COMPLEX NUMBERED LINEAR REGRESSION

James Tian

Main Supervisor: Dr Pierre Lafaye de Micheaux
Joint Supervisor: Frédéric Ouimet

School of Mathematics and Statistics
UNSW Sydney

November 2019

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: _____   Date: _____

Abstract

Complex numbered data crop up from many phenomena, so there will always be a need for statistical tools like linear regression to research them. Although it may be tempting to splice the real and imaginary or take the modulus, simplifications in expressions and higher accuracy can be found when sticking to the complex domain just like in other fields. In this work we look at how estimators can sometimes parallel linear regression in the real domain and give working estimators in R with nice statistical properties backed up by simulations. Applying these estimators to a wind case study which has both complex observations and data we look at the interpretation and different insights that complex estimates can have.

Contents

CHAPTER 1

Introduction

1.1 Motivation

Complex numbers are found in many fields of research, whether naturally measured or constructed organically to describe some phenomenon, nevertheless, we seek some statistical analysis to better understand what is happening. Our goal is to create a linear regression tool that can be used to analyse complex data, we'll be attempting to improve the effectiveness of current tools through using more of the data available and basing our estimators on a more generalised complex normal distribution. Historical methods using only the magnitude of complex data [1] have been improved on by including phase information [2], but further precision can be obtained using the second order information from the improper complex normal distribution [3].

We will be providing some closed forms of estimators and giving value extrinsically in unique interpretation of complex regression co-efficients, showing their parallel relationship with estimators in the real domain. Extending problems into complex numbers can be seen to be useful to unexpected areas in maths, with it's relationships with trigonometry and polar co-ordinates or otherwise, often they can provide great simplifications and cleanliness in statements, formulations and proofs [4]. These include and are not limited to; residue theorem in complex analysis, characteristic functions in statistics and solving low degree cases of Fermat's Last Theorem (unfortunately we are not going to be doing anything this flashy).

Accuracy of complex numbered statistical tools is necessary when looking at real world problems so improvements in these tools could amount to better detection for faults in nuclear reactor pipes when performing Eddy Current Testing [5], which is a process that returns naturally occurring complex numbers due to the nature of alternating current, increasing the safety of reactors. Or can come in the form of higher resolution imaging for fMRI scans, as regression tools check for significance in changes in brain activation. For our case study improvements in accuracy can lead to better decision making when dealing with wind, from wind turbine automatic orientation to professional sailing.

1.2 Outline

This paper will begin with a refresher to linear regression in the real case, looking at estimators and some of their properties in the first chapter, so you are very welcome to skip the first chapter if you are familiar with linear regression. Then in the next we focus on the complex linear regression model, following a similar structure to the first chapter, but with more information about the formulation of the complex normal distribution and how it ties in with the bi-variate normal. This chapter

then will provide the closed form estimates of the regression parameters and their important properties relevant for regression.

To verify our findings and in general test the properties in the fourth chapter, we run various simulations in R, looking at if we have the correct distributional properties and consistency of the estimators. To show these, we have simple graphical depictions with descriptions of what is going on, on-top of these we also compare the complex regression model with some common linear regression tools that deal with complex data.

In our application of complex regression we will be looking at modelling wind, we seek to test the effectiveness of our model on some real complex data. Data taken from NOAA (National Centers for Environmental Information, National Oceanic and Atmospheric Administration) US) [1] using wind speed as the magnitude and the direction of the wind as the argument, creating a complex data point [6] [7] and then using past winds to predict future winds. Then judging what could be added to that model, along with a potential application of prediction.

---

[1]https://www.ncdc.noaa.gov/data-access/land-based-station-data

CHAPTER 2

Real Numbered Linear Regression Model

This is a brief introduction or refresher, you can skip this chapter if you'd like. Our goal is to numerically describe the relationship between our measured data of interest y_i and p covariates x_{i1} through to x_{ip} which is something else we observe related to y_i. In creating a line of best fit through the data, it becomes the regression line that describes the linear relationship between y_i's and x_{ij}'s for $j = 1..p$. The subscript i denotes how many points of data we have, for our needs we take $i = 1, ..., n$, meaning there is a total of n data points.

Now comes the linear aspect, where we assume the relationship to be linear, at least on average

$$\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} = \mathbf{x}_i^T \boldsymbol{\beta}, \tag{2.0.1}$$

where $\beta_0, \beta_1, ..., \beta_p$ are constants and describe the intercept and the slope of the model respectively. However we must include some form of random error as naturally such perfect linear relationships would be too surprising. This error we expect to average out at 0, have some variation we'll get into later and are independent of one another, so we have our full model would be

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i. \tag{2.0.2}$$

To achieve the best linear relationship, we must find the values of $\boldsymbol{\beta}$ so our model most resembles the data $(\mathbf{x}_i, Y_i)$.

2.1   Real Model

For the real case the model is (2.0.2), the variable of interest y_i is a real number and so are $\boldsymbol{\beta}$, $\mathbf{x}_i$ and our error ϵ_i. However, we must make further assumptions about our model to develop our estimators so they best fit the data, they are as follows;

2.1.1   Assumptions

- The Y_i's come from a linear transformation of a $\mathbf{x}_i$.
- Our errors, ϵ_i are independent of one another.
- Furthermore ϵ_i are identically distributed, so that for all i, they come from the same distribution, that distribution being $\epsilon \sim \mathcal{N}(0, \sigma^2)$ (normal with mean 0, constant and finite variance σ^2).

2.2 Estimators

From this assumption that ϵ_i is distributed $\mathcal{N}(0, \sigma^2)$ and since $\mathbf{x}_i^T \boldsymbol{\beta}$ terms are all constants, the linear shift in the normal means y_i is distributed $\mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$ and are also independent of one another. So its probability density is;

$$f(y_i|\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2}{2\sigma^2}\right). \tag{2.2.1}$$

From this we can state the likelihood function as the joint density of the data, but as a function of $\boldsymbol{\beta}$ and σ^2 given the data,

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}, \sigma^2),$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2}{2\sigma^2}\right).$$

Since log is a monotonically increasing function, the maximum won't change after we take the log of both sides,

$$\log \mathcal{L}(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = \log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2}{2\sigma^2}\right)\right),$$

$$= \sum_{i=1}^{n} \left(\log\left[\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left(-\frac{\left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2}{2\sigma^2}\right)\right]\right),$$

$$\log \mathcal{L}(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = -\frac{n}{2} \log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2.$$

We then differentiate the log likelihood so we can obtain some $\boldsymbol{\beta}$ and σ^2 which maximise the likelihood,

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = \frac{\partial}{\partial \boldsymbol{\beta}} \left(-\frac{n}{2} \log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2\right),$$

$$= -\frac{1}{2\sigma^2} \frac{\partial}{\partial \boldsymbol{\beta}} \left(\sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2\right),$$

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{x}_i \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right). \tag{2.2.2}$$

Setting this to zero, we can obtain a point value for $\boldsymbol{\beta}$,

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{x}_i \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right) = 0,$$

$$0 = \sum_{i=1}^{n} \sigma^{-2} \mathbf{x}_i y_i - \sum_{i=1}^{n} \sigma^{-2} \mathbf{x}_i \mathbf{x}_i^T \boldsymbol{\beta},$$

$$\sum_{i=1}^{n} \mathbf{x}_i y_i = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \boldsymbol{\beta},$$

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{x}_i y_i\right). \tag{2.2.3}$$

