde{x}}_{i,j} \mathcal{V}\left(\underline{b}_i\right) + 2 \sum_{i \neq l} \underline{x}_{i,j} \underline{\tilde{x}}_{l,j} cov\left(\underline{b}_i, \underline{b}_l\right) + \varsigma^2_{\underline{\epsilon}} \quad , \tag{4.31}$$

where $cov(\cdot)$ is the direct covariance. The formulae (4.30) and (4.30) can then be used to calculate the conditional variances of real and imaginary parts of the response variable together with a conditional covariance between them. We can then either form a covariance matrix and produce a predictive ellipse using formulae discussed in Subsection 1.3.2 or consider each of the parts independently and create prediction intervals (also discussed in Subsection 1.3.2). Arguably, the latter is easier to work with than the former.

Finally, to produce confidence intervals, we can use the formulae (4.30) and (4.30), dropping the $\mathrm{V}\left(\underline{\epsilon}_j\right)$ and $\mathcal{V}\left(\underline{\epsilon}_j\right)$, which will then give us direct and conjugate conditional variances of the predicted value.

4.4.1 Demonstration in R

Building on the same example of artificial data from the previous section, we will see how predictions from the cLR can be made. To that extent, we will use the first 80 observations to estimate cLR using likelihood and then use the last 20 for prediction:

```
CLRArtificial <- clm(y~., dataArtificial,
                     loss="likelihood", subset=c(1:80))
```

After estimating the model, we produce forecasts with a 95% prediction interval (see Figure 4.1):

```
par(mfcol=c(2,1), mar=c(2,2,3,1))
predict(CLRArtificial, tail(dataArtificial,20),
        interval="prediction") |>
    plot()
```

Figure 4.1: Forecasts from cLR for the artificial data, given the values of explanatory variables in the holdout set.

If the confidence interval for the conditional mean is needed, then this is regulated with `interval="confidence"` parameter.

4.5 Dummy variables

In the real-valued regression analysis, dummy variables appear when a feature of an object can be measured only in a categorical scale. For example, we might be interested in sales of a red medium size t-shirt vs the sales of a blue small size one. In this case, the colour would be one of such characteristics (measured in the nominal scale), while the size would be the other one (in ordinal scale). In order to include such information in the regression model, a set of dummy variables is typically created. A dummy variable is the variable that equals to one, when the feature exists and zero otherwise. So, in our example, we would create dummy variables `colourRed`, `colourBlue`, and `colourGreen` to denote the first feature and `sizeSmall`, `sizeMedium`, and `sizeLarge` for the second one. These variables would be equal to one for the specific observations (t-shirts) in our data. And for obvious reasons, the t-shirt cannot be both red and blue or small and medium at the same time, so the respective variables will be self-exclusive. If you want to learn more, you can find some information on that in Chapter 13 of Svetunkov (2022).

In case of complex linear regression, it is possible to introduce dummy variables in the same way as in the conventional model, by adding real-valued variables. The model in this case becomes:

$$\underline{y}_j = \underline{\beta}_0 + \underline{\beta}_1 \underline{x}_{1,j} + \cdots + \underline{\beta}_{k-1} \underline{x}_{k-1,j} + \underline{\gamma}_1 d_{1,j} + \cdots + \underline{\gamma}_m d_{m,j} + \underline{\epsilon}_j, \qquad (4.32)$$

where $\underline{\gamma}_i$ is the complex parameter, $d_{i,j}$ is th i-th real-valued dummy variable and m is the number of dummy variables. In this case, each variable $d_{i,j}$ is multiplied by a complex coefficient, capturing the dummy effect of it on each part of the complex response variable. The effect of dummy variables on the response one is exactly the same as in the conventional real-valued regression - the intercept of the model, $\underline{\beta}_0$ will change by the value of $\underline{\gamma}_i$, when the variable $d_{i,j}$ equals to one.

It is also theoretically possible to encode a complex dummy variable, which would have two features in it at the same time, e.g. $colourRed + i \times sizeSmall$, where both `colourRed` and `sizeSmall` are dummy variables. The issue with this is that this encoding assumes very specific dynamic between the features and the response variable. In the example above, the sales of the first product will have $\gamma_{1,j} \times coolourRed_j - \gamma_{2,j} \times sizeSmall_j$, while the sales of the second one will have: $\gamma_{1,j} \times sizeSmall_j + \gamma_{2,j} \times coolourRed_j$. This means that the red colour should have exactly the same impact on sales of product one, as the small size has on the product two, while the effect of small size on product one is opposite to the effect of the red colour on the second product. It is difficult to find situations, where such effects would be meaningful, so we do not recommend that. Still, it is possible to formulate such a model and capture the qualitative features in a parsimonious way (in comparison with a simple introduction of variables to each equation).

Finally, if an interaction effect is needed (e.g. how the change of price on red product impacts its sales), this can be done in a similar way to the conventional real-valued regression. For example, here how it can be done for a variable $\underline{x}_{1,j}$:

$$\underline{y}_j = \underline{\beta}_0 + \underline{\beta}_1 \underline{x}_{1,j} + \cdots + \underline{\beta}_{k-1} \underline{x}_{k-1,j} + \underline{\gamma}_1 \underline{x}_{1,j} d_{1,j} + \underline{\epsilon}_j. \tag{4.33}$$

In this case, the specific complex effect of $\underline{x}_{1,j}$ on $\underline{y}_j$ will change when the dummy variable equals to one. The interaction effect between a complex variable $\underline{x}_{1,j}$ and a complex dummy variable $\underline{d}_{1,j}$ is also possible, but it will have a similar dilemma to the simple introduction of a dummy variable in a model.

4.5.1 Demonstration in R

The `clm()` function in the `complex` package in R supports categorical variables in a form of either a character vector, or a factor. Here is an example based on the same artificial data we used in Section 4.3:

```
set.seed(42)
# Create an artificial categorical variable
x3 <- rbinom(obs, 2, 0.5)
colour <- vector("character", obs)
colour[x3==0] <- "Red";
colour[x3==1] <- "Green";
colour[x3==2] <- "Blue";
# Add it to the data frame
# Note that it does not need to be factor()
dataArtificial <- data.frame(y=y, x1=x1, x2=x2, colour=colour)
```

We can then estimate the model using the same command as before:

```
clm(y~x1+x2+colour, dataArtificial, loss="likelihood") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: likelihood
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r  103.5161    17.1840    69.3946    137.6377 *
## (Intercept)_i -157.3312    12.5342  -182.2198   -132.4425 *
## x1_r             2.4141     0.0729     2.2693      2.5590 *
## x1_i             1.4995     0.0532     1.3939      1.6052 *
## x2_r             1.5754     0.0643     1.4478      1.7030 *
## x2_i            -0.6765     0.0469    -0.7695     -0.5834 *
## colourGreen_r    3.2016     2.4187    -1.6012      8.0044
## colourGreen_i    2.7011     1.7642    -0.8021      6.2043
## colourRed_r     -0.1795     2.7961    -5.7317      5.3727
## colourRed_i      1.9735     2.0395    -2.0763      6.0234
```

```
##
## Error covariance matrix:
##         e_r     e_i
## e_r 96.8123 52.4929
## e_i 52.4929 51.5081
##
## Sample size: 100
## Number of estimated parameters: 6.5
## Number of degrees of freedom: 93.5
## Information criteria:
##      AIC     AICc      BIC     BICc
## 1338.163 1339.217 1355.097 1357.524
```

The output above shows that R expanded the colour into a set of dummy variables and dropped one of them ("blue"). The complex coefficients for the dummy variables are all not significant on the 5% level, which in our case means that we cannot detect any strong relation between them and the response variable. This is because we did not use the variable in the data generation.

Chapter 5

Assumptions of Complex Linear Models

Similar to how it is done for the conventional real valued regression, we should discuss what assumptions are imposed on complex linear regression estimated using one of the methods discussed in the Section 4.2. We will group all regression assumptions to three parts (similar to how it was done by Svetunkov, 2023):

1. Model is correctly specified;
2. Residuals are independent and identically distributed (i.i.d.);
3. Explanatory variables are not correlated with anything but the response variable.

While the first two groups directly relate to the so called "true model", the last one refers to the estimation approaches discussed in the previous Chapter: if the latter are violated then the estimation procedure will lead to issues in estimates of parameters. We should also point out that many of the assumptions discussed in this Chapter are very similar to the assumptions in the conventional regression, which is why we do not plan to cover them in this monograph in detail, but rather to focus on their implications to complex-valued models. An reader interested in assumptions for the real-valued models is advised to read Chapter 15 of Svetunkov (2022).

Disclaimer: while in conventional econometrics, statistical tests are often used to check hypotheses of violation of specific assumptions, in this monograph we focus on the visual diagnostics. While tests might give ambiguous results and rely on a selected significance level, the visual inspection comes to expert judgment of an analyst. Yes, in some situations the subjectivity impacts the diagnostics, but at least it is not hidden behind p-values and a variety of outputs that might give contradictory results.

5.1 Model is correctly specified

This is one of the most general and most important groups of assumptions. It includes the following:

1. Model does not omit any important variables;
2. Model does not have redundant variables;
3. Variables are included in the model with appropriate transformations;
4. Residuals of the model do not contain outliers.

We briefly discuss all of them in this section.

5.1.1 Model does not omit any important variables

This assumption implies that we have all the variables that can substantially impact our response variable, and that we have included them in the model. If we do not do that then the estimates of parameters are known to be biased. In practice we always omit a lot of different variables that might potentially impact the response one, but do not have a large effect on it. Not including these variables is known not to cause serious issues in the estimation (**?**).

Formally speaking, the omitted variables should not be correlated with the ones that we include in the model, because otherwise the impact of the latter will not be captured correctly, causing bias in the estimates of parameters. It is not possible to test this assumption, so it can only be checked based on judgment of an analyst.

When it comes to complex-valued models, the relations become much more complicated and potentially non-linear than in the conventional models. The problem becomes then manifolds: we need to include the correct variables in the model, but also they need to be included correctly, because now any variable can be included in the real or in the imaginary part. Furthermore, the correlation in complex-valued models can be measured differently and implies a variety of effects between two variables.

Practically speaking, we consider two situations with omitted variables:

1. The model is estimated using OLS;
2. The model is estimated using CLS.

For the both cases, we assume that the true model is of the form:

$$\underline{\mathbf{y}} = \underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \underline{\mathbf{Z}}\underline{\boldsymbol{\gamma}} + \underline{\boldsymbol{\epsilon}}, \tag{5.1}$$

where the matrix $\underline{\mathbf{Z}}$ contains the variables omitted in the applied model and the vector $\underline{\boldsymbol{\gamma}}$ is the vector of the parameters for these variables in the true model.

We start with the effect of the omitted variables on the estimates of parameters using OLS. We refer to the derivations (4.20), which given (5.1) can be substituted

to:

$$\begin{aligned}\underline{\boldsymbol{b}}^{\mathrm{OLS}} &= \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\mathbf{y}} = \\ &\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\left(\underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \underline{\mathbf{Z}}\underline{\boldsymbol{\gamma}} + \underline{\boldsymbol{\epsilon}}\right) = \\ &\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\mathbf{Z}}\underline{\boldsymbol{\gamma}} + \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}} = \\ &\underline{\boldsymbol{\beta}} + \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\mathbf{Z}}\underline{\boldsymbol{\gamma}} + \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\boldsymbol{\epsilon}}.\end{aligned} \tag{5.2}$$

Taking the expectation of (5.2), we get:

$$\mathrm{E}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) = \underline{\boldsymbol{\beta}} + \mathrm{E}\left(\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\mathbf{Z}}\underline{\boldsymbol{\gamma}}\right) + 0, \tag{5.3}$$

with zero appearing because of the one of fundamental assumptions that the explanatory variables are not correlated with the error term in the true model. The expectation above shows that the estimates of the OLS parameters will be biased in case of omitted variables, and the bias will be proportional to the size of the $\underline{\mathbf{X}}'\underline{\mathbf{Z}}$ matrix, which consists of the covariances between the included and the omitted variables. The higher the correlation between these variables, the higher the bias will be. This agrees with the similar findings in the conventional real-valued econometrics.

When it comes to CLS, the logic is similar to the above. Based on the equations (4.21) and (5.1) we have:

$$\begin{aligned}\underline{\boldsymbol{b}}^{\mathrm{CLS}} &= \left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^{\top}\underline{\mathbf{y}} = \\ &\left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^{\top}\left(\underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \underline{\mathbf{Z}}\underline{\boldsymbol{\gamma}} + \underline{\boldsymbol{\epsilon}}\right) = \\ &\underline{\boldsymbol{\beta}} + \left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^{\top}\underline{\mathbf{Z}}\underline{\boldsymbol{\gamma}} + \left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^{\top}\underline{\boldsymbol{\epsilon}}\end{aligned}, \tag{5.4}$$

which simplifies (using similar logic as above) to:

$$\mathrm{E}\left(\underline{\boldsymbol{b}}^{\mathrm{CLS}}\right) = \underline{\boldsymbol{\beta}} + \mathrm{E}\left(\left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^{\top}\underline{\mathbf{Z}}\underline{\boldsymbol{\gamma}}\right), \tag{5.5}$$

Which also shows that the bias will increase with the increase of the values of the matrix $\underline{\mathbf{X}}^{\top}\underline{\mathbf{Z}}$. The main difference between the real-valued and the complex-valued econometrics is that the aforementioned matrices are complex and the respective covariances in OLS and CLS are respectively conjugate and direct. This means that:

1. if an omitted complex variable has high *direct correlation* with the included one but has the zero *conjugate correlation*, the OLS should give unbiased estimates of parameters, while the CLS would give the biased ones;
2. if an omitted complex variable has zero *direct correlation* with the included one but has high *conjugate correlation*, the CLS should give unbiased estimates of parameters, while the OLS would give the biased ones.

This property can be used in deciding which of the estimation methods to use if the researcher has some insights about the relation between the included and the omitted complex variables.