Checking that this point is the maximum, we take the second derivative of the log-likelihood,

$$\frac{\partial^2}{\partial \boldsymbol{\beta}^2} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = \frac{\partial}{\partial \boldsymbol{\beta}} \left(\frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{x}_i \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)\right),$$

$$= -\frac{1}{\sigma^2} \sum_{i=1}^{n} \left(\mathbf{x}_i \mathbf{x}_i^T\right)^T,$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}^2} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = -\frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T. \tag{2.2.4}$$

Provided that $\mathbf{x}_i \neq \mathbf{0}$, we can see that the second derivative is negative definite by taking an arbitrary non-zero vector $\mathbf{a} \in \mathbb{R}^p$,

$$\mathbf{a}^T \left(-\frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right) \mathbf{a} = -\frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{a}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{a},$$

$$= -\frac{1}{\sigma^2} \sum_{i=1}^{n} \left(\mathbf{x}_i^T \mathbf{a}\right)^2,$$

$$< 0$$

Hence $\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{x}_i y_i\right)$ maximises the likelihood. We do a similar procedure for σ^2,

$$\log \mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = -\frac{n}{2} \log \left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2,$$

$$\frac{\partial}{\partial \sigma^2} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = \frac{\partial}{\partial \sigma^2} \left(-\frac{n}{2} \log \left(2\pi\right) - \frac{n}{2} \log \left(\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2 \right),$$

$$\frac{\partial}{\partial \sigma^2} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2. \tag{2.2.5}$$

Setting this to zero, obtain a value for σ^2,

$$-\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2 = 0,$$

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}\right)^2.$$

Note there is another solution for σ^2, but that violates our constant variance assumption. Then we check that this is a maximum point,

$$\frac{\partial^2}{\partial \left(\sigma^2\right)^2} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = \frac{\partial}{\partial \sigma^2} \left(-\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2 \right),$$

$$\frac{\partial^2}{\partial \left(\sigma^2\right)^2} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = \frac{n}{2\left(\sigma^2\right)^2} - \frac{1}{\left(\sigma^2\right)^3} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2, \tag{2.2.6}$$

$$= \frac{1}{\left(\sigma^2\right)^2} \left(\frac{n}{2} - \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2 \right).$$

Since it is not clear that this quantity is negative for all values of σ^2, we substitute our stationary point $\widehat{\sigma^2}$ and look at the convexity there,

$$\frac{\partial^2}{\partial \left(\sigma^2\right)^2} \log \mathcal{L}(\boldsymbol{\beta}, \widehat{\sigma^2} | \mathbf{y}) \bigg|_{\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2} = \frac{1}{\left(\widehat{\sigma^2}\right)^2} \left(\frac{n}{2} - \frac{\sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2}{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2} \right),$$

$$= -\frac{n}{2\left(\widehat{\sigma^2}\right)^2},$$

$$< 0.$$

So we have our estimator for $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right)^{-1} \left(\sum_{i=1}^{n} \mathbf{x}_i y_i\right)$ and our estimator for the variance σ^2, $\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}\right)^2$, but we must look at how

useful they are for us, in terms of their efficiency and accuracy, so we look at their properties.

2.2.2   Properties of Estimators

Consistency

Consistency allows our estimators to measure exactly what we want them to measure given enough data, we want to show that our maximum likelihood estimators are consistent. That is to say $\hat{\theta}_{\text{MLE}} \to \theta_0$ in probability as $n \to \infty$ where θ_0 is the true value of the parameter we wish to estimate. In our case we seek to show that

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \widehat{\sigma^2} \end{pmatrix} \to \boldsymbol{\theta}_0 = \begin{pmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{pmatrix} \text{ as } n \to \infty.$$

To begin this we know that the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ maximises the following function,

$$\log \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(y_i|\boldsymbol{\theta}).$$

If we modify the log likelihood, by multiplying it by $\frac{1}{n}$, it would not change the maximum of the function nor the estimates and since this is dependent on the number of realisations, we denote our new function with $\mathcal{L}_n$ as,

$$\log \mathcal{L}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \log f(y_i|\boldsymbol{\theta}).$$

Then let us consider two new functions,

$$\log \mathcal{L}^*(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}_0} \left[\log f(\mathbf{y}|\boldsymbol{\theta}) \right],$$
$$= \int_{-\infty}^{\infty} \log f(\mathbf{y}|\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}_0) dy.$$

Here the $\mathbb{E}_{\boldsymbol{\theta}_0}$ denotes the expected value with respect to the true value of the parameter. By the law of large numbers we notice that,

$$\log \mathcal{L}_n(\boldsymbol{\theta}) \to \mathbb{E}_{\boldsymbol{\theta}_0} \left[\log f(\mathbf{y}|\boldsymbol{\theta}) \right] = \log \mathcal{L}^*(\boldsymbol{\theta}) \text{ for } n \to \infty.$$

So this is true for all values of $\boldsymbol{\theta}$, the final thing we require is that $\boldsymbol{\theta}_0$ maximises $\log \mathcal{L}^*(\boldsymbol{\theta})$, this can be done by showing that $\log \mathcal{L}^*(\boldsymbol{\theta}) \leq \log \mathcal{L}^*(\boldsymbol{\theta}_0)$. We can being this by considering the following difference,

$$\log \mathcal{L}^*(\boldsymbol{\theta}) - \log \mathcal{L}^*(\boldsymbol{\theta}_0),$$
$$= \mathbb{E}_{\boldsymbol{\theta}_0} \left[\log f(\mathbf{y}|\boldsymbol{\theta}) \right] - \mathbb{E}_{\boldsymbol{\theta}_0} \left[\log f(\mathbf{y}|\boldsymbol{\theta}_0) \right],$$
$$= \mathbb{E}_{\boldsymbol{\theta}_0} \left[\log f(\mathbf{y}|\boldsymbol{\theta}) - \log f(\mathbf{y}|\boldsymbol{\theta}_0) \right],$$
$$= \mathbb{E}_{\boldsymbol{\theta}_0} \left[\log \frac{f(\mathbf{y}|\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta}_0)} \right].$$