Furthermore, we can also see what happens with the direct and conjugate covariance matrices of parameters in case of omitted variables. In the case of the omitted variables the residuals become $\underline{\boldsymbol{\upsilon}} = \underline{\mathbf{Z}}\underline{\boldsymbol{\gamma}} + \underline{\boldsymbol{\epsilon}}$, implying that we estimate the following model instead of (5.1):

$$\underline{\mathbf{y}} = \underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\upsilon}}, \tag{5.6}$$

which means that we then make inference based on:

$$\underline{\mathbf{y}} = \underline{\mathbf{X}}\underline{\boldsymbol{b}} + \underline{\hat{\boldsymbol{\upsilon}}}. \tag{5.7}$$

The covariance matrix of the estimates of the parameter $\underline{\boldsymbol{\beta}}$ in this case comes to formulae discussed in Section 4.3:

$$\begin{aligned}
\mathrm{V}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) &= \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\tilde{\mathbf{X}}}\left(\underline{\mathbf{X}}^{\top}\underline{\tilde{\mathbf{X}}}\right)^{-1\prime}\sigma^2_{\underline{\upsilon}} \\
\mathcal{V}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) &= \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\varsigma^2_{\underline{\upsilon}} \\
\mathrm{V}\left(\underline{\boldsymbol{b}}^{\mathrm{CLS}}\right) &= \left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^{\top}\underline{\tilde{\mathbf{X}}}\left(\underline{\mathbf{X}}'\underline{\tilde{\mathbf{X}}}\right)^{-1\top}\sigma^2_{\underline{\upsilon}} \\
\mathcal{V}\left(\underline{\boldsymbol{b}}^{\mathrm{CLS}}\right) &= \left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\varsigma^2_{\underline{\upsilon}},
\end{aligned} \tag{5.8}$$

where $\sigma^2_{\underline{\upsilon}}$ is the conjugate and $\varsigma^2_{\underline{\upsilon}}$ is the direct variances of the residuals $\underline{\boldsymbol{\upsilon}}$. Now to better understand the effect of omitted variables on the covariance matrix of parameters, we should consider several situations:

1. Some parts of $\underline{\mathbf{Z}}$ are correlated with $\underline{\mathbf{X}}$, in which case applying the model (5.7) to the data will lead to the biased estimates of parameters, but would not have a large impact on the variance of the residuals of the model. The higher the correlation is, the lower the variances become in (5.8);
2. $\underline{\mathbf{Z}}$ is uncorrelated with $\underline{\mathbf{X}}$, which means that the effect of the omitted variable is not captured by the model and as a result the variance of the residual $\underline{\hat{\boldsymbol{\upsilon}}}$ will be inflated, increasing the standard errors of the estimates of parameters.

The case (1) corresponds to the classical problem of the omitted variables in econometrics, while the case (2) shows how the uncertainty about the estimates of parameters increases with omitted variables and might lead to the wider confidence intervals for the parameters.

5.1.2 Model does not have redundant variables

This is the situation opposite to the first one. It implies that we have included something that should not be there. Mathematically it can be represented by

two sets of equations, for the true and the applied models:

$$\begin{aligned}\underline{\mathbf{y}} &= \underline{\mathbf{X}}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\epsilon}} \\ \underline{\mathbf{y}} &= \underline{\mathbf{X}}\underline{\boldsymbol{b}} + \underline{\mathbf{X}}_{red}\underline{\boldsymbol{b}}_{red} + \underline{\boldsymbol{e}},\end{aligned} \tag{5.9}$$

where $\underline{\boldsymbol{b}}_{red}$ is the vector of the estimates of parameters for the redundant variables $\underline{\mathbf{X}}_{red}$ on a sample. It will not be zero in sample, because the values of $\underline{\mathbf{X}}_{red}$ can explain some randomness in $\underline{\mathbf{y}}$, thus reducing the size of variance of the residuals $\underline{\boldsymbol{e}}$ in comparison with the estimation of the correct model. This leads to the effect known as "overfitting" in forecasting.

The second equation in (5.9) can also be written in a compact form:

$$\underline{\mathbf{y}} = \underline{\mathbf{Z}}\underline{\boldsymbol{c}} + \underline{\boldsymbol{e}}, \tag{5.10}$$

where $\underline{\mathbf{Z}} = \begin{pmatrix}\underline{\mathbf{X}} & \underline{\mathbf{X}}_{red}\end{pmatrix}$ is the matrix that contains all variables under consideration, and $\underline{\boldsymbol{c}} = \begin{pmatrix}\underline{\boldsymbol{b}} \\ \underline{\boldsymbol{b}}_{red}\end{pmatrix}$. While it is hard to show explicitly using formulae of OLS and CLS, it is well known in statistics that the estimates of parameters in such model are unbiased, because the true $\underline{\boldsymbol{b}}_{red}$ is zero. At the same time, the inclusion of redundant variables increases the standard errors of parameters, because the covariance matrices from Section 4.3 would rely on different combinations of $\underline{\mathbf{Z}}$ and its transposition (direct and conjugate), which means that the final variances would be impacted by the variability of redundant variables $\underline{\mathbf{X}}_{red}$. The exact effect is difficult to capture and might depend on the direct/conjugate correlations between variables. We leave this task for future research.

5.1.3 Variables transformations

While the first two assumptions are universal for any statistical model, the third one has some special implications in case of cLR. This is because of the effect they have on the variables: transformations of separate parts of a complex variable are not equivalent to the transformations of the whole variable. For example, the logarithm of a complex variable $\underline{z}$ as shown in (1.12) is:

$$\ln \underline{z} = \ln r + i\phi,$$

which is not equivalent to the complex variable $\ln x_r + i \ln x_i$. In fact, there is no known transformation from $x_r + x_i$ to $\ln x_r + i \ln x_i$ except for transforming separately the real and imaginary parts of the variable. Still, logarithms can be used to linearise a non-linear complex model to simplify its estimation. For example, the following multiplicative model (based on discussion in Subsection 1.1.1):

$$\begin{aligned}y_r + iy_i =& (\beta_{0,r} + i\beta_{0,i}) \times (x_{1,r} + ix_{1,i})^{\beta_{1,r}+i\beta_{1,i}} \times \cdots \times \\ & (x_{k-1,r} + ix_{k-1,i})^{\beta_{k-1,r}+i\beta_{k-1,i}} \times (\epsilon_r + i\epsilon_i)\end{aligned}$$

can be linearised using natural logarithm to:

$$\begin{aligned}\ln(y_r + iy_i) =& \ln(\beta_{0,r} + i\beta_{0,i}) + (\beta_{1,r} + i\beta_{1,i}) \ln(x_{1,r} + ix_{1,i}) + \cdots + \\ & (\beta_{k-1,r} + i\beta_{k-1,i}) \ln(x_{k-1,r} + ix_{k-1,i}) + \ln(\epsilon_r + i\epsilon_i),\end{aligned}$$

which can then be estimated using OLS, CLS or likelihood maximisation as discussed in Section 4.2. On its own, the multiplicative models of complex variables have useful features, because they allow modelling highly non-linear processes due to the rotation as discussed in Section 1.1. For example, here how the linear increase of real and imaginary parts of a complex variable $\underline{x}$ leads to a non-linear transform of a variable $\underline{y} = \underline{x}^{0.5+0.5i}$:

```
x <- 1:100 + 1:100*1i
y <- x**(0.5+0.5i)
cplot(x, y)
```

Figure 5.1: Non-linear transformation of a complex variable that changes linearly.

Even if the imaginary part of $\underline{x}$ is zero (and as a result we deal with a real number, not a complex one), the complex power leads to highly non-linear transformation of the variable. This is a useful feature if the non-linearity is suspected in the data.

The diagnostics of correct transformations in cLR is challenging, but can be done by analysing the residuals. Consider the following example in R (using functions from the `complex` package in R):

```
# Set random seed for reproducibility
set.seed(41)
```

```
# Sample size
obs <- 1000
# Create the explanatory variable
x <- complex(real=rnorm(obs,100,10), imaginary=rnorm(obs,50,5))
# Generate parameters
b0 <- 1 - 1.5i
b1 <- 2.5 + 1.5i
# Generate error term from the complex normal distribution
e <- rcnorm(obs, 0, sigma2=0.25, varsigma2=0.16+0.09i)
# Generate the data using non-linear model
y <- exp(b0 + b1 * log(x) + e)
# Merge it to the matrix
complexData <- data.frame(y=y,x=x)
```

For demonstration purposes, we will first use a complex linear regression model on the data that was generated using a non-linear one:

```
# Apply a linear model
complexModel <- clm(y~x, complexData)
```

The issues of the model can be diagnosed using various plots. For example, here how the standardised residuals vs fitted would look for the model above (see Figure 5.2):

```
par(mfcol=c(1,2), mar=c(2,2,3,1))
plot(complexModel, which=2, main="")
```

Figure 5.2: Standardised residuals vs Fitted for the cLR on non-linear data.

The plots above demonstrate that there is a slight non-linear pattern in the residuals (especially for the real part of the model) and that their variances might not be constant. These are the indicators of a possible non-linearity in the data. In order to capture it correctly, we would need to transform both response and the explanatory complex variables and estimate the model in logarithms:

```
# Apply a model to log data
complexModelLogs <- clm(log(y)~log(x), complexData)
```

After which the same plot will look more reasonable, with residuals not exhibiting a strong u-shape anymore (see Figure 5.3):

```
par(mfcol=c(1,2), mar=c(2,2,3,1))
plot(complexModelLogs, which=2, main="")
```

Figure 5.3: Standardised residuals vs Fitted for the log-log cLR on non-linear data.

Finally, the `complex` package in R has `clog()` and `cexp()` functions, that support logarithmic and exponential transforms respectively for the separate real and imaginary parts of the complex variable. In this case, the functions require for both parts of the complex variable to be strictly positive. Here is an example of how they work:

```
# Generate data using exponent
y <- complex(real=exp(3 -0.02* c(1:100) + rnorm(100,0,0.1)),
             imaginary=exp(4 + 0.02* c(1:100) + rnorm(100,0,0.2)))

# Plot the data to see how it looks
```

```
layout(matrix(c(1,2,1,3),2,2))
par(mar=c(4,4,1,1))
plot(y)
plot(Re(y), type="l", xlab="Time")
plot(Im(y), type="l", xlab="Time")
```

Figure 5.4: Plot of the original non-linear data.

Figure 5.4 shows the generated data, which exhibits exponential trajectories (by design). After taking logarithms for individual real and imaginary parts we get (Figure 5.5):

```
yLog <- clog(y)

layout(matrix(c(1,2,1,3),2,2))
par(mar=c(4,4,1,1))
plot(yLog)
plot(Re(yLog), type="l", xlab="Time")
plot(Im(yLog), type="l", xlab="Time")
```

We see that the tendencies in the data became linear. Compare this with the conventional logarithm of a complex variable (Figure 5.6):

Figure 5.5: Plot of the data in logarithms.

Figure 5.6: Plot of the data in logarithms.

5.1.4 No outliers

Finally, the presence of outliers in the residuals typically implies that either there is an error in recording the data, or the model omits an important variable (e.g. a dummy variable for an external event). In that case the effect on estimates of parameters will be similar to the one discussed in Subsection 5.1.1: the estimates of parameters will be in general biased with some specific effects depending on the direct/conjugate correlation and the estimation method used.

The simplest way to detect outliers is to produce a diagnostic plot of fitted vs standardised residuals (similar to the one in Figure 5.3) and analyse those values that lie outside of the constructed confidence interval. If the value of an outlier cannot be explained (e.g. this is not a calendar event, this is not a promotion etc) then it can be removed or interpolated. In the latter case, creating a dummy variable (see Section 4.5) that equals to one on that specific observation and to zero on all the others is one of the simplest way of dealing with the outlier.

5.2 Residuals are i.i.d.

The next group of assumptions has five elements in it:

1. Residuals are homoscedastic;
2. No autocorrelation in residuals;
3. Expectation of residuals is zero (no matter what);
4. Residuals follow an assumed distribution;
5. Residuals follow a circular distribution.

Some of these assumptions are universal for statistical models, while the others (e.g. the last one) are specific to cLR.

5.2.1 Homoscedastic residuals

The first assumption implies that both direct and conjugate variances of the residuals are constant and do not change with any changes of variables. This follows directly from the formulae for the direct and conjugate covariances from Section 4.3: if either of the covariances is not constant, and we use the formulae for calculating the standard errors of parameters, we will obtain averaged-out values, that would be lower than needed for some observations and higher than needed for the others.

In case of the formulation of cLR as a vector model and the likelihood estimation, the assumption implies that the covariance matrix of residuals Σ_ϵ stays the same no matter what. This assumption aligns well with a similar assumption of conventional multivariate models, such as VAR (**?**) or VETS (**?**).

The homoscedasticity assumption can be typically diagnosed visually by producing the plot of absolute residuals vs fitted. In case of the cLR, this can be modified to producing plots for each part of the complex residuals. Figure

5.7 shows how the residuals of the linear model applied to the non-linear data (example from Subsection 5.1.3) look:

```
par(mfcol=c(1,2), mar=c(2,2,3,1))
plot(complexModel, which=4, main="")
```

Figure 5.7: Absolute residuals vs fitted

The plots in Figure 5.7 show that the variability of residuals (and the local mean) increases with the increase of fitted values. This is a signal of the heteroscedasticity in the data. Some non-linear transformations (e.g. taking logarithm) typically resolve the issue.

5.2.2 No autocorrelation in residuals

This assumption only applies to time series data. In case of the real-valued models, it means that the residuals in the past should not impact the ones in the future. Typically, this effect appears because of the wrong specification of the model (e.g. the appropriate transformations are not done, or some autoregressive elements are missing). In case of the cLR, the idea is very similar, but now we are talking about complex relations between the residuals, which could arise, again, because of the wrong transformations or because of missing elements (e.g. complex autoregression, discussed later in this book).

To diagnose this, we can analyse the autocorrelation functions. In the real-valued domain, Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are typically used. In the complex-valued domain, these transform to the direct and conjugate complex ACF/PACF (or cACF/cPACF, based on the

correlations discussed in Chapter 3), giving an analyst four instruments instead of two (Yao et al., 2010; Yao and Lu, 2011; Zhang et al., 2022). Furthermore, using the idea with correlation on MDS (see Section 3.5), we can calculate ACF/PACF based on Pearson's correlation between the complex variables reduced to real ones.

In general, the idea with direct and conjugate ACF/PACF for complex variables is similar to the one for the conventional ACF/PACF for the real ones. The complex ACF measures the correlations between the variable on observation t and the one with lag $\tau = \{1, \dots, T\}$, where T is the maximum lag we calculate the correlation for. If we calculate the conjugate correlations, we obtain a vector of values $\rho_1, \dots, \rho_T$ (specific formulae are discussed in Section 3.3). In case of the direct one (see Section 3.4), we get a vector $\varrho_1, \dots, \varrho_T$. In both cases the index j in subscription refers to the specific lag, so, for example, ρ_1 will measure the conjugate autocorrelation for lag 1, i.e. conjugate correlation between $\underline{y}_t$ and $\underline{y}_{t-1}$.

When it comes to PACF, the logic is to estimating recursively regression models of the $\underline{y}_t$ from all the previous $\underline{y}_{t-j}$ for $j = \{1, \dots, J\}$ for each J from 1 to T and collecting respective estimates of parameters in front of $\underline{y}_{t-J}$. This way, the coefficients are clear of the interim effects and we get a clear correlation between $\underline{y}_t$ and $\underline{y}_{t-J}$. In case of the direct complex PACF, we use the CLS, for the conjugate one we use OLS and for the Pearson's, we do MDS and then calculate the conventional PACF (see Chapter 3 for discussion of different correlation coefficients).

In R, there are functions `cacf()` and `cpacf()` in the `complex` package that implement respective complex ACF/PACF with a parameter `method` specifying, which of the correlations to calculate. To demonstrate how they work, we generate a data from a complex autoregressive model of order one and apply it to the data (note that for diagnostic purposes, the functions below should be applied to the residuals of the model):

```
# Generate complex data with an autocorrelation of order 1
# Random seed for reproducibility
set.seed(41)
# Number of observations
obs <- 100
# Vector of the final variable
y <- vector("complex", obs)
# Initial values for the first observation
y[1] <- 300 + 250i
for(i in 2:obs){
    y[i] <- 100 + 200i + (0.5 - 0.25i) * y[i-1] +
        complex(real=rnorm(1,0,10), imaginary=rnorm(1,0,5))
}
```

The conjugate ACF (due to its nature) is represented by a set of real values (Figure 5.8):

```
cacf(y, method="conjugate", plot=TRUE, main="")
```

Figure 5.8: Conjugate complex autocorrelation function

The plot in Figure 5.8 shows that the autocorrelation for lag one is statistically different from zero on the 5% significance level (the latter is regulated by `level` parameter in the function). The non-rejection region on the plot is costructed based on the Folded t-distribution (which is an approximation of the square root of convolution of two Fisher distributions) and helps testing the null hypothesis that autocorrelation coefficient for each lag equals to zero (with the alternative being not equal). The value of the coefficient itself does not provide any specific information about the autoregressive relation in the data, but merely shows that the correlation between variables is mediocre.

In contrast, the conjugate PACF provides more details and shows the specific values for the real and the imaginary parts of the coefficient in the process (see Figure 5.9):

```
cpacf(y, method="conjugate", plot=TRUE, main="")
```

Figure 5.9 shows three plots: real vs imaginary complex ACF plot at the top and two separate plots for the real and the imaginary parts of the complex PACF at the bottom. As shown in the plots both the real and the imaginary parts of the complex coefficient are statistically different from zero on the 5% significance level for the first lag and not significant for the rest of the lags. This agrees with the data generated process that we used. Furthermore, the real part of the cPACF is close to the true value of 0.5, while the imaginary one is close to the

Figure 5.9: Conjugate complex partial autocorrelation function

-0.25, which were used in the data generation.

Similarly, we can analyse the direct complex ACF of a variable, which will generate a plot similar to the ones in Figure 5.9 (see Figure 5.11):

```
cacf(y, method="direct", plot=TRUE, main="")
```

While the dynamics of direct cACF is similar to the one of the conjugate cPACF, they show different things: the latter is clear of the interim relations between variables, while the former shows the specific correlations between the actual value and each of its lags. Similarly to how we did that before, we can also produce the direct cPACF (Figure 5.11):

```
cpacf(y, method="direct", plot=TRUE, main="")
```

Because we generated the data from the complex autoregressive model of order one, the cACF and cPACF give us roughly the same message: the lag one is significant on the 5% level, while the rest of values are not. Note however that the direct cACF/cPACF shows that there are other lags that are significantly different from zero. This appears by chance and should be ignored.

Finally, we can generate ACF/PACF for the variable after its MDS transformation (Figure 5.12):

```
par(mfcol=c(2,1), mar=c(2,2,2,1))
cacf(y, method="pearson", plot=TRUE)
```

Figure 5.10: Direct complex autocorrelation function

Figure 5.11: Direct complex partial autocorrelation function

```
cpacf(y, method="pearson", plot=TRUE)
```

Figure 5.12: ACF and PACF of the complex variable after the MDS.

Similarly to the conjugate cACF, the plots in Figure 5.12 show that there is a significant spike for the lag one, implying that the variable is autocorrelated. However, they do not give any specific information about the value of the parameter (like conjugate and direct pACF do).

Furthermore, there is another presentation for complex ACF/PACF - in the polar coordinates one. In that case instead of plotting separate real and imaginary parts of the function, we produce the magnitude (absolute value) and the angle (argument) of the complex variables. In that presentation, the higher the magnitude is, the higher overall the value of a complex variable is. Figure 5.13 demonstrates how this looks for the conjugate cPACF:

```
cpacf(y, method="conjugate", plot=FALSE) |>
    plot(2)
```

The distribution of the absolute value of a bivariate normal complex variable follows Hoyt's distribution, which can be used for the confidence interval construction for the plot above. However, the derivations become complicated and we do not provide them here (we might provide them in the future editions of this monograph).

Finally, one can analyse the ACF/PACF of the separate real and imaginary

Figure 5.13: Complex ACF in polar coordinates.

parts of the residuals, thus ignoring the potential dynamics between them. The analysis in this case becomes trivial and has been discussed in many textbooks and monographs (for example, see Svetunkov, 2023). In R, if a model is estimated using the `clm()` function from the `complex` package, the `plot()` method supports producing ACF/PACF for each separate variable. For example, for the same model from the Subsection 5.1.3, we have (see Figure 5.13):

```
par(mfcol=c(2,2))
plot(complexModel,which=c(10,11))
```

The residuals of this model show that there are no important lags on the 5% significance level. The fact that there are some correlations with values lying outside of the non-rejection region can be explained by random chance. We should take such cases with a pinch of salt not to overcomplicate a model.

In addition, we can do more thorough diagnostics based on the complex ACF/PACF (as discussed above) for any complex-valued model. In that case we need to extract residuals and then apply `cacf()` or `cpacf()` functions. For example, using this code:

```
residuals(complexModel) |>
    cacf(method="direct")
```

After doing the diagnostics of the residuals of a cLR, we can change its form

Figure 5.14: ACF/PACF for the real (Series 1) and imaginary (Series 2) parts of the complex residuals.

by introducing complex AR (autoregression) or MA (moving average) elements, depending on what lags seem significant on the selected level. The complex AR/MA models are outside of the scope of this monograph and are not discussed in detail here. It suffices to say that in order to construct the complex AR, one needs to generate lagged response variable and add it to the matrix with the explanatory variables. The estimation of such model should be trivial.

5.2.3 Expectation of residuals is zero

While many other assumptions can be diagnosed in one way or another, this assumption in many cases cannot be. For example, if we use OLS, the unconditional mean of the residuals will be equal to zero by design, which does not provide any useful information about the "true" model. However, this assumption can be analysed conditionally, e.g. conditional on the fitted values:

$$\mathrm{E}(\underline{e}_j|\underline{\mathbf{x}}_j) = 0, \tag{5.11}$$

meaning that the expectation of complex residual conditional on the values of explanatory variables should be equal to zero. If the conditional mean is not constant and not zero, then we can conclude that some important elements were omitted in the model. In a way, the assumption becomes similar to the assumption about the omitted variables, wrong transformations, and/or autocorrelated residuals. To that extent, we do not discuss it in more detail

here.

5.2.4 Residuals follow an assumed distribution

In the classical real-valued regression, this assumption is not considered to be crucial, and if a model is estimated using OLS, it typically comes to the "residuals follow the Normal distribution". In context of complex variables, the equivalent assumption would be that the complex residuals follow the Complex Normal distribution discussed in Subsection 1.3.1. If the model is formulated using matrix notations (4.5), the assumption transforms to "Multivariate Normal" one, discussed in Section 1.3.2. In statistics, this assumption is typically considered as not critical in comparison with the others. This is because it can be shown that in case of OLS/CLS even when it does not hold, but all the others do, the estimates of parameters should be unbiased, efficient and consistent. However, this assumption becomes essential when prediction interval needs to be generated from the model. In that case we want to capture the uncertainty about the data correctly, and if the residuals do not follow the Normal distribution, the prediction interval will be miscalibrated.

One of the simplest ways of checking the assumption is by producing QQ-plots for each part of the complex residuals. This approach ignores the multivariate nature of the residuals (e.g. then can be correlated), but given that each part of the c.r.v. following Complex Normal distribution is Normal as well, this should suffice for diagnostic purposes.

In R, this can be done using the following command (for the model discussed in Subsection 5.1.3, see Figure 5.15):

```
par(mfcol=c(1,2))
plot(complexModelLogs,which=6)
```

Similarly to the conventional diagnostics related to distributional assumptions, the QQ-plots in Figure 5.15 can be interpreted as follows. If all the points (empirical quantiles) lie on or close to the lines (theoretical quantiles) then the empirical distribuion looks similar to the theoretical one. In the example in Figure 5.15, we see that the points are indeed close to the lines for both real (called "Series 1") and imaginary (called "Series 2" in the plot) parts of the residuals. There is a slight deviation in the right tail of the imaginary part, but it is not substantial. Based on this diagnostics we can conclude that the residuals look Normal or at least very close to it.

5.2.5 Residuals follow a circular distribution

Finally, in signal processing literature, there is an assumption that residuals follow so called "circular distribution". This means that the covariance between the residuals equals to zero and that the variance of the real part is equal to the variance of the imaginary one. This also implies that the direct variance of residuals is equal to zero. While this might be a suitable assumption in

Figure 5.15: QQ-plots for the real (Series 1) and imaginary (Series 2) parts of the residuals.

signal processing, there is no good rationale for this to hold universally for all cLR models. In fact, the power of the cLR comes from different variances and potentially non-zero covariance between the residuals, because in that case the model can capture a complex dynamics between the parts of the complex variable. So, we suggest ignoring this assumption, especially given that it does not effect estimates of parameters of the forecasts from the model.

However, it is important to note that the used estimation methods might imply some parts of this assumption. For example, OLS minimises the sum of squared residuals of the real and imaginary parts, completely neglecting the covariance between them (see Subsection 2.2.1). At the same time, the CLS (Subsection 2.2.2) ignores the specific sizes of the squared residuals and focuses on making their squares equal, minimising at the same time the covariance between them. If we believe that the residuals might have a more complicated distribution (elliptic Normal) then MLE (Subsection 2.2.3) might be the best option for the model estimation.

5.3 Explanatory variables are not correlated with anything but the response variable

The two assumptions in this group relate to the estimation of model rather than anything else:

1. No endogeneity

2. No multicollinearity

The first assumption comes to the idea that a linear regression model can only capture a one-directional relation, where an explanatory variable impacts the response variable. If the relation in the real life is bi-directional then the estimates of parameters of a regression model will be biased. In that case, the relation needs to be somehow made one-directional, for example by substituting the explanatory variable that causes a problem with something similar that is not impacted by the response variable. In case of the complex linear regression, the situation would be similar to the conventional one. We do not plan to discuss how to solve the problem in case of cLR in this monograph. But we believe that the methods from econometrics of real-valued models should be widely applicable here as well (e.g. using proxies or instrumental variables).

As for the multicollinearity issue, in the real-valued statistics, it arises when least squares or maximum likelihood is used for estimation. It is, in a way, a technical issue: if the explanatory variables are linearly related it becomes difficult to split the impact of each of the explanatory variables on the response variable and thus challenging to correctly capture the relations between the explanatory variables and the response one.

In case of an OLS applied to a conventional real-valued regression model, the issue comes to inverting the matrix $(\mathbf{X}'\mathbf{X})^{-1}$. If some of variables in $\mathbf{X}$ are strongly linearly related, the determinant of the matrix becomes close to zero and thus the inversion becomes challenging. Even if it is still possible to invert the matrix, the estimates of parameters become inefficient because small changes in the related explanatory variables may lead to substantial changes in the inversion.

In case of cLR, the situation is similar, but with some specific features. First, it comes to the estimation methods, with the following options (as discussed in Section 4.2):

- OLS: $\underline{\boldsymbol{b}}^{OLS} = \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}'\underline{\mathbf{y}}$;
- CLS: $\underline{\boldsymbol{b}}^{CLS} = \left(\underline{\mathbf{X}}^{\top}\underline{\mathbf{X}}\right)^{-1}\underline{\mathbf{X}}^{\top}\underline{\mathbf{y}}$;
- Likelihood: no analytical solution, comes to the minimisation of the Generalised Variance $|\hat{\boldsymbol{\Sigma}}_{\epsilon}|$.

If the OLS is used for parameters estimation then a potential issue might arise from the inversion of $\left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1}$. In one of the special cases, this comes to the conjugate covariance between variables. If it is too high in comparison to the conjugate variances, then we would face the multicollinearity issue, which would make the estimates of parameters inefficient. More generally speaking, if a complex linear regression between an explanatory variable and other explanatory variables estimated using OLS captures the relations between variables well (i.e. having a high coefficient of determination), then we might be facing a multicollinearity problem.

Similar problems might appear in case of CLS, but they come to a slightly different inversion, $\left(\underline{\mathbf{X}}^\top \underline{\mathbf{X}}\right)^{-1}$ and thus relates to direct covariances rather than the conjugate ones.

And when it comes to the likelihood, due to the model formulation, the conventional real-valued relations between explanatory variables should be considered when multicollinearity is suspected.

All of this means that a choice of an estimator can be dictated by the strength of the conventional Pearson's, direct and/or conjugate correlations (and more widely conventional linear model and/or complex linear model for explanatory variables estimated using OLS/CLS).

In addition to the estimation issues, there might be some related to the calculation of the covariance matrix of parameters (discussed in Section 4.3), where we need both direct and conjugate covariance matrices:

- OLS:
  - Conjugate covariance matrix: $\mathrm{V}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) = \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1} \underline{\mathbf{X}}'\underline{\tilde{\mathbf{X}}} \left(\underline{\mathbf{X}}^\top \underline{\tilde{\mathbf{X}}}\right)^{-1\prime} \sigma^2_{\underline{\epsilon}}$
  - Direct covariance matrix: $\mathcal{V}\left(\underline{\boldsymbol{b}}^{\mathrm{OLS}}\right) = \left(\underline{\mathbf{X}}'\underline{\mathbf{X}}\right)^{-1} \varsigma^2_{\underline{\epsilon}}$;
- CLS:
  - Conjugate covariance matrix: $\mathrm{V}\left(\underline{\boldsymbol{b}}^{\mathrm{CLS}}\right) = \left(\underline{\mathbf{X}}^\top \underline{\mathbf{X}}\right)^{-1} \underline{\mathbf{X}}^\top \underline{\tilde{\mathbf{X}}} \left(\underline{\mathbf{X}}'\underline{\tilde{\mathbf{X}}}\right)^{-1\top} \sigma^2_{\underline{\epsilon}}$
  - Direct covariance matrix: $\mathcal{V}\left(\underline{\boldsymbol{b}}^{\mathrm{CLS}}\right) = \left(\underline{\mathbf{X}}^\top \underline{\mathbf{X}}\right)^{-1} \varsigma^2_{\underline{\epsilon}}$;
- Likelihood: no closed form, comes to calculating Hessian, which involves the inversion of a matrix based on the matrix of explanatory variables.

In this case, the linear relation between explanatory variables (in one form or the other) will impact the elements of the final covariance matrix inflating them and thus leading to the higher uncertainty of the estimates of parameters. Given that we need both direct and conjugate covariance matrices for the calculation of the standard errors of the individual parameters, the choice of estimation method would not resolve the multicollinearity issue. We can expect that the standard errors of parameters will be higher than in the situation without the multicollinearity, no matter what estimation method we use.

Note that the standard issues that arise in conventional regression analysis due to linear relation between variables, might not apply to the cLR. For example, if a real part of one variable is strongly correlated with either a real or an imaginary part of the other one, the model might still be estimable and would not suffer from multicollinearity. The effect of multicollinearity in cLR is much more pronounced, when the complex variables are linearly related, rather than their separate parts.

The diagnostics of the multicollinearity in cLR can be done by calculating the correlation matrix between the explanatory variables to see whether there is a strongly related couple of them. Given the nature of the model, this means

that we might need to estimate and analyse the direct and conjugate correlation matrices together with the conventional one after transforming the complex variables to the vector of real ones.