Using the inequality $\log x \leq x - 1$, we arrive at,

$$\leq \mathbb{E}_{\boldsymbol{\theta}_0} \left[\frac{f(\mathbf{y}|\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta}_0)} - 1 \right],$$
$$= \int_{-\infty}^{\infty} \left(\frac{f(\mathbf{y}|\boldsymbol{\theta})}{f(y|\boldsymbol{\theta}_0)} - 1 \right) f(\mathbf{y}|\boldsymbol{\theta}_0) dy,$$
$$= \int_{-\infty}^{\infty} f(\mathbf{y}|\boldsymbol{\theta}) - f(\mathbf{y}|\boldsymbol{\theta}_0) dy,$$
$$= \int_{-\infty}^{\infty} f(\mathbf{y}|\boldsymbol{\theta}) dy - \int_{-\infty}^{\infty} f(\mathbf{y}|\boldsymbol{\theta}_0) dy,$$
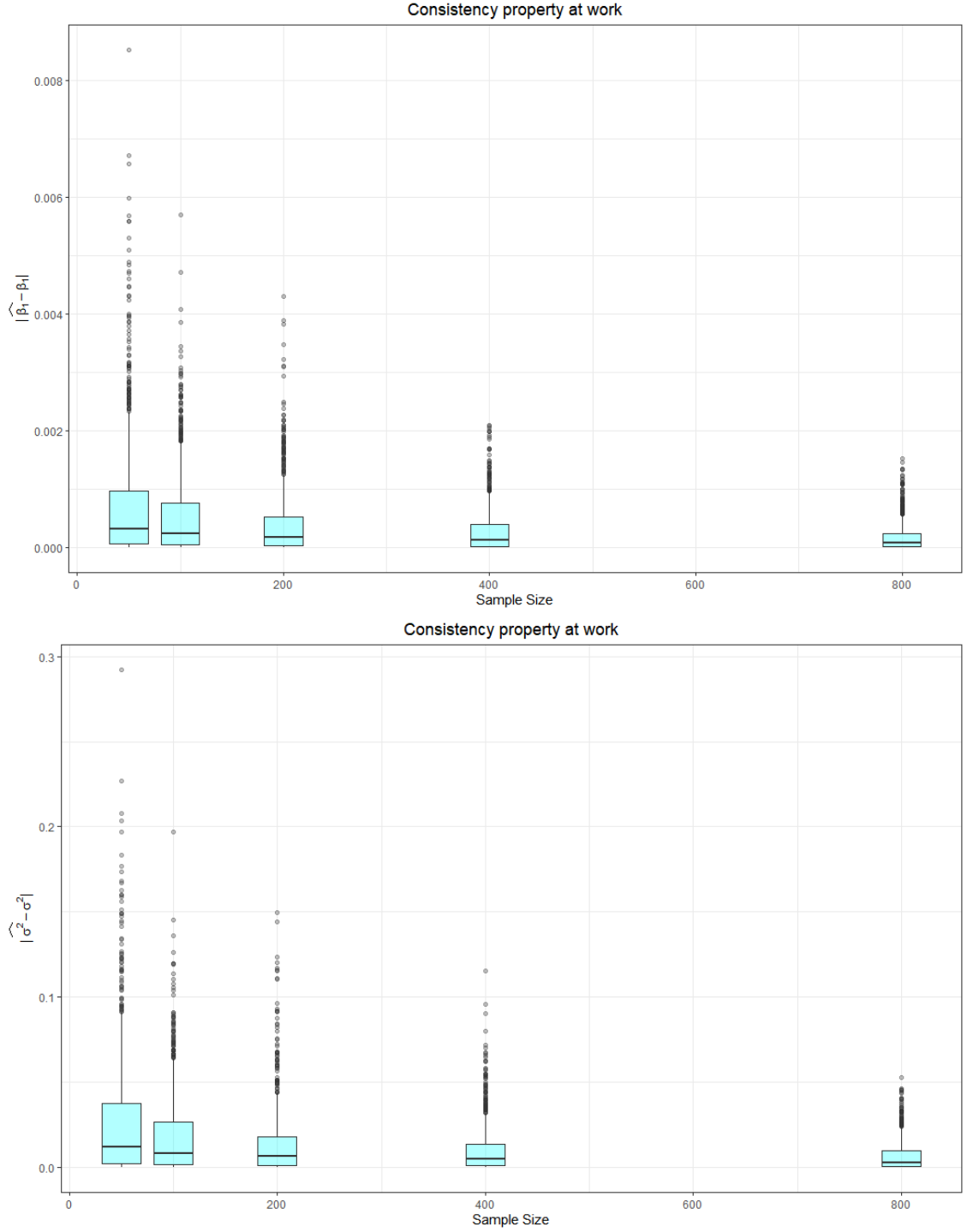$$= 1 - 1 = 0.$$

Since both are valid probability density functions, they integrate to 1. We therefore have $\log \mathcal{L}^*(\boldsymbol{\theta}) \leq \log \mathcal{L}^*(\boldsymbol{\theta}_0)$.

From these, we can conclude that $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ is consistent. As $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ maximises $\log \mathcal{L}_n(\boldsymbol{\theta})$, $\boldsymbol{\theta}_0$ maximises $\log \mathcal{L}^*(\boldsymbol{\theta})$ and as the sample size increases ($n \to \infty$) the function of the data $\log \mathcal{L}_n(\boldsymbol{\theta}) \to \log \mathcal{L}^*(\boldsymbol{\theta})$ in probability.

We can numerically show the consistency of the estimators, we take the simple case of simple linear regression where $p = 1$ and look at each of the estimators as the sample size increases. For $\hat{\beta}_0$, we can see that the absolute value difference between the estimator and the true value decreases as n increases, meaning the estimator appears to converge to the true value. This process involved generating some random error ϵ_i and some random data x_i and choosing some random $\boldsymbol{\beta}$ for a sample size of 50, obtaining the estimate for the regression parameters and computing the difference, doing that 1000 times then increasing the sample size till we reach a sample size of 800.

Consistency property at work
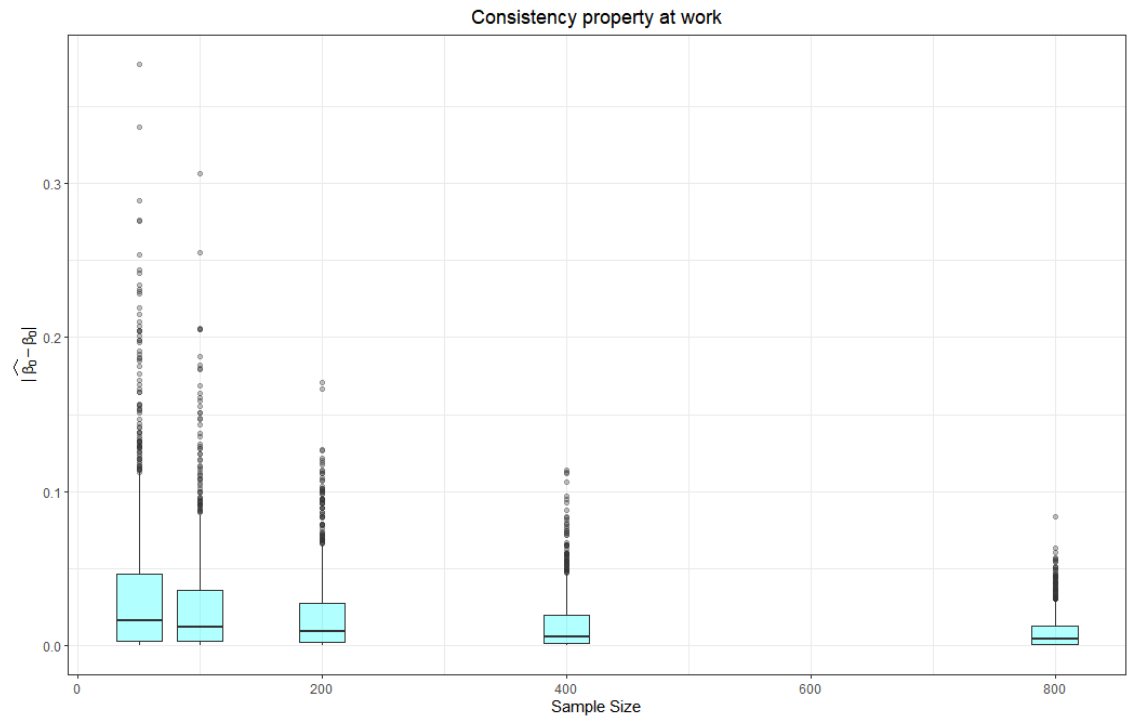
It is almost an identical situation for the $\hat{\beta}_1$ estimator as well as $\widehat{\sigma^2}$,

Consistency property at work

Consistency property at work

Fisher Information

To look at the distributional properties of the estimators, we can start by calculating the Fisher information, which measures the sensitivity of the estimates, acting as a way to judge the quality of the estimator. We define $\nabla_{\mathbf{x}}$ to be the gradient, eg