5.3.1 Demonstration in R

For demonstration purposes we generate data from a model:

$$\underline{y} = \underline{\beta}_0 + \underline{\beta}_1 \underline{x}_1 + \underline{\beta}_2 \underline{x}_2 + \underline{\epsilon}$$

for several scenarios:

1. $\underline{x}_1$ is strongly correlated with $\underline{x}_2$;
2. $\mathcal{R}(\underline{x}_1)$ is strongly correlated with $\mathcal{R}(\underline{x}_2)$;
3. $\mathcal{R}(\underline{x}_1)$ is strongly correlated with $\mathcal{I}(\underline{x}_2)$.

These three cases should cover the main possible issues in the model, and we will see what different estimators give in terms of parameters and their standard errors. We do not conduct a proper simulation experiment, but instead focus on a demonstration for one data series.

The R code below creates three datasets for the three respective scenarios (the coefficients have been selected arbitrarily):

```
set.seed(41)
# Sample size
obs <- 100
# Generate parameters
b0 <- 100 - 150i
b1 <- 2.5 + 1.5i
b2 <- 1.5 - 0.75i
a1 <- 1.3 + 1.5i
# Create the explanatory variables
x1 <- rcnorm(obs, mu=100+50i, sigma2=125, varsigma2=75)
# Generate error term from the complex normal distribution
e <- rcnorm(obs, mu=0, sigma2=40^2, varsigma2=-500)
# The second variable

## Scenario 1
x2 <- a1*x1 + rnorm(obs,0,1)
# Generate the response variable
y <- b0 + b1 * x1 + b2 * x2 + e
# Form a data frame with the variables
data1 <- data.frame(y=y,x1=x1,x2=x2)

## Scenario 2
x2 <- complex(real=Re(x1)*Re(a1) + rnorm(obs,0,1),
              imaginary=rnorm(obs,10,5))
```

```
# Generate the response variable
y <- b0 + b1 * x1 + b2 * x2 + e
# Form a data frame with the variables
data2 <- data.frame(y=y,x1=x1,x2=x2)

## Scenario 3
x2 <- complex(real=rnorm(obs,10,5),
              imaginary=Re(x1)*Re(a1) + rnorm(obs,0,1))
# Generate the response variable
y <- b0 + b1 * x1 + b2 * x2 + e
# Form a data frame with the variables
data3 <- data.frame(y=y,x1=x1,x2=x2)
```

Scenario 1

First, we do diagnostics based on correlation matrices that we discussed earlier in this section, including conjugate, direct and MDS-based correlations and the conventional covariance matrix of parameters.

We start with the conjugate complex correlation matrix:

```
ccor(data1, method="conjugate") |> round(3)
```

```
##        y    x1    x2
## y  1.000 0.855 0.853
## x1 0.855 1.000 0.999
## x2 0.853 0.999 1.000
```

The output above shows that $\underline{x}_1$ is strongly correlated with $\underline{x}_2$ - the correlation coefficient between them is 0.999. This agrees with how we generated the data, and we can expect issues when estimating the model and/or generating the standard errors of parameters.

The second output is for the direct complex correlation matrix:

```
ccor(data1, method="direct") |> round(3)
```

```
##               y           x1            x2
## y  1.000+0.000i 1.001-0.081i  0.998-0.078i
## x1 1.001-0.081i 1.000+0.000i  1.000+0.002i
## x2 0.998-0.078i 1.000+0.002i -1.000+0.000i
```

While some correlation coefficients in the output above are hard to interpret (because some of them are above one), the real part of the one between the explanatory variables equals to one (after rounding), which also indicates a strong relation between them.

We also calculate Pearson's correlation matrix for the MDS-transformed complex variables to get another look at the problem:

```
ccor(data1, method="pearson") |> round(3)
```

```
##        y    x1    x2
## y  1.000 0.897 0.895
## x1 0.897 1.000 0.999
## x2 0.895 0.999 1.000
```

And as we can see from the output above, the explanatory variables $\underline{x}_1$ and $\underline{x}_2$ are indeed strongly linearly related.

Finally, we calculate the covariance matrix of parameters between individual variables:

```
covar(data1) |> cov2cor() |> round(3)
```

```
##         y_r   y_i   x1_r   x1_i   x2_r  x2_i
## y_r   1.000 0.389  0.888 -0.211  0.856 0.719
## y_i   0.389 1.000  0.524  0.553  0.149 0.713
## x1_r  0.888 0.524  1.000 -0.017  0.850 0.905
## x1_i -0.211 0.553 -0.017  1.000 -0.537 0.411
## x2_r  0.856 0.149  0.850 -0.537  1.000 0.547
## x2_i  0.719 0.713  0.905  0.411  0.547 1.000
```

If shows that there is a strong correlation between $\mathcal{R}(\underline{x}_1)$ and $\mathcal{R}(\underline{x}_2)$ and $\mathcal{I}(\underline{x}_2)$, while the correlation between the imaginary part of $\underline{x}_1$ with the other variables is not as strong.

All of the outputs above tell us that if we want to fit the cLR, we will face difficulties due to the strong multicollinearity caused by the strong correlation between the explanatory variables. We estimate the model using the three different methods and produce their summary outputs. We start with OLS:

```
clm(y~., data1, loss="OLS") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: OLS
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r   81.2499    27.8800    25.9158    136.5840 *
## (Intercept)_i -162.3298    35.2247  -232.2410    -92.4185 *
## x1_r             4.2390     5.7400    -7.1532     15.6313
## x1_i             8.9223     7.2521    -5.4712     23.3157
## x2_r            -1.8366     2.8949    -7.5822      3.9091
## x2_i            -2.6157     3.6576    -9.8750      4.6436
##
## Error covariance matrix:
##          e_r       e_i
```

```
## e_r 687.7480  108.1091
## e_i 108.1091 1097.8374
##
## Sample size: 100
## Number of estimated parameters: 3
## Number of degrees of freedom: 97
```

The output above shows that the estimates of parameters are wrong: the complex parameter for the first variable should be $2.5 + 1.5i$, while for the second one should be $1.5 - 0.75i$. Furthermore, the standard errors of parameters are high, leading to the high uncertainty about the parameters, so that we cannot be sure whether the true effect is positive or negative for the estimated model.

The CLS gives a very similar picture:

```
clm(y~., data1, loss="CLS") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: CLS
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r   89.0089    54.0839   -18.3327    196.3505
## (Intercept)_i -198.3455    61.3562  -320.1206    -76.5704 *
## x1_r             4.3945     6.2816    -8.0728     16.8617
## x1_i             8.8479     6.8397    -4.7271     22.4228
## x2_r            -1.7130     2.9266    -7.5214      4.0955
## x2_i            -2.4582     3.6720    -9.7461      4.8297
##
## Error covariance matrix:
##          e_r       e_i
## e_r 696.9762  106.4108
## e_i 106.4108 1101.1012
##
## Sample size: 100
## Number of estimated parameters: 3
## Number of degrees of freedom: 97
```

However, we should note that the standard errors of parameters are in some cases higher than in the OLS, but not everywhere. For example, the parameter for $\mathcal{I}(\underline{x}_1)$ estimated using CLS has a lower standard error than in the case of OLS. Nonetheless, the estimates of parameters are clearly biased and not efficient.

Finally, we estimate the model using likelihood:

```
clm(y~., data1, loss="likelihood") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: likelihood
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r   81.2499    28.0973    25.4734    137.0263 *
## (Intercept)_i -161.9716    35.5009  -232.4450    -91.4981 *
## x1_r             4.2390     5.7847    -7.2443     15.7224
## x1_i             8.9223     7.3090    -5.5869     23.4314
## x2_r            -1.8383     2.9175    -7.6299      3.9533
## x2_i            -2.6157     3.6863    -9.9334      4.7020
##
## Error covariance matrix:
##          e_r       e_i
## e_r 698.5102  109.8689
## e_i 109.8689 1115.1225
##
## Sample size: 100
## Number of estimated parameters: 4.5
## Number of degrees of freedom: 95.5
## Information criteria:
##      AIC     AICc      BIC     BICc
## 1922.373 1922.897 1934.096 1935.303
```

Because we generated the data with the uncorrelated errors, the output of the likelihood is very similar to the one from the OLS. In the other situation it might have produced different estimates of parameters.

All of the above happened because of the strong multicollinearity in the data: the complex variables $\underline{x}_1$ and $\underline{x}_2$ were strongly correlated by design, which caused the aforementioned issues.

The Scenario 1 demonstrates the behaviour of estimators as expected: we would not expect to estimate the models successfully, when the complex variables are so strongly correlated.

As a side note, in this scenario, we would not be able to get efficient and unbiased estimates of parameters for the individual real-valued models for $\mathcal{R}\left(\underline{y}\right)$ and $\mathcal{I}\left(\underline{y}\right)$ with this data, because the covariance matrix with individual variables demonstrated a high correlation between $\mathcal{R}\left(\underline{x}_1\right)$ and $\mathcal{I}\left(\underline{x}_2\right)$.

Scenario 2

In this scenario, we have strong correlation between real parts of explanatory variables, but not between the complex ones. We would expect the real-valued regression to fail in this case, but the cLR might still be estimable. We start the demonstration with the same set of outputs as before:

```
ccor(data2, method="conjugate") |> round(3)
```

```
##        y    x1    x2
## y  1.000 0.723 0.687
## x1 0.723 1.000 0.791
## x2 0.687 0.791 1.000
```

The conjugate correlation above shows that there is a linear relation between $\underline{x}_1$ and $\underline{x}_2$, but it is not sever, so we should be able to estimate the model and get reasonable estimates of parameters.

```
ccor(data2, method="direct") |> round(3)
```

```
##               y           x1           x2
## y  1.000+0.000i 1.152-0.146i 1.219-0.062i
## x1 1.152-0.146i 1.000+0.000i 1.330+0.012i
## x2 1.219-0.062i 1.330+0.012i 1.000+0.000i
```

The direct correlation agrees with the conjugate one in general, but the fact that the correlation between $\underline{x}_1$ and $\underline{x}_2$ has the real part above one, implies that the relation between the variables is stronger than expected by the previous output.

```
ccor(data2, method="pearson") |> round(3)
```

```
##        y    x1    x2
## y  1.000 0.854 0.858
## x1 0.854 1.000 0.997
## x2 0.858 0.997 1.000
```

Pearson's correlations on the MDS variables shows that there is a very strong relation between $\underline{x}_1$ and $\underline{x}_2$, although we cannot say for sure what causes it.

```
covar(data2) |> cov2cor() |> round(3)
```

```
##         y_r   y_i   x1_r   x1_i   x2_r   x2_i
## y_r   1.000 0.151  0.857 -0.135  0.861 -0.012
## y_i   0.151 1.000  0.161  0.337  0.163  0.112
## x1_r  0.857 0.161  1.000 -0.017  0.997 -0.016
## x1_i -0.135 0.337 -0.017  1.000 -0.015 -0.136
## x2_r  0.861 0.163  0.997 -0.015  1.000 -0.022
## x2_i -0.012 0.112 -0.016 -0.136 -0.022  1.000
```

Finally, as expected, the correlations between the specific parts of variables shows that $\mathcal{R}(\underline{x}_1)$ and $\mathcal{R}(\underline{x}_2)$ have almost functional relation. This might mean that we could face difficulties estimating the model.

The model estimated using OLS produces the following output:

```
clm(y~., data2, loss="OLS") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: OLS
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r   58.7490    32.3010    -5.3596    122.8575
## (Intercept)_i -182.8910    41.0770  -264.4174   -101.3646 *
## x1_r             2.7010     0.3975     1.9121      3.4900 *
## x1_i             0.9558     0.5055    -0.0475      1.9591
## x2_r             1.5106     0.3245     0.8666      2.1546 *
## x2_i            -0.1824     0.4126    -1.0014      0.6365
##
## Error covariance matrix:
##          e_r       e_i
## e_r 678.7893  108.4556
## e_i 108.4556 1097.7400
##
## Sample size: 100
## Number of estimated parameters: 3
## Number of degrees of freedom: 97
```

The parameters seem to be less efficient than they would have been without the strong correlation between parts of explanatory variables, but notably all signs are correct and the specific values are not too far from the ones used in the data generation ($\underline{\beta}_1 = 2.5 + 1.5i$, $\underline{\beta}_2 = 1.5 - 0.75i$). But it would be interesting to compare the output from OLS with the one from the CLS:

```
clm(y~., data2, loss="CLS") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: CLS
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r   46.5476    37.5007   -27.8808    120.9760
## (Intercept)_i -175.9491    47.1085  -269.4465    -82.4518 *
## x1_r             2.8511     0.5146     1.8299      3.8724 *
## x1_i             0.9655     0.4235     0.1250      1.8059 *
## x2_r             1.4826     0.4179     0.6531      2.3120 *
## x2_i            -0.2975     0.3625    -1.0170      0.4221
##
## Error covariance matrix:
##          e_r       e_i
## e_r 681.2265  110.0717
## e_i 110.0717 1099.2501
##
```

```
## Sample size: 100
## Number of estimated parameters: 3
## Number of degrees of freedom: 97
```

The estimates of parameters in CLS are similar to the OLS, but there are several interesting observations:

1. The standard errors for the real parts of parameters in CLS are higher than the similar estimates in OLS;
2. The standard errors for the imaginary parts of parameters in CLS are lower than the respective ones in OLS;
3. The imaginary parts of the parameters are closer to the true ones for the CLS, while the real ones are closer to the true ones in OLS.

But overall, both methods produce biased, potentially not too efficient, but reasonable estimates of parameters. This was not the case for the Scenario 1.

We skip the likelihood estimation, because for our DGP, it would produce estimates similar to the ones from OLS due to the zero correlation between the real and imaginary parts of the error term.

Finally, we should note that the conventional real-valued regression would not be estimable in this situation due to the high correlation between some parts of the variables. This demonstrates the difference between the cLR and the LR.

Scenario 3

As a reminder, in the Scenario 3 the real part of $\underline{x}_1$ is strongly correlated with the imaginary one of $\underline{x}_2$. We skip the correlation coefficients because arguably they will provide the information similar to the one from Scenario 2, showing the strong correlation between $\mathcal{R}(\underline{x}_1)$ and $\mathcal{I}(\underline{x}_2)$ as intended.

The OLS estimate of the model gives:

```
clm(y~., data3, loss="OLS") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: OLS
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r   51.9756    36.0017   -19.4778    123.4290
## (Intercept)_i -174.5119    45.5491  -264.9143    -84.1095 *
## x1_r             2.6536     0.3928     1.8740      3.4331 *
## x1_i             1.0941     0.4969     0.1078      2.0803 *
## x2_r             1.9229     0.3179     1.2921      2.5538 *
## x2_i            -0.8264     0.4022    -1.6246     -0.0282 *
##
## Error covariance matrix:
```

```
##          e_r      e_i
## e_r 686.3546  114.530
## e_i 114.5300 1098.659
##
## Sample size: 100
## Number of estimated parameters: 3
## Number of degrees of freedom: 97
```

While the CLS one is:

```
clm(y~., data3, loss="CLS") |>
    summary()
```

```
## Complex Linear Regression estimated via clm()
## Response variable: y
## Loss function used in estimation: CLS
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r   49.5640    41.2109   -32.2282    131.3562
## (Intercept)_i -168.4209    51.4915  -270.6172    -66.2247 *
## x1_r             2.6548     0.5073     1.6479      3.6616 *
## x1_i             1.0889     0.4179     0.2595      1.9184 *
## x2_r             1.8815     0.3543     1.1783      2.5847 *
## x2_i            -0.8449     0.4079    -1.6544     -0.0354 *
##
## Error covariance matrix:
##          e_r       e_i
## e_r 686.6725  114.6494
## e_i 114.6494 1098.7677
##
## Sample size: 100
## Number of estimated parameters: 3
## Number of degrees of freedom: 97
```

And again, we see similar picture to the Scenario 2: the real parts of parameters have lower standard errors in the model estimate using OLS than in the model estimated with CLS, while the imaginary parts have them the other way around. And similarly to the Scenario 2, we would expect the conventional real-valued regression to fail due to high correlation between some parts of variables.

While these three scenarios were supposed to show how different types of correlations impact the estimates of parameters, they are here for demonstration purposes only. They show that multicollinearity in the cLR is a different beast than the one in the conventional real-valued model. In real life, we would not have variables that have close to functional relation (with correlation close to one), instead they might have strong relation and depending on the specific situation, different estimators will give different result.

In general, getting rid of the multicollinearity is not a trivial task on its own. It can be done either by dropping some of the most correlated variables, or by applying some techniques for dimensionality reduction (such as Principle Components Analysis). These methods are standard, they lie outside of the scope of this monograph, and we do not discuss them here.

Chapter 6

Complex Dynamic Models

In this chapter we consider the application of the complex linear model to time series data. There are different ways how to do that, and we only discuss a few of the straight forward options of how to construct dynamic models. We start with a model with a trend, which has its own features, then discuss complex Autoregression (cAR), then move to the complex Moving Average (cMA) and complex Autoregression with Moving Average (cARMA) and then finish the discussion with the model applied to differences of the data, cARIMA. cARIMA and its sub-models have a lot in common with the vector ARIMA models applied to two time series. The main difference between them is in the matrix of the parameters: in case of the vector models, it is estimated in full (four parameters for two time series), while in case of the complex ARIMA, it contains two times fewer parameters to estimate. Many of the properties of the vector models can be transferred to cARIMA, and many of them have been explained in detail by **?**. Still, there are some of them that are specific for the complex models.

In this chapter, we will use index t instead of j to denote that we are applying the model to time series, not to the cross-sectional data.

6.1 Complex model with trend

One of the simplest ways to introduce dynamics in a complex model is to include a trend. In the real-valued domain this means adding an explanatory variable with the index of the observation:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_{k-1} x_{k-1,t} + \gamma t + \epsilon_t, \tag{6.1}$$

where γ is the parameter for the trend and the rest of the equation is just a conventional real-valued regression. The trend in this situation is deterministic and linear, meaning that it starts from the beginning of the sample and increases

or decreases (depending on the value of the parameter γ) indefinitely. Adding such trend to the model allows capturing the deterministic trend and neglecting its effect on the values of other parameters in the model. In cLR, the addition of the trend in the additive model is similar and has the same meaning - it will increase linearly indefinitely. Mathematically, cLR with trend can be written as:

$$\underline{y}_t = \underline{\beta}_0 + \underline{\beta}_1 \underline{x}_{1,t} + \cdots + \underline{\beta}_{k-1} \underline{x}_{k-1,t} + \underline{\gamma} t + \underline{\epsilon}_t. \tag{6.2}$$

The more interesting case arises when a multiplicative cLR is considered, i.e. the model with logarithmic transform (as discussed in Subsection 1.1.1 and Section 5.1.3):

$$\ln \underline{y}_t = \underline{\beta}_0 + \underline{\beta}_1 \ln \underline{x}_{1,t} + \cdots + \underline{\beta}_{k-1} \ln \underline{x}_{k-1,t} + \underline{\gamma} \ln t + \underline{\epsilon}_t, \tag{6.3}$$

which is equivalent to:

$$\underline{y}_t = \exp \underline{\beta}_0 \times \underline{x}_{1,t}^{\underline{\beta}_1} \times \cdots \times \underline{x}_{k-1,t}^{\underline{\beta}_{k-1}} \times t^{\underline{\gamma}} \times \exp \underline{\epsilon}_t. \tag{6.4}$$

The trend in (6.4) is now non-linear and depending on the specific values of the complex parameter $\underline{\gamma}$ can exhibit exponential, trigonometric or inverse trajectories. Figure 6.1 shows several trajectories of the real and the imaginary parts of the response variable with different values of the parameter $\underline{\gamma}$.

Figure 6.1: Different trajectories of trend with different values of the parameter $\underline{\gamma}$: 0.5-1.5i, 1.5-0.5i, and 0.5+0.2i.

As we can see, the non-linear transformation gives the cLR flexibility in capturing various types of trends in the data, making it a much more flexible instrument

for modelling than the real-valued regressions. Furthermore, this flexibility in the trend can be used in the real-valued models as well, if we substitute all the complex variables and parameters (except for $\underline{\gamma}$) by their real-valued counterparts. The model will then produce a complex response variable, the imaginary part of which can be dropped.

To better understand how the cLR model with the trend works, we can represent it in the exponential form:

$$t^{\underline{\gamma}} = \exp(\underline{\gamma} \ln t) = t^{\gamma_r} \exp(i\gamma_i \ln t), \tag{6.5}$$

where $\underline{\gamma} = \gamma_r + i\gamma_i$. The form (6.5) shows that the real part of the parameter $\underline{\gamma}$ controls the magnitude of the resulting complex variable, while the imaginary one controls how fast the argument (angle) changes. So, the more stable trajectories (closer to the linear ones) will be obtained with γ_i close to zero, while the speed of the change in the trajectory is regulated by γ_r.

Another way to look at the complex-valued trend is to use the trigonometric form of complex variable:

$$t^{\underline{\gamma}} = t^{\gamma_r} \left(\cos(\gamma_i \ln t) + i \sin(\gamma_i \ln t)\right). \tag{6.6}$$

This form tells us that fundamentally any trajectory from the complex-valued trend is trigonometric, but γ_i regulates the strength of the cosine/sine waves for the respective real and imaginary parts of the complex response variable.

6.1.1 Example in R

In R, the trend is supported in the `clm()` function from the `complex` package via the formula. Here is an example:

```
# Set random seed for reproducibility
set.seed(41)
# Sample size
obs <- 100
# Create an explanatory variable
x <- complex(real=rnorm(obs,100,10), imaginary=rnorm(obs,50,5))
# Generate parameters
b0 <- 2 + 1.5i
b1 <- 0.8 - 0.4i
b2 <- 0.5 + 0.2i
# Generate error term from the complex normal distribution
e <- rcnorm(obs, 0, sigma2=0.25, varsigma2=0.16+0.09i)
# Generate the data using non-linear model
y <- exp(b0 + b1 * log(x) + b2 * log(c(1:obs)) + e)
# Merge it to the matrix
complexData <- data.frame(y=y,x=x)
```

Figure 6.2: Generated time series with a trend.

The code above will generate data with a slowly increasing trend (similar to the one shown in Figure 6.1), which is shown in Figure 6.2

Applying a complex linear regression to the data with the same formula as in the DGP gives us the following estimates of parameters:

```
# Apply a model with the trend in logarithms
clrModel <- clm(log(y)~log(x)+log(trend), complexData, subset=1:80)
summary(clrModel)
```

```
## Complex Linear Regression estimated via clm()
## Response variable: logy
## Loss function used in estimation: likelihood
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r   3.8248     2.7186    -1.5904      9.2400
## (Intercept)_i   5.6307     1.3591     2.9235      8.3380 *
## log(x)_r        0.3237     0.5685    -0.8087      1.4561
## log(x)_i       -1.2205     0.2842    -1.7867     -0.6544 *
## log(trend)_r    0.5124     0.0604     0.3921      0.6327 *
## log(trend)_i    0.1836     0.0302     0.1235      0.2437 *
##
## Error covariance matrix:
```

```
##        e_r    e_i
## e_r 0.2416 0.0735
## e_i 0.0735 0.0604
##
## Sample size: 80
## Number of estimated parameters: 4.5
## Number of degrees of freedom: 75.5
## Information criteria:
##     AIC    AICc     BIC    BICc
## 78.6004 79.2649 89.3195 90.7753
```

The resulting model fit and the forecast for the next 20 observations are shown in Figure 6.3:

```
# Extract fitted values and transform to the original scale
fitted(clrModel) |> exp() -> yFitted
yForecast <- predict(clrModel, newdata=tail(complexData, 20))

# Setup the canvas
par(mfcol=c(2,1), mar=c(4,4,1,1))
# Plot the real part
plot(Re(complexData$y), type="l",
     xlab="Time", ylab="Real part")
lines(Re(yFitted), col="purple", lwd=2, lty=2)
exp(yForecast$mean) |> Re() |>
    lines(x=c(81:100), col="blue", lwd=2, lty=2)

# Plot the imaginary part
plot(Im(complexData$y), type="l",
     xlab="Time", ylab="Imaginary part")
lines(Im(yFitted), col="purple", lwd=2, lty=2)
exp(yForecast$mean) |> Im() |>
    lines(x=c(81:100), col="blue", lwd=2, lty=2)
```

As we can see, this example shows how a non-linear trend can be estimated and then used for forecasting using a complex linear regression model.

6.2 Complex AR

One of the possible ways of modelling dynamics is by saying that future values of the variable depend on its past ones. **?**, who have developed a theory and methodology of so called "AutoRegressive Integrated Moving Average" (or ARIMA), provide an example of a series of CO_2 output by a furnace with variable gas rate. In that situation the future amount of carbon dioxide will depend on the amount the furnace produced on the previous observations. This is one of example of an AR process. An interested reader is directed to Chapter 8 of

Figure 6.3: cLR applied to the generated data with a trend.

Svetunkov (2023), where the ARIMA model is discussed in more detail.

Mathematically the cAR model can be represented in a vector form, which will make it look more similar to vector AR rather than to the univariate one. However, the cAR model is restricted with just two time series and thus can be represented using complex variables (this is also discussed in Chandna, 2014):

$$\underline{y}_t = \underline{\beta}_0 + \underline{\phi}_1 \underline{y}_{t-1} + \cdots + \underline{\phi}_p \underline{y}_{t-p} + \underline{\epsilon}_t, \tag{6.7}$$

where ϕ_j is the j-th parameter, and p is the order of the model, complex autoregression, cAR(p). The same model can be rewritten in the conventional polynomial form using a backshift operator B:

$$\left(1 - \underline{\phi}_1 B - \cdots - \underline{\phi}_p B^p\right) \underline{y}_t = \underline{\beta}_0 + \underline{\epsilon}_t. \tag{6.8}$$

This form is convenient for what follows, e.g. for the discussion of unit roots calculation. But also, if we add the already discussed static elements to the right-hand side of (6.7) or (6.8), namely $\underline{\beta}_1 \underline{x}_{1,t} + \cdots + \underline{\beta}_{k-1} \underline{x}_{k-1,t}$, then the model will transform into cARX(p), combining the properties of the cLR and cAR.

Similarly to the conventional AR, it can be shown how complex ACF/PACF (discussed in Subsection 5.2.2) behave for different orders of cAR. In this monograph, for simplicity, we only provide an example of cAR(1) without an intercept,

noting that the logic for the higher orders is similar (Svetunkov, 2023, has similar derivations for the real-valued AR(p) in Section 8.3).

The cAR(1) model is defined as:

$$\underline{y}_t = \underline{\phi}_1 \underline{y}_{t-1} + \underline{\epsilon}_t. \tag{6.9}$$

The covariance between the most recent and the previous observations can then be calculated as (note here that we provide the formula for the conjugate covariance, but the calculations for the direct one will be similar):

$$\text{cov}\left(\underline{y}_t, \underline{y}_{t-1}\right) = \text{cov}\left(\underline{\phi}_1 \underline{y}_{t-1} + \underline{\epsilon}_t, \underline{y}_{t-1}\right), \tag{6.10}$$

which given the basic assumptions of models (Section 5.2) equals to:

$$\text{cov}\left(\underline{y}_t, \underline{y}_{t-1}\right) = \underline{\phi}_1 \text{cov}\left(\underline{y}_{t-1}, \underline{y}_{t-1}\right). \tag{6.11}$$

Now, depending on the type of covariance we need, we get:

1. For the conjugate one:

$$\text{cov}\left(\underline{y}_t, \underline{y}_{t-1}\right) = \underline{\phi}_1 \sigma^2_{y_t}; \tag{6.12}$$

2. For the direct one:

$$cov\left(\underline{y}_t, \underline{y}_{t-1}\right) = \underline{\phi}_1 \varsigma^2_{y_t}. \tag{6.13}$$

If we now insert (6.12) and (6.13) into the respective formulae for the conjugate and direct correlations (Section 3.2, equations (3.1) and (3.2)), we will get the values of the cACF for the first lag according to:

1. Conjugate correlation:

$$\rho(1) = \frac{\sqrt{\underline{\phi}_1 \sigma^2_{y_t} \underline{\phi}_1 \sigma^2_{y_t}}}{\sigma^2_{y_t}} = |\underline{\phi}_1|; \tag{6.14}$$

2. Direct correlation:

$$\varrho(1) = \frac{\underline{\phi}_1 \varsigma^2_{y_t}}{\varsigma^2_{y_t}} = \underline{\phi}_1. \tag{6.15}$$

As we can see from the formula (6.14), the specific value of the cAR(1) parameter is lost because of the geometric mean in the formula: we end up with a magnitude of the complex variable instead of it real and imaginary parts. On the other hand, the direct cACF (6.15) keeps the information about the cAR(1) parameter. However, the two formulae above are calculated in terms of expectations and their sample estimates might be slightly different. Specifically, the direct correlation (6.15) relies on the sample estimate of the direct variance of the response variable y_t, which might become equal to zero if the variances of the real and the imaginary

parts of the complex variable are similar (as discussed in Section 3.4). This means that in some situations the values of the direct cACF might become greater than one, leading to potential losses in information. But this also means that both conjugate and direct cACFs should be used for the analysis of complex time series: each of them has issues, but both of them give much clearer information about the potential process (Zhou and Pan, 2015, provide some arguments to support this claim).

If we calculate the covariance for the cAR(1) between the values y_t and y_{t-2}, we will get the following (the reader is encouraged to do the calculations manually to check the correctness of the final result, but also see Knight and Nunes, 2019):

1. Conjugate correlation:
$$\rho(2) = |\underline{\phi}_1^2|; \tag{6.16}$$

2. Direct correlation:
$$\varrho(2) = \underline{\phi}_1^2. \tag{6.17}$$

In principles, it can be shown that the main properties of real-valued AR discussed by **?** hold widely for the cAR processes: cACF will decline exponentially starting from the lag p. Note however that the direct cACF involves the power of a complex number, which means that in some situations, for some values of the complex parameter $\underline{\phi}_1$, the cACF will decline harmonically rather than exponentially.