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix} \text{ where } \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n, \text{ because our vector of parameters is}$$

size $p + 2$, we define the $p + 2$ parameter Fisher information by the following,

10

$$\mathcal{I}(\boldsymbol{\theta}) = \mathbb{E}\left((\nabla_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta}))(\nabla_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta}))^T\right). \tag{2.2.7}$$

For simplicity of calculation we wish to show the following result,

$$-\mathbb{E}\left(\nabla^2_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta})\right) = \mathcal{I}(\boldsymbol{\theta}). \tag{2.2.8}$$

To show this, we begin with a simpler result,

$$\int_{-\infty}^{\infty} f(\mathbf{y}|\boldsymbol{\theta})dy = 1,$$

$$\nabla_{\boldsymbol{\theta}} \int_{-\infty}^{\infty} f(\mathbf{y}|\boldsymbol{\theta})dy = \nabla_{\boldsymbol{\theta}} 1,$$

$$\int_{-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} f(\mathbf{y}|\boldsymbol{\theta})dy = \mathbf{0},$$

$$\nabla_{\boldsymbol{\theta}} \int_{-\infty}^{\infty} \nabla_{\boldsymbol{\theta}} f(\mathbf{y}|\boldsymbol{\theta})dy = \nabla_{\boldsymbol{\theta}} \mathbf{0},$$

$$\int_{-\infty}^{\infty} \nabla^2_{\boldsymbol{\theta}} f(\mathbf{y}|\boldsymbol{\theta})dy = 0 \in \mathcal{M}_{(p+2)\times(p+2)}. \tag{2.2.9}$$

This is a useful result we will apply later on, but to prove the equivalence between 2.2.7 and 2.2.8 we will begin with the hessian from the latter,

$$\begin{aligned}
\nabla^2_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta}) &= \nabla_{\boldsymbol{\theta}}\left(\nabla_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta})\right), \\
&= \nabla_{\boldsymbol{\theta}}\left((\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}))\mathcal{L}^{-1}(\boldsymbol{\theta})\right), \\
&= \left(\nabla^2_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})\right)\mathcal{L}^{-1}(\boldsymbol{\theta}) - (\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}))(\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}))^T\mathcal{L}^{-2}(\boldsymbol{\theta}), \\
-\mathbb{E}\left(\nabla^2_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta})\right) &= -\mathbb{E}\left(\left(\nabla^2_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})\right)\mathcal{L}^{-1}(\boldsymbol{\theta}) - (\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}))(\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}))^T\mathcal{L}^{-2}(\boldsymbol{\theta})\right), \\
&= \mathbb{E}\left((\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}))(\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}))^T\mathcal{L}^{-2}(\boldsymbol{\theta})\right) - \mathbb{E}\left(\left(\nabla^2_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})\right)\mathcal{L}^{-1}(\boldsymbol{\theta})\right), \\
&= \int_{-\infty}^{\infty}\left((\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}))(\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}))^T\mathcal{L}^{-2}(\boldsymbol{\theta})\right)f(\mathbf{y}|\boldsymbol{\theta})dy- \\
&\qquad \int_{-\infty}^{\infty}\left(\left(\nabla^2_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})\right)\mathcal{L}^{-1}(\boldsymbol{\theta})\right)f(\mathbf{y}|\boldsymbol{\theta})dy.
\end{aligned}$$

Here we have to note that $f(y|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta})$ as they are the density and the likelihood of the same distribution, this allows for some simplification,

$$-\mathbb{E}\left(\nabla_{\boldsymbol{\theta}}^2 \log \mathcal{L}(\boldsymbol{\theta})\right) = \int_{-\infty}^{\infty} \left(\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})\mathcal{L}^{-1}(\boldsymbol{\theta})\right)\left(\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})\mathcal{L}^{-1}(\boldsymbol{\theta})\right)^T f(\mathbf{y}|\boldsymbol{\theta})dy-$$

$$\int_{-\infty}^{\infty} \nabla_{\boldsymbol{\theta}}^2 \mathcal{L}(\boldsymbol{\theta})dy,$$

$$= \int_{-\infty}^{\infty} \left(\nabla_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta})\right)\left(\nabla_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta})\right)^T f(\mathbf{y}|\boldsymbol{\theta})dy-$$

$$\int_{-\infty}^{\infty} \nabla_{\boldsymbol{\theta}}^2 f(\mathbf{y}|\boldsymbol{\theta})dy.$$

Using 2.2.9, and reverting the integral back into an expectation,

$$-\mathbb{E}\left(\nabla_{\boldsymbol{\theta}}^2 \log \mathcal{L}(\boldsymbol{\theta})\right) = \mathbb{E}\left(\left(\nabla_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta})\right)\left(\nabla_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta})\right)^T\right) - 0,$$

$$= \mathcal{I}(\boldsymbol{\theta}).$$

With that proven, we can calculate the Fisher information for the maximum likelihood estimates,

$$\mathcal{I}(\boldsymbol{\beta}, \sigma^2) = -\mathbb{E}\begin{pmatrix} \frac{\partial^2}{\partial \boldsymbol{\beta}^2} \log \mathcal{L} & \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \sigma^2} \log \mathcal{L} \\ \frac{\partial^2}{\partial \sigma^2 \partial \boldsymbol{\beta}} \log \mathcal{L} & \frac{\partial^2}{\partial (\sigma^2)^2} \log \mathcal{L} \end{pmatrix}.$$

Some of parts of this expression we have already calculated earlier, so we will substitute (2.2.4) and (2.2.6), but we need to derive one more time (2.2.2) and (2.2.5) before substituting those. From (2.2.2),

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbf{x}_i \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right),$$

$$\frac{\partial^2}{\partial \sigma^2 \partial \boldsymbol{\beta}} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2|\mathbf{y}) = -\frac{1}{(\sigma^2)^2} \sum_{i=1}^{n} \mathbf{x}_i \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta}\right),$$

$$= \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \sigma^2} \log \mathcal{L}(\boldsymbol{\beta}, \sigma^2|\mathbf{y}).$$

So we can construct the Fisher information matrix,

$$\mathcal{I}(\boldsymbol{\beta}, \sigma^2) = -\mathbb{E}\begin{pmatrix} -\frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T & -\frac{1}{(\sigma^2)^2}\sum_{i=1}^n \mathbf{x}_i\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right) \\ -\frac{1}{(\sigma^2)^2}\sum_{i=1}^n \mathbf{x}_i\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right) & \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^n\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2 \end{pmatrix},$$

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T & \mathbb{E}\left(\frac{1}{(\sigma^2)^2}\sum_{i=1}^n \mathbf{x}_i\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)\right) \\ \mathbb{E}\left(\frac{1}{(\sigma^2)^2}\sum_{i=1}^n \mathbf{x}_i\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)\right) & \mathbb{E}\left(-\frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}\sum_{i=1}^n\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2\right) \end{pmatrix},$$

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T & \frac{1}{(\sigma^2)^2}\sum_{i=1}^n \mathbf{x}_i\left(\mathbb{E}(y_i) - \mathbf{x}_i^T\boldsymbol{\beta}\right) \\ \frac{1}{(\sigma^2)^2}\sum_{i=1}^n \mathbf{x}_i\left(\mathbb{E}(y_i) - \mathbf{x}_i^T\boldsymbol{\beta}\right) & -\frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}\sum_{i=1}^n\mathbb{E}\left(\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2\right) \end{pmatrix},$$

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T & \frac{1}{(\sigma^2)^2}\sum_{i=1}^n \mathbf{x}_i\left(\mathbf{x}_i^T\boldsymbol{\beta} - \mathbf{x}_i^T\boldsymbol{\beta}\right) \\ \frac{1}{(\sigma^2)^2}\sum_{i=1}^n \mathbf{x}_i\left(\mathbf{x}_i^T\boldsymbol{\beta} - \mathbf{x}_i^T\boldsymbol{\beta}\right) & -\frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}\sum_{i=1}^n(\sigma^2) \end{pmatrix},$$

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T & 0 \\ 0 & -\frac{n}{2(\sigma^2)^2} + \frac{n}{(\sigma^2)} \end{pmatrix}.$$

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{pmatrix}. \tag{2.2.10}$$

Asymptotic Normality

We want to also know the distributional properties of $\widehat{\boldsymbol{\beta}}$, so we can apply statistical tests to it, find out how it is distributed looking at 2.2.3,

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\sum_{i=1}^n \mathbf{x}_iy_i\right).$$

From the right hand side, y_i is the only random variable, with the following distribution,

$$y_i \sim \mathcal{N}\left(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2\right).$$

We can then transform the left hand side using properties of the normal distribution,

$$\mathbf{x}_iy_i \sim \mathcal{N}\left(\mathbf{x}_i\mathbf{x}_i^T\boldsymbol{\beta}, \mathbf{x}_i\sigma^2\mathbf{x}_i^T\right),$$

$$\sum_{i=1}^n \mathbf{x}_iy_i \sim \mathcal{N}\left(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\right),$$

$$\left(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\sum_{i=1}^n \mathbf{x}_iy_i\right) \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2\left(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\right),$$

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2\left(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\right).$$

We see that the covariance here does coincide with the well known result that the variance of the parameter is the inverse $\boldsymbol{\beta}$ term of the fisher information in 2.2.10.

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right)^{-1}\right),$$

$$\frac{1}{\sigma}\left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right)^{\frac{1}{2}} \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \sim \mathcal{N}\left(\mathbf{0}, \mathbb{1}_p\right).$$

This is for a known σ^2, we'll leave it here, as to parallel with our findings in the complex regression case, as we do not deal with the distribution of the estimate for the covariance there.

Complex Numbered Regression Model

3.1 Complex Model

Now we seek to construct the complex number model, find out how to estimators look like and observe their properties. Making the move over from the real scenario, our observations become complex along with the data and errors, complex meaning they have the form $a + jb$ where j is the complex number $\sqrt{-1}$. Ultimately for the i-th observation out of n, our underlying model will look like,

$$Z_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i.$$

Where the complex vector $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip})^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$. However to follow Definition 1 from Ducharme [8] we will augment our model, not removing anything, but only adding conjugates of each of the parameters so we can use the nice form of the density, our model for each observations becomes,

$$\underset{\sim}{Z}_i = \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} + \underset{\sim}{\epsilon}_i. \tag{3.1.1}$$

And the full model with n observations,

$$\mathbf{\underset{\sim}{Z}} = \underset{\sim}{X} \underset{\sim}{\boldsymbol{\beta}} + \boldsymbol{\underset{\sim}{\epsilon}}. \tag{3.1.2}$$

There is a lot of notation here to unpack and we'll go through them sequentially. Firstly we let z be a complex random variable, so $\mathbf{z}$ is a complex vector, with $\mathbf{z}^*$ denoting the conjugate of that vector and $\mathbf{z}^{\mathcal{H}} = (\mathbf{z}^*)^T = (\mathbf{z}^T)^*$ to denote the complex conjugate transpose of $\mathbf{z}$.

We then define $\underset{\sim}{z}_i = (z_i, z_i^*)^T$, a sort of stacked complex vector, different from stacking the real parts and imaginary parts of a complex vector [6]. Then when we put this in vector form we obtain, $\underset{\sim}{z} = (z_1, z_1^*, \ldots, z_n, z_n^*)^T$.

Almost similarly we can define this stacking operator for a matrix X, first we look a single data point $x_i = (1, x_{i1}, \ldots, x_{ip})^T$, when stacked this becomes,

$$\underset{\sim}{x}_i = \begin{pmatrix} 1 & x_{i1} & \ldots & x_{ip} & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 1 & x_{i1}^* & \ldots & x_{ip}^* \end{pmatrix}.$$

Then for the full matrix of data,

$$\underset{\sim}{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & x_{11}^* & \dots & x_{1p}^* \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & x_{n1}^* & \dots & x_{np}^* \end{pmatrix}.$$

This looks like a mess, although it is, when you take in consideration that the vector of regression parameters $\underset{\sim}{\boldsymbol{\beta}}$ will look like $\underset{\sim}{\boldsymbol{\beta}} = \left(\beta_0, \dots, \beta_p, \beta_0^*, \dots, \beta_p^*\right)^T$, an expression like $\left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}}\right)$ makes sense.

3.1.1 Assumptions

Different model means different assumptions that need to be made, the assumptions would mostly be clear and similar to the real case.

- Each augmented pair of errors are independent and identically distributed complex improper normal with 0 mean and 2 by 2 covariance matrix Γ which we will get into in a later subsection. Eg $\epsilon_i \sim \mathcal{CN}(0, \Gamma)$.
- All these parameters in model 3.1.1 are complex, eg random observations Z_i, data vector $\mathbf{x}_i$, regression parameter $\boldsymbol{\beta}$ and the error ϵ_i are all complex.
- There is a linear relationship between the observation Z_i and the data $\mathbf{x}_i$.

3.1.2 Improper Complex Normal Distribution

We will mostly use the notation borrowed from the Ducharme paper [8] however changing one or two things to suit our needs. One way to look at complex normal random variables, is that the real and imaginary parts of Z are from some normal distribution and that there is also some correlation between the two parts. So each observation $Z_i = \mathcal{X}_1 + j\mathcal{X}_2$ is really from a two dimensional multivariate normal,

$$\boldsymbol{\mathcal{X}} = (\mathcal{X}_1, \mathcal{X}_2)^T \sim \mathcal{N}_2(\boldsymbol{\mu}_\mathcal{X}, \Sigma_\mathcal{X}).$$

Where $\Sigma_\mathcal{X}$ is a positive definite covariance matrix. A transformation can be made from $\boldsymbol{\mathcal{X}}$ to $\underset{\sim}{Z}_i$ using a handy matrix found in both [8] and [3],

$$M = \frac{1}{2}\begin{pmatrix} \mathbb{1}_n & \mathbb{1}_n \\ -j\mathbb{1}_n & j\mathbb{1}_n \end{pmatrix}, \qquad M^{-1} = \begin{pmatrix} \mathbb{1}_n & j\mathbb{1}_n \\ \mathbb{1}_n & -j\mathbb{1}_n \end{pmatrix} = 2M^{\mathcal{H}},$$

$$\boldsymbol{\mathcal{X}} = M\underset{\sim}{Z}_i = M\begin{pmatrix} Z_i \\ Z_i^* \end{pmatrix}.$$

Here the $\mathbb{1}_n$ denotes the identity matrix of size n, and n in this case is 1. We can then use this matrix to show a relationship between $\Sigma_\mathcal{X}$ and Γ, Γ which we will define as,

$$\Gamma = \mathbb{E}\left[\underline{\epsilon_i}\underline{\epsilon_i}^{\mathcal{H}}\right],$$

$$= \mathbb{E}\left[\begin{pmatrix}\epsilon_i \\ \epsilon_i^*\end{pmatrix}\begin{pmatrix}\epsilon_i \\ \epsilon_i^*\end{pmatrix}^{\mathcal{H}}\right],$$

$$= \begin{pmatrix}\mathbb{E}\left[\epsilon_i\epsilon_i^*\right] & \mathbb{E}\left[\epsilon_i\epsilon_i\right] \\ \mathbb{E}\left[\epsilon_i^*\epsilon_i^*\right] & \mathbb{E}\left[\epsilon_i^*\epsilon_i\right]\end{pmatrix}, \tag{3.1.3}$$

$$= \begin{pmatrix}K & J \\ J^* & K\end{pmatrix}. \tag{3.1.4}$$

On a side, we can get complex valued J, however K is always real valued. Commonly K is referred to as the covariance and J is called pseudo-covariance, in the proper complex normal distribution J is set to 0, hence we use the improper complex normal to add some level of generalisation and capture more information that might otherwise be left out of the proper complex normal distribution.