The more useful (at least potentially) property for cAR(p) is the behaviour of the cPACF. This function shows the relations between specific lags without the effect of all the interim lags, i.e. cleaned from the autocorrelations between the neighbouring lags. One of ways of calculating the cPACF for a lag τ is to estimate the following model:

$$\underline{y}_t = \underline{\phi}_1 \underline{y}_{t-1} + \cdots + \underline{\phi}_\tau \underline{y}_{t-\tau} + \underline{\epsilon}_t \tag{6.18}$$

and getting the coefficient $\underline{\phi}_\tau$. Repeating this procedure for all lags of interest, we get a cPACF. The model (6.18) can be estimated using OLS or CLS, which will result in respective conjugate and direct cPACF.

It can be shown that for the cAR(p) process, both conjugate and direct cPACF will drop to zero abruptly after the lag p and that both of them will produce complex numbers, reflecting the respective parts of the coefficients of the cAR(p) model. However, the direct cPACF might have issues similar to the ones the direct cACF has if the variances of the real and the imaginary parts of the response variable are close to each other.

As we see, the complex AR(p) has similar properties to the real-valued one, but it should be analysed using both conjugate and direct cACF/cPACF.

Finally, in Box-Jenkins methodology, it is important for the AR(p) processes to be stationary, otherwise they become explosive and cannot be efficiently identified at the model building stage. The stationarity condition for AR(p) in **?**

is that the roots of the polynomial formed from the parameters of the AR model should all lie outside the unit circle. For AR(1) this simplifies to $|\phi_1| < 1$. The same conditions hold for the cAR(p) processes, meaning that the magnitudes of the complex parameters are considered. For the cAR(1), the stationarity condition is $|\underline{\phi}_1| < 1$. But more widely, for cAR(p), it is that the roots of the following polynomial:

$$1 - \underline{\phi}_1 x - \underline{\phi}_2 x^2 - \cdots - \underline{\phi}_p x^p = 0 \tag{6.19}$$

should all lie outside the unit circle (be greater than one by absolute value).

6.2.1 Example in R

For demonstration purposes we consider the cAR(1) process to see how cACF/cPACF and the data will look

```
# Sample size
set.seed(41)
# Number of observations
obs <- 110
# Parameters
b0 <- 100+50i
phi1 <- 0.5 + 0.2i
# Complex white noise
e <- rcnorm(obs, 0, sigma2=25, varsigma2=16+9i)
y <- vector("complex", obs)
# Initial value
y[1] <- b0 + e[1]
# The cAR(1)
for(i in 2:obs){
    y[i] <- phi1 * y[i-1] + b0 + e[i]
}
# Drop the first 10 observations as a burn-in period
y <- y[-c(1:10)]
```

The conjugate cACF/cPACF are shown in Figures 6.4 and 6.5:

```
cacf(y, method="conjugate", main="")
```

As we see from Figure 6.4, the cACF declines after the lag 1. Notably, the cACF for lag one equals to 0.546, which is close to the true value of the magnitude of the parameter $\phi_1 = 0.5 + 0.2i$ is 0.5385165.

```
cpacf(y, method="conjugate", main="")
```

The cPACF in Figure 6.5 demonstrates that both the real and the imaginary values drop to zero after the first observation, as expected for the AR(1) process. Notably, the value of the first cPACF is 0.495+0.236i, which is close to the true value of the parameter.

Figure 6.4: Conjugate cACF of the complex AR(1).

Figure 6.5: Conjugate cPACF of the complex AR(1).

The direct cACF and cPACF are shown in Figures 6.6 and 6.7, roughly giving the same information as their conjugate counterparts. We should note however that the direct cACF has the more detailed information about the dynamic relations in the data than the conjugate one.

Figure 6.6: Direct cACF of the complex AR(1).

Although, both plots contain values outside of the 95% non-rejection region, these could be considered as happening at random and can be neglected.

It is worth noting that in both Figures 6.6 and 6.7 the first values of the cACF/cPACF are close to the true value of $0.5 + 0.2i$.

Finally, the plot in Figure 6.8 shows the generated data on the complex plane and the dynamics of both the real and the imaginary parts.

When it comes to applying a model and to forecasting, the same `clm()` function from the `complex` package in R can be used:

```
# Fit the model with intercept and AR(1)
cAR1Model <- clm(y~1, orders=c(1,0,0), subset=c(1:80))
# Generate the summary
summary(cAR1Model)
```

```
## cARIMA(1,0,0) estimated via clm()
## Response variable: y
## Loss function used in estimation: likelihood
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
```

Figure 6.7: Direct cPACF of the complex AR(1).

Figure 6.8: Visualisation of the cAR(1) process.

```
## (Intercept)_r 104.0038    17.8382    68.4797    139.5278 *
## (Intercept)_i  51.6885     9.1839    33.3991     69.9778 *
## yLag1_r         0.4825     0.0856     0.3122      0.6529 *
## yLag1_i         0.2075     0.0440     0.1198      0.2952 *
##
## Error covariance matrix:
##         e_r    e_i
## e_r 17.4734 4.1781
## e_i  4.1781 4.6316
##
## Sample size: 80
## Number of estimated parameters: 3.5
## Number of degrees of freedom: 76.5
## Information criteria:
##      AIC     AICc      BIC     BICc
## 785.9517 786.3689 794.2888 795.2029
```

The output above shows the estimates of parameters of the model, demonstrating that the OLS produced estimates close to the true values of parameters (as expected). After that we can generate forecast for the next 20 observations:

```r
# Dummy data. 20 rows tells the function what the horizon is
complexData20 <- matrix(rep(1,20),20,1)
# Produce forecasts for the next 20 steps
yForecast <- predict(cAR1Model, newdata=complexData20)
```

The only catch in the code above is to create a dummy `newdata` with the number of rows equal to the forecast horizon. This is a fix, needed because the `clm()` function focuses on explanatory variables. The forecasts from the function are shown in Figure 6.9.

```r
# Produce the plot of the data and the forecasts
par(mfcol=c(2,1),mar=c(4,4,1,1))
plot(yForecast)
```

As we can see, the forecast trajectory from the cAR(1) model corresponds to the dampening line, which is a behaviour similar to the conventional real-valued AR(1) model (e.g. discussed in Subsection 8.1.1 of Svetunkov, 2023).

6.3 Complex MA

Another type of the dynamic model discussed by **?**, is called "Moving Average". In case of complex variables, it can be formulated as:

$$\underline{y}_t = \underline{\beta}_0 + \underline{\theta}_1 \underline{\epsilon}_{t-1} + \cdots + \underline{\theta}_q \underline{\epsilon}_{t-q} + \underline{\epsilon}_t, \tag{6.20}$$

Figure 6.9: Forecasts for the cAR(1) process.

where θ_j is the j-th parameter, and q is the order of the model complex MA(q) model. Similar to the cAR, it can be written via polynomials:

$$\underline{y}_t = \underline{\beta}_0 + \left(1 + \underline{\theta}_1 B + \cdots + \underline{\theta}_q B^q\right) \underline{\epsilon}_t. \tag{6.21}$$

The idea of this model is that the new actual observation is formed as a linear combination of the past white noise. **?** continue the example with a CO_2 output, showing that some random disturbances in the system in the past can impact the volume of the carbon dioxide at the current moment.

Using similar logic as in case of cAR, we will show how cACF and cPACF behave for cMA(1) without intercept, and then we will generalise their behaviour for cMA(q).

The cMA(1) model can be defined as:

$$\underline{y}_t = \underline{\theta}_1 \underline{\epsilon}_{t-1} + \underline{\epsilon}_t. \tag{6.22}$$

For the cACF, we need to calculate the covariance between the values on observations t and $t-1$. Again, the formulae for the conjugate and the direct covariances will be similar at this stage:

$$\text{cov}\left(\underline{y}_t, \underline{y}_{t-1}\right) = \text{cov}\left(\underline{\theta}_1 \underline{\epsilon}_{t-1} + \underline{\epsilon}_t, \underline{\theta}_1 \underline{\epsilon}_{t-2} + \underline{\epsilon}_{t-1}\right). \tag{6.23}$$

Given that the basic assumptions discussed in Chapter 5 hold for the "true model", the covariance is simplified to:

$$\text{cov}\left(\underline{y}_t, \underline{y}_{t-1}\right) = \underline{\theta}_1 \text{cov}\left(\underline{\epsilon}_{t-1}, \underline{\epsilon}_{t-1}\right). \tag{6.24}$$

Based on this, the conjugate and the direct covariances will be equal to respectively:

$$\text{cov}\left(\underline{y}_t, \underline{y}_{t-1}\right) = \underline{\theta}_1 \sigma^2_{y_t}. \tag{6.25}$$

and

$$cov\left(\underline{y}_t, \underline{y}_{t-1}\right) = \underline{\theta}_1 \varsigma^2_{y_t}. \tag{6.26}$$

Note that in the formulae above, we use the fact that the variance of the error term will coincide with the variance of the response variable $\underline{y}_t$, which becomes apparent from the original model formulation (6.22), assuming that the error term has a zero expectation. Inserting (6.25) and (6.26) in the formulae for the conugate and direct correlations, we get that for the cMA(1), the first lag of the cACF equals to:

1. Conjugate correlation:
$$\rho(1) = |\underline{\theta}_1|; \tag{6.27}$$

2. Direct correlation:
$$\underline{\varrho}(1) = \underline{\theta}_1. \tag{6.28}$$

Similarly to how it was with the cAR(1) process, we see that the conjugate cACF is based on the magnitude of the complex parameter, while the direct one has the value itself in it. This comes with the same in-sample limitations as above with the direct correlation being sensitive to the similarity of the variances of the real and the imaginary parts of the response variable.

However, when it comes to the second lag of the cACF, the situation differs from the one with the cAR(1):

$$\text{cov}\left(\underline{y}_t, \underline{y}_{t-2}\right) = \text{cov}\left(\underline{\theta}_1 \underline{\epsilon}_{t-1} + \underline{\epsilon}_t, \underline{\theta}_1 \underline{\epsilon}_{t-3} + \underline{\epsilon}_{t-2}\right) = 0. \tag{6.29}$$

This agrees with the behaviour of the cACF in case of the conventional real-valued MA(q) processes (**?**): it falls to zero abruptly right after the lag q.

When it comes to the cPACF of cMA(1), its behaviour becomes similar to the behaviour of cACF for AR(1). This is because the model (6.22) can be represented as an infinite cAR:

$$\underline{y}_t = \sum_{j=1}^{\infty} -1^{j-1} \underline{\theta}_1^j \underline{y}_{t-j} + \underline{\epsilon}_t, \tag{6.30}$$

which is based on the fact that $\underline{\epsilon}_t = \underline{y}_t - \underline{\theta}_1 \underline{\epsilon}_{t-1}$. In this situation, both conjugate and direct cPACF will decline either exponentially or harmonically, depending

on the specific value of the complex parameter. In general, this ones again agrees with the behaviour of the conventional PACF in case of the real-valued MA(q) process.

As we see, the behaviour of both cACF and cPACF in case of cMA(q) model agrees with their behaviour in the real-valued model. Still, similarly to the cAR(p) process, both conjugate and direct cPACF should be used for time series analysis.

Continuing the similarities between the complex and the real-valued models, it can be shown that the property of the invertibility is easily transferable to the cMA(q) model from the conventional MA(q). The model will be invertible if all the roots of the following polynomial equation lie outside of the unit circle:

$$1 + \underline{\theta}_1 x + \underline{\theta}_2 x^2 + \cdots + \underline{\theta}_p x^p = 0. \tag{6.31}$$

When it comes to estimating the model on a sample of data, we face several restrictions: the cMA part needs to be constructed recursively, because the new error from observation t is used on the next observations. This means that the analytical solutions according to OLS or CLS cannot be applied anymore, and we need to revert to the numeric optimisation. To that extent, CLS can no longer be used, because the standard solvers require the real-valued loss function, while the one from CLS is a complex one. So, when estimating cMA(q) model, we have to use either OLS or likelihood.

6.3.1 Example in R

Similarly to the cAR(p), we consider an R example, analysing how the cACF and cPACF behave in a special case of cMA(1) with an intercept:

```
# Sample size
set.seed(41)
# Number of observations
obs <- 110
# Parameters
b0 <- 100 + 50i
theta1 <- -0.8 + 0.2i
# Complex white noise
e <- rcnorm(obs, 0, sigma2=25, varsigma2=16+9i)
y <- vector("complex", obs)
# Initial value
y[1] <- b0 + e[1]
# The cAR(1)
for(i in 2:obs){
    y[i] <- theta1 * e[i-1] + b0 + e[i]
}
# Drop the first 10 observations as a burn-in period
y <- y[-c(1:10)]
```

Because we use the same random seed value as before and the same value of the parameter as in case of cAR(1), we get time series, which look similar to the one generated from cAR(1) (see Figure 6.10, compared to Figure 6.8).

Figure 6.10: Visualisation of the cMA(1) process.

However, when we analyse cACF of the time series, we will spot an important difference - its values drop to zero abruptly right after the first lag.

```
cacf(y, method="conjugate", main="")
```

As we see from Figure 6.11, the only statistically significant lag (on 5%) is the first one, all the others lie in the non-rejection region. We observe a similar behaviour in case of the direct cACF (Figure 6.12), although it is worth noting that the imaginary part of the direct cACF has all lags lying in the non-rejection region.

On the other hand, the conjugate and direct cPACF will decline exponentially, as discussed earlier. Figure 6.13 shows how the Direct cPACF looks for the complex MA(1), while Figure 6.14 shows the conjugate one. While the specific values are different for the two functions, they give roughly the same message: the cPACFs decline over time.

```
cpacf(y, method="conjugate", main="")
```

There are also some values outside of the 95% non-rejection region on the direct cPACF. However, they can be considered as happening at random (we expect

Figure 6.11: Conjugate cACF of the complex MA(1).

Figure 6.12: Direct cACF of the complex MA(1).

Figure 6.13: Direct cPACF of the complex MA(1).

Figure 6.14: Conjugate cPACF of the complex MA(1).

them to lie outside in 5% cases anyway) and can be neglected.

Finally, we can apply the model to the data:

```
# Fit the model with intercept and MA(1)
cMA1Model <- clm(y~1, orders=c(0,0,1),
                 subset=c(1:80), loss="likelihood")
# Generate the summary
summary(cMA1Model)
```

```
## cARIMA(0,0,1) estimated via clm()
## Response variable: y
## Loss function used in estimation: likelihood
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r 100.0041     0.4064    99.1947    100.8134 *
## (Intercept)_i  50.0262     0.2210    49.5862     50.4663 *
## eLag1_r        -0.2597     0.0418    -0.3429     -0.1764 *
## eLag1_i        -0.1278     0.0427    -0.2127     -0.0428 *
##
## Error covariance matrix:
##         e_r    e_i
## e_r 27.0792 3.0137
## e_i  3.0137 6.0300
##
## Sample size: 80
## Number of estimated parameters: 3.5
## Number of degrees of freedom: 76.5
## Information criteria:
##      AIC     AICc      BIC     BICc
## 856.9714 857.3886 865.3085 866.2226
```

In the code above, we specify the `loss="likelihood"` because the parameters of cMA(q) cannot be estimated analytically, and likelihood, having good statistical properties (efficiency and consistency), is appropriate. Note that the estimates of parameters for the complex moving average are far from the true ones, which is due to the sample size and the specific random sample we selected. Asymptotically, the estimator should give parameters closer to the "true" ones.