This also means that Γ is a 2 by 2 matrix, however when considering an augmented vector of $\underline{\mathbf{z}} = (z_1, z_1^*, ..., z_n, z_n^*)^T$, Γ will be block diagonal matrix with it's 2 by 2 self as the blocks and in another case if $\underline{\mathbf{z}} = (\mathbf{z}^T, \mathbf{z}^{*T})^T$ then the $2n$ by $2n$ matrix Γ using the notation in 3.1.4 will be in weird form, K's on the main diagonal and J's on the n-th super-diagonal and J^*'s on the n-th sub-diagonal, here's an illustration to better explain,

$$\Gamma = \begin{bmatrix}K\mathbb{1}_n & J\mathbb{1}_n \\ J^*\mathbb{1}_n & K\mathbb{1}_n\end{bmatrix}.$$

This was not super important, but the notation used for the three different Γ will not be differentiated, because generally they are dependent on circumstance and their meanings are similar and are all hermitian. Moving on from this digression, the relationship between Σ_χ and Γ, taking the M matrix with n = 1 and the 3.1.3 definition of Γ (2 by 2),

$$M\Gamma M^{\mathcal{H}} = \frac{1}{4}\begin{pmatrix}1 & 1 \\ -j & j\end{pmatrix}\begin{pmatrix}\mathbb{E}\left[\epsilon_i\epsilon_i^*\right] & \mathbb{E}\left[\epsilon_i\epsilon_i\right] \\ \mathbb{E}\left[\epsilon_i^*\epsilon_i^*\right] & \mathbb{E}\left[\epsilon_i^*\epsilon_i\right]\end{pmatrix}\begin{pmatrix}1 & j \\ 1 & -j\end{pmatrix},$$

$$= \frac{1}{4}\left(\begin{array}{c}\mathbb{E}\left[\epsilon_i\epsilon_i^*\right] + \mathbb{E}\left[\epsilon_i\epsilon_i\right] + \mathbb{E}\left[\epsilon_i^*\epsilon_i^*\right] + \mathbb{E}\left[\epsilon_i^*\epsilon_i\right] \\ -j\mathbb{E}\left[\epsilon_i\epsilon_i^*\right] + j\mathbb{E}\left[\epsilon_i\epsilon_i\right] - j\mathbb{E}\left[\epsilon_i^*\epsilon_i^*\right] + j\mathbb{E}\left[\epsilon_i^*\epsilon_i\right]\end{array}\right.$$

$$\left.\begin{array}{c}j\mathbb{E}\left[\epsilon_i\epsilon_i\right] - j\mathbb{E}\left[\epsilon_i\epsilon_i^*\right] + j\mathbb{E}\left[\epsilon_i^*\epsilon_i^*\right] - j\mathbb{E}\left[\epsilon_i^*\epsilon_i\right] \\ \mathbb{E}\left[\epsilon_i\epsilon_i^*\right] - \mathbb{E}\left[\epsilon_i\epsilon_i\right] - \mathbb{E}\left[\epsilon_i^*\epsilon_i^*\right] + \mathbb{E}\left[\epsilon_i^*\epsilon_i\right]\end{array}\right).$$

This mess here, in-fact can be simplified, first by taking a look at the three different expectations and using $\epsilon_i = \text{Re}\left(\epsilon_i\right) + j\text{Im}\left(\epsilon_i\right)$, the expectations can be written as,

$$\mathbb{E}\left[\epsilon_i \epsilon_i\right] = \mathbb{E}\left[\text{Re}(\epsilon_i)^2\right] + 2j\mathbb{E}\left[\text{Re}(\epsilon_i)\text{Im}(\epsilon_i)\right] - \mathbb{E}\left[\text{Im}(\epsilon_i)^2\right],$$
$$\mathbb{E}\left[\epsilon_i^* \epsilon_i^*\right] = \mathbb{E}\left[\text{Re}(\epsilon_i)^2\right] - 2j\mathbb{E}\left[\text{Re}(\epsilon_i)\text{Im}(\epsilon_i)\right] - \mathbb{E}\left[\text{Im}(\epsilon_i)^2\right],$$
$$\mathbb{E}\left[\epsilon_i^* \epsilon_i\right] = \mathbb{E}\left[\epsilon_i \epsilon_i^*\right] = \mathbb{E}\left[\text{Re}(\epsilon_i)^2\right] + \mathbb{E}\left[\text{Im}(\epsilon_i)^2\right].$$

We can this substitute all these into the $M\Gamma M^{\mathcal{H}}$ matrix multiplication, netting the very nice result,

$$M\Gamma M^{\mathcal{H}} = \begin{pmatrix} \mathbb{E}\left[\text{Re}(\epsilon_i)^2\right] & \mathbb{E}\left[\text{Re}(\epsilon_i)\text{Im}(\epsilon_i)\right] \\ \mathbb{E}\left[\text{Re}(\epsilon_i)\text{Im}(\epsilon_i)\right] & \mathbb{E}\left[\text{Im}(\epsilon_i)^2\right] \end{pmatrix},$$
$$M\Gamma M^{\mathcal{H}} = \Sigma_\chi. \tag{3.1.5}$$

As this is the definition of the multivariate covariance, as χ_1 and χ_2 are the pair of random variables with variances $\text{Re}(\epsilon_i)$ and $\text{Im}(\epsilon_i)$ respectively. This result also leads to the following conclusions, using Leibniz Formula for determinants midway,

$$
\begin{aligned}
|\Sigma_\chi| &= |M\Gamma M^{\mathcal{H}}|, \\
&= |M| \cdot |\Gamma| \cdot |M^{\mathcal{H}}|, \\
&= \left| \frac{1}{2}\begin{pmatrix} \mathbb{1}_n & \mathbb{1}_n \\ -j\mathbb{1}_n & j\mathbb{1}_n \end{pmatrix} \right| \cdot |\Gamma| \cdot \left| \frac{1}{2}\begin{pmatrix} \mathbb{1}_n & j\mathbb{1}_n \\ \mathbb{1}_n & -j\mathbb{1}_n \end{pmatrix} \right|, \\
&= 2^{-2n}\left| \mathbb{1}_n - \mathbb{1}_n \left(j\mathbb{1}_n\right)^{-1}\left(-j\mathbb{1}_n\right) \right| \cdot |j\mathbb{1}_n| \cdot |\Gamma| \\
&\qquad\qquad \cdot 2^{-2n} \cdot \left| \mathbb{1}_n - j\mathbb{1}_n\left(-j\mathbb{1}_n\right)^{-1}\mathbb{1}_n \right| \cdot |-j\mathbb{1}_n|, \\
&= 2^{-2n} \cdot |\mathbb{1}_n + \mathbb{1}_n| \cdot j^n \cdot |\Gamma| \cdot 2^{-2n} \cdot |\mathbb{1}_n + \mathbb{1}_n| \cdot (-j)^n, \\
&= 2^{-2n} \cdot 2^n \cdot j^n \cdot |\Gamma| \cdot 2^{-2n} \cdot 2^n \cdot (-j)^n, \\
&= 2^{-2n} \cdot |\Gamma|.
\end{aligned}
$$