Similarly to how we did it with cAR(1), we can produce forecasts from the model using the dummy data:

```
# Dummy data. 20 rows tells the function what the horizon is
complexData20 <- matrix(1,20,1)
# Produce forecasts for the next 20 steps
yForecast <- predict(cMA1Model, newdata=complexData20)
```

These forecasts are shown in Figure 6.15.

```
# Produce the plot of the data and the forecasts
par(mfcol=c(2,1),mar=c(4,4,1,1))
plot(yForecast)
```

Figure 6.15: Forecasts for the cMA(1) process.

As expected, the forecasts from the cMA(1) converge to the intercept value abruptly right after the $h = 1$.

6.4 Complex ARMA and ARIMA

Uniting the cAR(p) with cMA(q) gives us the CARMA(p,q) model, which can be written as (in the polynomial form):

$$\left(1 - \underline{\phi}_1 B - \cdots - \underline{\phi}_p B^p\right) \underline{y}_t = \underline{\beta}_0 + \left(1 + \underline{\theta}_1 B + \cdots + \underline{\theta}_q B^q\right) \underline{\epsilon}_t. \tag{6.32}$$

This model combines the properties of the both models, making it even more flexible. However, the cACF/cPACF become even more complicated than before for some cARMA models, because they show complex interactions between the cAR and cMA. We do not discuss specific examples here and leave them as a home task for an interested reader.

Furthermore, in order for cARMA model to be identifiable, we need to apply it to the stationary data - all the examples of cAR and cMA above assumed

that implicitly. There are several ways of achieving this, including the addition of trend component as in Section 6.1 and taking differences of the original data. The latter implies the introduction of the difference polynomial giving us cARIMA(p,d,q), which is analogous to the real-valued ARIMA:

$$\left(1 - \underline{\phi}_1 B - \cdots - \underline{\phi}_p B^p\right)(1-B)^d \underline{y}_t = \underline{\beta}_0 + \left(1 + \underline{\theta}_1 B + \cdots + \underline{\theta}_q B^q\right)\underline{\epsilon}_t. \quad (6.33)$$

There are several ways how to construct such a model. First one is to take the differences of the original data and then apply cARMA. But this implies that we estimate the following model:

$$\Delta_{\underline{y}_t} = (1-B)^d \underline{y}_t = \underline{\beta}_0 + \left(\underline{\phi}_1 B + \cdots + \underline{\phi}_p B^p\right)\Delta_{\underline{y}_t} + \left(\underline{\theta}_1 B + \cdots + \underline{\theta}_q B^q\right)\underline{\epsilon}_t + \underline{\epsilon}_t. \quad (6.34)$$

But working with this model might be challenging because in order to produce forecasts from it, one needs to take the inverse of differences to get back to the original units. Instead, we argue that the following form (in principle) is easier to implement and use:

$$\underline{y}_t = \underline{\beta}_0' + \frac{\left(1 + \underline{\theta}_1 B + \cdots + \underline{\theta}_q B^q\right)}{\left(1 - \underline{\phi}_1 B - \cdots - \underline{\phi}_p B^p\right)(1-B)^d}\underline{\epsilon}_t, \quad (6.35)$$

where we now can produce future $\hat{\underline{y}}_t$ directly instead of $\hat{\Delta}_{\underline{y}_t}$. This formulation is implemented and used in the `clm()` function from the `complex` package in R via the `orders` parameter.

When it comes to deciding whether to take the differences of the time series or not, the standard ADF (Dickey and Fuller, 1979) and KPSS (**?**) tests can be applied independently to the real and imaginary parts to see whether they are stationary or not. This approach assumes that both parts contain variables that have similar dynamics in principle. But there might be some situations, when one part is stationary, while the other is not. In that case, the decision about the order of differencing d might not be straight forward. A solution in this case is to take differences, although this might lead to the overdifferencing one of the parts of the complex variable. The alternative approach is to apply the test to the two parts jointly and a relatively simple solution in this case is to use the MDS to scale the complex variable into the real-valued one. After that the conventional ADF/KPSS can be applied to determine whether the differences are needed to make the variable stationary "overall".

When it comes to the selection of orders of cARIMA(p,d,q), using cACF/cPACF for this purpose becomes extremely difficult, because of all the possibilities for the behaviour of these functions depending on the order of the model and/or specific values of parameters. For example, in some situations cACF will decline harmonically for an AR(p) process, while cPACF will do the same for an MA(q) one. The combination of cAR and cMA might produce really puzzling cACF/cPACF, which might become useless for diagnostics or order selection.

Furthermore, as discuss in Section 8.3 of Svetunkov (2023), the fact that some specific orders of cARMA(p,q) generate some specific shapes of cACF/cPACF does not mean that if we observe that shape then the data is generated from that specific model. After all, there can be many reasons why we observe the specific cACF/cPACF, and conclusions about the order of the appropriate cARMA model just based on them become unreliable. As a result, we recommend using other principles for the selection of the appropriate p and q. One of those is to do that based on the information criteria, such as AIC (**?**) or BIC (**?**). But in order to be able to do that, one needs to use likelihood for the model estimation.

Chapter 7

Examples of application

In this chapter we consider examples of application for cLR and the complex ARIMA models.

7.1 cARIMA applied to Box-Jenkins Sales data

In this example, we consider the Box-Jenkins Sales data with a leading indicator. While typically the indicator is treated as an explanatory variable, we will treat the two as one joint complex variable. Separately, they have the dynamics shown in Figure 7.1.

As we can see they seem to change over time similarly, which is probably because the indicator drives the sales. But if we unite the two in one complex variable and plot their joint dynamics we will see that not only the values change over time, but also the relation between the variables (see Figure 7.2).

While judgmentally we can conclude that both parts of this complex variable are non-stationary, we will conduct ADF and KPSS tests using `adf.test()` and `kpss.test()` functions from the `tseries` package after applying MDS to it. We will use 1% significance level in the hypotheses testing.

```
# Normalise the variables
complex(real=BJsales, imaginary=BJsales.lead) |>
    cscale(scaling="norm") -> y
# Scale the complex variable into the real-valued one
complex2vec(y) |> dist() |> cmdscale(k=1) -> yScaled
```

In the code above we use the `cscale()` function from the `complex` package to scale both parts of the complex variable. We can then conduct the ADF test:

```
# Apply ADF test
tseries::adf.test(yScaled)
```

Figure 7.1: Box-Jenkins Sales data with a leading indicator.

Figure 7.2: Box-Jenkins Sales data. Complex dynamics. The red circle depicts the first observation, while the blue trianle is the last one.

```
##
##  Augmented Dickey-Fuller Test
##
## data:  yScaled
## Dickey-Fuller = -2.1558, Lag order = 5, p-value = 0.5115
## alternative hypothesis: stationary
```

The output above shows that on 1% level we fail to reject the null hypothesis that the data is not stationary according to the ADF test. With KPSS, the final message is similar, because we reject the null hypothesis on the 1% significance level:

```
tseries::kpss.test(yScaled)
```

```
## Warning in tseries::kpss.test(yScaled): p-value smaller than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  yScaled
## KPSS Level = 2.6339, Truncation lag parameter = 4, p-value = 0.01
```

If we take the first differences, the conclusions become contradictory:

```
diff(yScaled) |>
    tseries::adf.test()
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(yScaled)
## Dickey-Fuller = -2.5661, Lag order = 5, p-value = 0.3406
## alternative hypothesis: stationary
```

In the ADF test, we still fail to reject the null hypothesis on the 1% level (see output above), while in the KPSS test, we now fail to reject the null as well. This means that the ADF detects non-stationarity in the data, while the KPSS does not.

```
diff(yScaled) |>
    tseries::kpss.test()
```

```
## Warning in tseries::kpss.test(diff(yScaled)): p-value greater than printed
## p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  diff(yScaled)
## KPSS Level = 0.13191, Truncation lag parameter = 4, p-value = 0.1
```

Given this contradiction, we plot the differences of the scaled data to make the conclusion judgmentally (Figure 7.3):

Figure 7.3: First differences of the MDS of the complex variable of the Box-Jenkins Sales data.

Analysing the series in Figure 7.3, it might be the case that the data contains a strong AR component or that it is non-stationary. Using this information, we will construct two models: cARIMA(p,2,q) and cARIMA(p,1,q) - and see, which of them performs better. Note that while in general the models are not comparable if different orders d are applied due to the loss of the number of in-sample observations, the `clm()` function takes care of potential missing values and extrapolates them back, making the models comparable.

Now we apply the cARIMA model to the newly constructed complex variable. We will try several special cases of cARIMA and choose the one that has the lowest information criterion. The code below shows how the process can be automated:

```
# Create all combinations of cARIMA orders to consider
# Here, we set d={1, 2} based on the earlier data exploration
expand.grid(c(0:3),c(1:2),c(0:3)) |>
    as.matrix() -> orders
colnames(orders) <- c("cAR","cI","cMA")
nModels <- nrow(orders)
# Prepare the list of all models under consideration
cARIMABJ <- vector("list",nModels)
names(cARIMABJ) <- paste0("cARIMA(",orders[,1],",",
                          orders[,2],",",orders[,3],")");
```

```r
# Construct models
for(i in 1:nModels){
    cARIMABJ[[i]] <- clm(y~1, orders=orders[i,], subset=c(1:130))
}
```

The script above will produce 32 cARIMA models of different orders. To choose the best one based on an information criterion, we will use the `AICc()` function from the `greybox` package in R:

```r
# Extract AICc values
cARIMABJAICc <- sapply(cARIMABJ, AICc)
# Fix names. Sometimes R makes silly things...
names(cARIMABJAICc) <- names(cARIMABJ)
# Record the index of the model with the lowest AICc
i <- which.min(cARIMABJAICc)
```

The summary of the best performing model is shown below:

```r
summary(cARIMABJ[[i]])
```

```
## cARIMA(3,1,1) estimated via clm()
## Response variable: y
## Loss function used in estimation: likelihood
## Coefficients:
##                Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r   0.0014     0.0026    -0.0038      0.0066
## (Intercept)_i   0.0047     0.0052    -0.0055      0.0149
## yLag1_r         0.3009     0.1564    -0.0087      0.6104
## yLag1_i        -0.1360     0.1925    -0.5171      0.2451
## yLag2_r         0.0989     0.2409    -0.3779      0.5757
## yLag2_i         0.0057     0.2198    -0.4293      0.4407
## yLag3_r         0.0823     0.0333     0.0163      0.1483 *
## yLag3_i        -0.2276     0.0688    -0.3638     -0.0913 *
## eLag1_r        -0.3726     0.0831    -0.5371     -0.2080 *
## eLag1_i         0.0946     0.0489    -0.0022      0.1914
##
## Error covariance matrix:
##        e_r    e_i
## e_r 0.0005 0.0006
## e_i 0.0006 0.0070
##
## Sample size: 130
## Number of estimated parameters: 6.5
## Number of degrees of freedom: 123.5
## Information criteria:
##       AIC      AICc       BIC      BICc
## -901.0591 -900.2632 -882.4202 -880.4831
```

The output above shows the name of the model and the standard statistics we have already seen before. The forecast from this model for the next 20 observations is shown in Figure 7.4.

Figure 7.4: Forecast for the Box-Jenkins series.

As we see, the model managed to capture the dynamics of the original data well, producing reasonable point forecasts for both parts. If we want to return to the original scale, we can use the `cdescale()` function from the `complex` package:

```
yForecast <- predict(cARIMABJ[[i]],
                     newdata=matrix(NA,20,1))
cdescale(yForecast$mean,
         complex(real=BJsales, imaginary=BJsales.lead),
         scaling="norm")
```

```
##  [1] 257.6162+13.1087i 258.7371+13.1270i 257.4338+13.1111i 257.5523+13.1418i
##  [5] 257.7756+13.1524i 257.8027+13.1935i 258.1313+13.2268i 258.4488+13.2549i
##  [9] 258.8752+13.2862i 259.3374+13.3120i 259.7896+13.3359i 260.2593+13.3575i
## [13] 260.7118+13.3771i 261.1492+13.3960i 261.5722+13.4142i 261.9793+13.4323i
## [17] 262.3764+13.4506i 262.7655+13.4691i 263.1503+13.4881i 263.5336+13.5074i
```

7.2 Seatbels law in Great Britain

In another example, we consider the time series data with several explanatory variables. This is the Road Casualties in Great Britain from 1969 to 1984, available in the `datasets` package in R. We do not need all the variables in the data, but we will use the following ones:

1. drivers - the number of car drivers killed or seriously injured;
2. killed - car drivers killed (from the `DriversKilled` variable);
3. injured - drivers injured (calculated as `drivers` - `DriversKilled`);
4. kms - distance driven by drivers;
5. PetrolPrice - average price of petrol in each month;
6. law - binary variable, indicating when the law for mandatory seatbelts was in effect.

The variables (2) and (3) will form our complex response variable, while (4), (5) and (6) will be used as explanatory ones. Note that in this setting we do not form complex variables from the explanatory ones and only have the complex response variable. The code below shows how this can be done:

```
SeatbeltsData <- data.frame(killed=Seatbelts[,"DriversKilled"],
                            injured=Seatbelts[,"drivers"]-
                                Seatbelts[,"DriversKilled"],
                            kms=Seatbelts[,"kms"],
                            PetrolPrice=Seatbelts[,"PetrolPrice"],
                            law=Seatbelts[,"law"])

SeatbeltsComplex <-
    data.frame(y=complex(real=SeatbeltsData$killed,
                         imaginary=SeatbeltsData$injured),
               kms=SeatbeltsData$kms,
               PetrolPrice=SeatbeltsData$PetrolPrice,
               law=SeatbeltsData$law)
```

Looking at the response variable, we would argue that a multiplicative model is needed, because the amplitude of seasonality in both variables changes when the seatbelts law is introduced (see Figure 7.5).

Taking logarithm of the complex response variable would not be reasonable in this situation, because that would not fix the issue with the amplitude of seasonality. So, we can use the `clog()` function from the `complex` package to take logarithms of real and imaginary parts independently. Furthermore, given that both the real and the imaginary parts of the response variable are in the same units (number of drivers), we do not need to scale them. The only additional thing we need to do is to introduce a variable for seasonality, which can be done using `temporaldummy()` function from the `greybox` package:

Figure 7.5: Dynamics of number of drivers killed and injured in the Great Britain.

```
SeatbeltsComplex$seasonal <-
    temporaldummy(SeatbeltsComplex[,1],
                  type="month", of="year", factors=TRUE)
```

We can then estimate a complex-valued model with real-valued explanatory variables, which should be equivalent to a multivariate linear regression. Note that we take logarithms of the numerical explanatory variables to capture the non-linearity correctly:

```
SeatbeltscLR01 <-
    clm(I(clog(y))~log(kms)+log(PetrolPrice)+law+seasonal,
        SeatbeltsComplex)
```

The output above shows the estimates of the parameters of the model, demonstrating how the model captured the relations in the data. The model fit is shown in Figure 7.6.

While the model has already captured the relations in the data well, there is still some structure left, which becomes apparent from the analysis of the ACF/PACF in the residuals (Figure 7.7).