And since Γ and Σ_χ are positive definite, they are invertible,

$$
\begin{aligned}
\Sigma_\chi^{-1} &= \left(M\Gamma M^{\mathcal{H}}\right)^{-1}, \\
&= M^{-1\mathcal{H}}\Gamma^{-1}M^{-1}, \\
&= 4M\Gamma^{-1}M^{\mathcal{H}}.
\end{aligned}
$$

With these nice relations that can be generalised to the $2n$-th dimension, we can look at rewriting the multivariate normal density, suppose $\boldsymbol{X} \sim \mathcal{N}_{2n}\left(\boldsymbol{\mu}_\chi, \Sigma_\chi\right)$, recall that $\mathbf{x} = M\underset{\sim}{\mathbf{z}}$ and we'll denote $\mathbb{E}\left[\underset{\sim}{\mathbf{z}}\right] = \underset{\sim}{\boldsymbol{\mu}}$, so $\boldsymbol{\mu}_\chi = M\underset{\sim}{\boldsymbol{\mu}}$,

$$f_\chi(\mathbf{x}) = \frac{1}{(2\pi)^n |\Sigma_\chi|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_\chi)^T \Sigma_\chi^{-1} (\mathbf{x} - \boldsymbol{\mu}_\chi)\right),$$

$$= \frac{1}{(2\pi)^n (2^{-2n}|\Gamma|)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_\chi)^T M^{-1\mathcal{H}}\Gamma^{-1}M^{-1} (\mathbf{x} - \boldsymbol{\mu}_\chi)\right),$$

$$= \frac{1}{\pi^n|\Gamma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \left(M^{-1*}\mathbf{x}^* - M^{-1*}\boldsymbol{\mu}_\chi^*\right)^T \Gamma^{-1} \left(M^{-1}\mathbf{x} - M^{-1}\boldsymbol{\mu}_\chi\right)\right),$$

$$f_Z(\mathbf{z}) = \frac{1}{\pi^n|\Gamma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \left(\underset{\sim}{\mathbf{z}}^* - \underset{\sim}{\boldsymbol{\mu}}^*\right)^T \Gamma^{-1} \left(\underset{\sim}{\mathbf{z}} - \underset{\sim}{\boldsymbol{\mu}}\right)\right),$$

$$f_Z(\mathbf{z}) = \frac{1}{\pi^n|\Gamma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \left(\underset{\sim}{\mathbf{z}} - \underset{\sim}{\boldsymbol{\mu}}\right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{\mathbf{z}} - \underset{\sim}{\boldsymbol{\mu}}\right)\right).$$

Which leads us to the definition of the complex normal distribution where $Z \sim \mathcal{CN}_n (\boldsymbol{\mu}, \Gamma)$.

So in the case of our complex model which has observations that are Z with mean $X\boldsymbol{\beta}$, it is distributed $Z \sim \mathcal{CN}_n (\boldsymbol{\mu}, \Gamma)$ with a density of,

$$f_Z(\mathbf{z}) = \frac{1}{\pi^n|\Gamma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \left(\underset{\sim}{\mathbf{z}} - \underset{\sim}{X\boldsymbol{\beta}}\right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{\mathbf{z}} - \underset{\sim}{X\boldsymbol{\beta}}\right)\right)$$

Then for each observation $Z_i \sim \mathcal{CN}(x_i\boldsymbol{\beta}, \Gamma)$

$$f_{Z_i}(z_i) = \frac{1}{\pi|\Gamma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i\boldsymbol{\beta}}\right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i\boldsymbol{\beta}}\right)\right). \tag{3.1.6}$$

Since each of the ϵ_i were independent, so are the z_i, this will be helpful for analytically solving the likelihood.

3.1.3 Likelihood

The next step to obtaining estimators for parameters in our model is to formulate the likelihood, we shall use a simplification to help ease notation letting $\boldsymbol{\theta} = \left(\underset{\sim}{\boldsymbol{\beta}}, \underset{\sim}{\boldsymbol{\beta}}^*, \Gamma\right)$.

$$\mathcal{L}(\boldsymbol{\theta}) = f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta}),$$

$$= \prod_{i=1}^n f_{Z_i}(z_i|\boldsymbol{\theta}),$$

$$= \prod_{i=1}^n \left(\frac{1}{\pi|\Gamma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i\boldsymbol{\beta}}\right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i\boldsymbol{\beta}}\right)\right)\right).$$

Since this is a distribution, we needn't worry about taking the log of the likelihood since it is real,

$$\log \mathcal{L}(\boldsymbol{\theta}) = \log \prod_{i=1}^{n} \left(\frac{1}{\pi |\Gamma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right) \right) \right),$$

$$\log \mathcal{L}(\boldsymbol{\theta}) = -n \log \left(\pi |\Gamma|^{\frac{1}{2}} \right) - \frac{1}{2} \sum_{i=1}^{n} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right). \qquad (3.1.7)$$

Now we can get maximising and obtaining maximum likelihood estimates.

3.2 Estimators

3.2.1 Derivation

Calculation of the estimators involves complex vector calculus specifically the following identities [9], where we take $\mathbf{z}$ to be the complex vector variable, $\mathbf{c}$ to be a constant vector, $\mathbf{z}^*$ to denote the complex conjugate of $\mathbf{z}$ and $\mathbf{c}^T$ to be the transpose of $\mathbf{c}$.

$$\frac{\partial}{\partial \mathbf{z}} \mathbf{c}^T \mathbf{z} = \mathbf{c}, \qquad\qquad \frac{\partial}{\partial \mathbf{z}} \mathbf{c}^T \mathbf{z}^* = \mathbf{0},$$
$$\frac{\partial}{\partial \mathbf{z}^*} \mathbf{c}^T \mathbf{z} = \mathbf{0}, \qquad\qquad \frac{\partial}{\partial \mathbf{z}^*} \mathbf{c}^T \mathbf{z}^* = \mathbf{c}. \qquad (3.2.1)$$