It is hard to understand what specifically is missing in the residuals, so we will fit several cARIMA models with maximum order of 1 for each p, d, and q:

Figure 7.6: Number of drivers killed and injured and the fitted values.

Figure 7.7: ACF/PACF of the residuals of the first model.

```
# Create all possible orders
cARIMAOrders <- as.matrix(expand.grid(c(0,1),c(0,1),c(0,1)))
colnames(cARIMAOrders) <- c("AR","I","MA")
# Create a list for all models
SeatbeltscLRList <- vector("list",nrow(cARIMAOrders))
names(SeatbeltscLRList) <-
    paste0("ARIMA(",apply(cARIMAOrders, 1,
                          paste0, collapse=","),")")

# Loop to estimate models.
for(i in 1:nrow(cARIMAOrders)){
    SeatbeltscLRList[[i]] <-
        clm(I(clog(y))~log(kms)+log(PetrolPrice)+law+seasonal,
            SeatbeltsComplex, orders=cARIMAOrders[i,])
}
```

After that we can see which of the models is more appropriate for the data based on the AICc:

```
# Select the best model
iBest <- which.min(sapply(SeatbeltscLRList, AICc))
# Produce summary.
# We specify stepSize for the Hessian calculation
# to make sure that we get reasonable standard errors
summary(SeatbeltscLRList[[iBest]], stepSize=1e-6)
```

```
## cARIMAX(1,0,1) estimated via clm()
## Response variable: Iclog(y)
## Loss function used in estimation: likelihood
## Coefficients:
##                     Estimate Std. Error Lower 2.5% Upper 97.5%
## (Intercept)_r         3.0425     0.1210     2.8036      3.2814 *
## (Intercept)_i         5.0013     0.1351     4.7347      5.2679 *
## log(kms)_r           -0.0693     0.0282    -0.1250     -0.0136 *
## log(kms)_i           -0.1142     0.0354    -0.1841     -0.0444 *
## log(PetrolPrice)_r   -0.2915     0.0221    -0.3351     -0.2479 *
## log(PetrolPrice)_i   -0.2393     0.1077    -0.4519     -0.0268 *
## law_r                -0.0958     0.0744    -0.2425      0.0510
## law_i                -0.0982     0.0946    -0.2849      0.0884
## seasonal2_r          -0.0325     0.0885    -0.2072      0.1422
## seasonal2_i          -0.0389     0.0682    -0.1736      0.0958
## seasonal3_r           0.0275     0.0182    -0.0083      0.0634
## seasonal3_i           0.0704     0.0793    -0.0861      0.2268
## seasonal4_r           0.0164     0.0391    -0.0607      0.0936
## seasonal4_i          -0.0215     0.0362    -0.0930      0.0500
## seasonal5_r           0.0679     0.0507    -0.0322      0.1681
```

```
## seasonal5_i          0.1105     0.0020     0.1065      0.1145 *
## seasonal6_r          0.1044     0.1126    -0.1179      0.3267
## seasonal6_i          0.0344     0.0982    -0.1595      0.2282
## seasonal7_r          0.0872     0.0512    -0.0140      0.1883
## seasonal7_i          0.1126     0.0816    -0.0485      0.2737
## seasonal8_r          0.0941     0.0580    -0.0204      0.2086
## seasonal8_i          0.1064     0.0749    -0.0415      0.2542
## seasonal9_r          0.1596     0.0469     0.0669      0.2522 *
## seasonal9_i          0.1230     0.0059     0.1113      0.1346 *
## seasonal10_r         0.2532     0.0385     0.1771      0.3292 *
## seasonal10_i         0.1775     0.0578     0.0634      0.2917 *
## seasonal11_r         0.2691     0.0590     0.1525      0.3856 *
## seasonal11_i         0.2445     0.0594     0.1272      0.3618 *
## seasonal12_r         0.2864     0.1137     0.0620      0.5107 *
## seasonal12_i         0.2543     0.0788     0.0987      0.4099 *
## yLag1_r              0.3702     0.1162     0.1409      0.5996 *
## yLag1_i              0.0168     0.0263    -0.0352      0.0688
## eLag1_r             -0.1513     0.0435    -0.2372     -0.0655 *
## eLag1_i              0.0658     0.0159     0.0344      0.0973 *
##
## Error covariance matrix:
##        e_r    e_i
## e_r 0.0157 0.0050
## e_i 0.0050 0.0057
##
## Sample size: 192
## Number of estimated parameters: 18.5
## Number of degrees of freedom: 173.5
## Information criteria:
##       AIC      AICc       BIC      BICc
## -765.7492 -761.5665 -705.4855 -694.4905
```

The estimates of parameters in the output above have a clear interpretation:

1. With the increase of distance driven by 1%, the number of killed and injured drivers tends to decrease on average by 0.069% and 0.114% respectively;
2. The 1% increase in average petrol price leads to the decrease in the number of killed and injured drivers by 0.292% and 0.239% respectively;
3. The introduction of law on mandatory seatbelts has reduced the number of killed and injured drivers by approximately 10% each.

While we could fit two independent regression models instead of one complex-valued, it would not be able to capture the cARIMA elements the way they are captured in the model above. Arguably, number of killed and injured drivers have a complex dynamic relation, which is captured by the cARIMA(1,0,1) part of the model.

Finally, we can see how the model fits the data and compare it with the previous

one (Figure 7.8).

Figure 7.8: Number of drivers killed and injured and the fitted values from the cARIMA(1,0,1) model.

Arguably, the model fit is better due to the capture of dynamic elements that were missing in the cLR before.

Conclusions

Now that we have reached the final chapter of this book, we want to discuss the main limitations and future work in the area of complex-valued economics. We focused this book on the technical side of modelling with complex-valued regressions, discussing the estimation methods, elements of correlation analysis and the main aspects of the regression analysis. We briefly discussed the complex ARIMA model, only touching on some of the important properties of it. But there is much more left outside.

For example, we discussed that complex-valued models share some properties with the multivariate ones, such as Vector AR (VAR). This means that we can use the tools developed for those models for the complex-valued ones with minor modifications. For instance, we did not discuss how to calculate Impulse Response Function, which are widely used for analysis of effects based on VAR in Economics and Marketing Analytics. We expect that these tools can be transferred to the complex-valued models easily, but they lie outside of the scope of this monograph and can be considered as a future work.

Furthermore, a curious reader might wonder what to do when we are interested in modelling more than two variables at the same time. The first instinct would be to switch to the multivariate models (VAR). But the main issue of those models is the high dimensionality. For example, VAR(p) on 10 time series requires the estimation of $p \times 10^2$ parameters, which is not a simple task. However, **?** has developed a solution in this situation which involves complex variables. The model based on this is called vector cARIMA, VcARIMA(p,d,q) and involves stacking several complex response variables in a vector. This model is currently researched, and Sergey Svetunkov is preparing a paper explaining its properties and showing how it can be used for time series analysis and in forecasting.

Moreover, as discussed in this book several times, some of the largest benefits of complex variables arise when a non-linear model is considered, e.g. a log-log model. We have discussed what happens with the cLR in this situation (Subsection 5.1.3 and Section 6.1), but we have not yet study how the cARIMA behaves in this situation. We expect that it should be able to capture non-linear dynamics, which would be beneficial when applied to large samples of data, such as, for example, in energy domain.

Bibliography

Adali, T., Li, H., Novey, M., Cardoso, J. F., 2008. Complex ica using nonlinear functions. IEEE Transactions on Signal Processing 56, 4536–4544.

Adali, T., Schreier, P. J., Scharf, L. L., 2011. Complex-valued signal processing: The proper way to deal with impropriety. IEEE Transactions on Signal Processing 59, 5101–5125.

Amblard, P. O., Gaeta, M., Lacoume, J. L., 1996. Statistics for complex variables and signals - part i: Variables. Signal Processing 53, 1–13.

Arens, R., 1957. Complex processes for envelopes of normal noise. IRE Transactions on Information Theory 3, 204–207.

Bombelli, R., 1572. L'algebra opera.

Brown, W. M., Crane, R. B., 1969. Conjugate linear filtering. IEEE Transactions on Information Theory 15, 462–465.

Buzzi, S., Lops, M., Sardellitti, S., 2006. Widely linear reception strategies for layered space-time wireless communications. IEEE Transactions on Signal Processing 54, 2252–2262.

Chandna, S., 2014. Frequency domain analysis and simulation of multi-channel complex-valued time series, 207.

Chargé, P., Wang, Y., Saillard, J., 2001. A non-circular sources direction finding method using polynomial rooting. Signal Processing 81, 1765–1770.

Chevalier, P., Pipon, F., 2006. New insights into optimal widely linear array receivers for the demodulation of bpsk, msk, and gmsk signals corrupted by noncircular interferences - application to saic. IEEE Transactions on Signal Processing 54, 870–883.

Chung, K., 2020. Pseudo complex correlation coefficient : with application to correlated information sources for noma in 5g systems. International Journal of Advanced Smart Convergence 9, 42–51.

Delmas, J. P., 2004. Asymptotically minimum variance second-order estimation for noncircular signals with application to doa estimation. IEEE Transactions on Signal Processing 52, 1235–1241.

Dickey, D. A., Fuller, W. A., 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association 74 (366a), 427 – 431.

Eriksson, J., Koivunen, V., 2006. Complex random vectors and ica models: Identifiability, uniqueness, and separability. IEEE Transactions on Information Theory 52, 1017–1029.

Gao, R., Ahmadzade, H., Esfahani, M., 2019. Covariance and pseudo-covariance of complex uncertain variables. Journal of Intelligent and Fuzzy Systems 36, 241–251.

Gardner, W. A., 1993. Cyclic wiener filtering: Theory and method. IEEE Transactions on Communications 41, 151–163.

Gerstacker, W. H., Schober, R., Lampe, A., 2003. Receivers with widely linear processing for frequency-selective channels. IEEE Transactions on Communications 51, 1512–1523.

Goh, S. L., Popović, D. H., Mandic, D. P., 2004. Complex-valued estimation of wind profile and wind power. Proceedings of the Mediterranean Electrotechnical Conference - MELECON 3, 1037–1040.

Goodman, N. R., 1963. Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). The Annals of Mathematical Statistics 34, 152–177.

Halliwell, L. J., 2015. Complex random variables. Casualty Actuarial Society E-Forum Fall, 1–66.

Javidi, S., Mandic, D. P., Took, C. C., Cichocki, A., 2011. Kurtosis-based blind source extraction of complex non-circular signals with application in eeg artifact removal in real-time. Frontiers in Neuroscience, 1–25.

Javidi, S., Pedzisz, M., Goh, S. L., Mandic, D. P., 2008. The augmented complex least mean square algorithm with application to adaptive prediction problems. Proceedings of the 1st IARP Workshop on Cognitive Information Processing, 54–57.

Jia, Y., Yang, X., 2016. Adaptive complex-valued independent component analysis based on second-order statistics. Journal of Electrical and Computer Engineering 2016, 20–22.

Knight, M. I., Nunes, M. A., 2019. Long memory estimation for complex-valued time series. Statistics and Computing 29, 517–536.

Kociuba, M. C., Rowe, D. B., 2016. Complex-valued time-series correlation increases sensitivity in fmri analysis. Magnetic Resonance Imaging 34, 765–770.

Kourentzes, N., Svetunkov, I., Trapero, J. R., 2021. Connecting forecasting and inventory performance: a complex task. SSRN Electronic Journal.

Mandic, D. P., Javidi, S., Souretis, G., Goh, V. S. L., 2007. Why a complex valued solution for a real domain problem. IEEE, 384–389.

Miyabe, S., Ono, N., Makino, S., 2015. Estimating Correlation Coefficient Between Two Complex Signals Without Phase Observation. Vol. 9237. pp. 421–428.

Neeser, F. D., Massey, J. L., 1993. Proper complex random processes with applications to information theory. IEEE Transactions on Information Theory 39, 1293–1302.

Panchev, S., 1971. Second- order statistics of complex signals. Pergamon Press.

Percival, D. B., 2006. Exact simulation of complex-valued gaussian stationary processes via circulant embedding. Signal Processing 86, 1470–1476.

Picinbono, B., Bondon, P., 1997. Second-order statistics of complex signals. IEEE Transactions on Signal Processing 45, 411–420.

Reed, I. S., 1962. On a moment theorem for complex gaussian processes. IRE Transactions on Information Theory 8, 194–195.

Rubin-Delanchy, P., Walden, A. T., 2007. Simulation of improper complex-valued sequences. IEEE Transactions on Signal Processing 55, 5517–5521.

Schreier, P. J., Scharf, L. L., 2010. Statistical Signal Processing of Complex-Valued Data. Cambridge University Press.

Svetunkov, I., 2022. Statistics for business analytics. Lecture notes. OpenForecast, (version: 23.04.2023).
URL https://openforecast.org/sba/

Svetunkov, I., 2023. Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM). Chapman and Hall/CRC.
URL https://openforecast.org/adam/

Svetunkov, I., 2024. complex: Time Series Analysis and Forecasting Using Complex Variables. R package version 0.0.2.41021.
URL https://github.com/config-i1/complex

Svetunkov, I., Kourentzes, N., Ord, J. K., 12 2022. Complex exponential smoothing. Naval Research Logistics (NRL) 69, 1108–1123, we introduce CES.

Svetunkov, S., 2012. Complex-Valued Modeling in Economics and Finance. Springer New York, New York, NY.

Tamari, B., 1997. Conservation and Symmetry Laws and Stabilization Programs in Economics. Econometry Ltd.

Tauböck, G., 2007. Complex noise analysis of dmt. IEEE Transactions on Signal Processing 55, 5739–5754.

Tavares, G. N., Tavares, L. M., 2007. On the statistics of the sum of squared complex gaussian random variables. IEEE Transactions on Communications 55, 1857–1862.

Trampitsch, S., 2013. Complex-valued data estimation, 129.

van den Bos, A., 1994. Estimation of complex parameters. Elsevier, pp. 1429–1433.

van den Bos, A., 1995. The multivariate complex normal distribution - a generalization. IEEE Transactions on Information Theory 41, 537–539.

Walden, A. T., Rubin-Delanchy, P., 2009. On testing for impropriety of complex-valued gaussian vectors. IEEE Transactions on Signal Processing 57, 825–834.

Wooding, R. A., 1956. The multivariate distribution of complex normal variables. Biometrika 43, 212–215.

Yao, Z., Cui, X., Lu, M., Feng, Z., Yang, J., 2010. Pseudo-correlation-function-based unambiguous tracking technique for sine-boc signals. IEEE Transactions on Aerospace and Electronic Systems 46, 1782–1796.

Yao, Z., Lu, M., 2011. Side-peaks cancellation analytic design framework with applications in boc signals unambiguous processing. pp. 775–785.

Zhang, Y., Ma, L., Zhang, L., Xu, Z., 2022. Acquisition algorithm of beidou b1c signal based on improved pseudo correlation function. IEEE, pp. 23–28.

Zhou, Y., Pan, Y., 2015. Unambiguous tracking method for alternative binary offset carrier modulated signals based on pseudo correlation function technique. IEEE Communications Letters 19, 371–374.