We can then proceed to differentiate the log-likelihood with respect to our complex vector,

$$\log \mathcal{L}(\boldsymbol{\theta}) = -n \log \left(\pi |\Gamma|^{\frac{1}{2}} \right) - \frac{1}{2} \sum_{i=1}^{n} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right),$$

$$\frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \log \mathcal{L}(\boldsymbol{\theta}) = \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(-n \log \left(\pi |\Gamma|^{\frac{1}{2}} \right) - \frac{1}{2} \sum_{i=1}^{n} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right) \right),$$

$$= -\mathbf{0} - \frac{1}{2} \cdot \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(\sum_{i=1}^{n} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z_i} - \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right) \right),$$

$$= -\frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(z_i^{\mathcal{H}} \Gamma^{-1} z_i - z_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} - \underset{\sim}{\boldsymbol{\beta}}^{\mathcal{H}} \underset{\sim}{x_i}^{\mathcal{H}} \Gamma^{-1} z_i + \underset{\sim}{\boldsymbol{\beta}}^{\mathcal{H}} \underset{\sim}{x_i}^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right),$$

$$= -\frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(z_i^{\mathcal{H}} \Gamma^{-1} z_i \right) + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(z_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right)$$

$$\qquad + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(\underset{\sim}{\boldsymbol{\beta}}^{\mathcal{H}} \underset{\sim}{x_i}^{\mathcal{H}} \Gamma^{-1} z_i \right) - \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(\underset{\sim}{\boldsymbol{\beta}}^{\mathcal{H}} \underset{\sim}{x_i}^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x_i}\underset{\sim}{\boldsymbol{\beta}} \right),$$

$$= -\mathbf{0} + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(\left(\underset{\sim}{x_i}^{T} \left(\Gamma^{-1} \right)^{T} z_i^{*} \right)^{T} \underset{\sim}{\boldsymbol{\beta}} \right) + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(\underset{\sim}{\boldsymbol{\beta}}^{\mathcal{H}} \underset{\sim}{x_i}^{\mathcal{H}} \Gamma^{-1} z_i \right)$$

$$\qquad - \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(\left(\underset{\sim}{x_i}^{T} \left(\Gamma^{-1} \right)^{T} \underset{\sim}{x_i}^{*} \underset{\sim}{\boldsymbol{\beta}}^{*} \right)^{T} \underset{\sim}{\boldsymbol{\beta}} \right).$$

Using 3.2.1, when differentiating with respect to $\underset{\sim}{\boldsymbol{\beta}}$ we keep $\underset{\sim}{\boldsymbol{\beta}^*}$ constant,

$$\frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \log \mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^{n} \left(\underset{\sim}{x}_i^T \left(\Gamma^{-1}\right)^T \underset{\sim}{z}_i^* \right) + \mathbf{0} - \frac{1}{2} \sum_{i=1}^{n} \left(\underset{\sim}{x}_i^T \left(\Gamma^{-1}\right)^T \underset{\sim}{x}_i^* \underset{\sim}{\boldsymbol{\beta}}^* \right).$$

When we set,

$$\frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \log \mathcal{L}(\boldsymbol{\theta}) = \mathbf{0},$$

$$\mathbf{0} = \frac{1}{2} \sum_{i=1}^{n} \underset{\sim}{x}_i^T \left(\Gamma^{-1}\right)^T \underset{\sim}{z}_i^* - \frac{1}{2} \sum_{i=1}^{n} \underset{\sim}{x}_i^T \left(\Gamma^{-1}\right)^T \underset{\sim}{x}_i^* \underset{\sim}{\boldsymbol{\beta}}^*,$$

$$\sum_{i=1}^{n} \underset{\sim}{x}_i^T \left(\Gamma^{-1}\right)^T \underset{\sim}{z}_i^* = \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^T \left(\Gamma^{-1}\right)^T \underset{\sim}{x}_i^* \right) \underset{\sim}{\boldsymbol{\beta}}^*,$$

$$\widehat{\underset{\sim}{\boldsymbol{\beta}}^*} = \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^T \left(\Gamma^{-1}\right)^T \underset{\sim}{x}_i^* \right)^{-1} \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^T \left(\Gamma^{-1}\right)^T \underset{\sim}{z}_i^* \right). \qquad (3.2.2)$$

Similarly we could differentiate the log-likelihood with respect to $\underset{\sim}{\boldsymbol{\beta}^*}$ the conjugate of $\underset{\sim}{\boldsymbol{\beta}}$,

$$\log \mathcal{L}(\boldsymbol{\theta}) = -n \log \left(\pi |\Gamma|^{\frac{1}{2}} \right) - \frac{1}{2} \sum_{i=1}^{n} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} \right),$$

$$\frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}^*} \log \mathcal{L}(\boldsymbol{\theta}) = \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}^*} \left(-n \log \left(\pi |\Gamma|^{\frac{1}{2}} \right) - \frac{1}{2} \sum_{i=1}^{n} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} \right) \right),$$

$$= -\mathbf{0} - \frac{1}{2} \cdot \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}^*} \left(\sum_{i=1}^{n} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} \right) \right),$$

$$= -\frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}} \left(\underset{\sim}{z}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{z}_i - \underset{\sim}{z}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} - \underset{\sim}{\boldsymbol{\beta}}^{\mathcal{H}} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{z}_i + \underset{\sim}{\boldsymbol{\beta}}^{\mathcal{H}} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} \right),$$

$$= -\frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}^*} \left(\underset{\sim}{z}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{z}_i \right) + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}^*} \left(\underset{\sim}{z}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} \right)$$

$$+ \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}^*} \left(\left(\underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{z}_i \right)^{T} \underset{\sim}{\boldsymbol{\beta}}^* \right)$$

$$- \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}^*} \left(\left(\underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} \right)^{T} \underset{\sim}{\boldsymbol{\beta}}^* \right),$$

$$\frac{\partial}{\partial \underset{\sim}{\boldsymbol{\beta}}^*} \log \mathcal{L}(\boldsymbol{\theta}) = -\mathbf{0} + \mathbf{0} + \frac{1}{2} \sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{z}_i - \frac{1}{2} \sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \underset{\sim}{\boldsymbol{\beta}} = \mathbf{0},$$

$$\widehat{\underset{\sim}{\boldsymbol{\beta}}} = \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \right)^{-1} \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{z}_i \right). \tag{3.2.3}$$

We should notice here that the estimators are indeed conjugates of each other,

$$\left(\widehat{\underset{\sim}{\boldsymbol{\beta}}} \right)^* = \left(\left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \right)^{-1} \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{z}_i \right) \right)^*,$$

$$= \left(\left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \right)^{-1} \right)^* \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{z}_i \right)^*,$$

$$= \left(\left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \right)^* \right)^{-1} \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{T} \left(\Gamma^{-1} \right)^* \underset{\sim}{z}_i^* \right),$$

$$= \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{T} \left(\Gamma^{-1} \right)^* \underset{\sim}{x}_i^* \right)^{-1} \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \left(\Gamma^{-1} \right)^* \underset{\sim}{z}_i^* \right),$$

Looking at the covariance matrix,

22

$$\left(\Gamma^{-1}\right)^* = \left(\Gamma^*\right)^{-1} = \left(\begin{pmatrix} K & J \\ J^* & K \end{pmatrix}^*\right)^{-1} = \begin{pmatrix} K^* & J^* \\ J^{**} & K^* \end{pmatrix}^{-1} = \begin{pmatrix} K & J^* \\ J & K \end{pmatrix}^{-1} = \left(\Gamma^{-1}\right)^T.$$

So the next step is,

$$\left(\widehat{\underset{\sim}{\beta}}\right)^* = \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^T \left(\Gamma^{-1}\right)^T \underset{\sim}{x}_i^*\right)^{-1} \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \left(\Gamma^{-1}\right)^T \underset{\sim}{z}_i^*\right) = \widehat{\underset{\sim}{\beta^*}}.$$

Showing that calculating just of these is effectively the same as calculating the other and maximises the likelihood in both ways.

The next step is to calculate the maximum likelihood of the covariance, or at least the covariance parameters. From the 3.1.7 log-likelihood we have,

$$\log \mathcal{L}\left(\boldsymbol{\theta}\right) = -n \log\left(\pi |\Gamma|^{\frac{1}{2}}\right) - \frac{1}{2} \sum_{i=1}^{n} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\beta}\right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\beta}\right),$$

$$= -n \log \pi - \frac{n}{2} \log |\Gamma| - \frac{1}{2} \sum_{i=1}^{n} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\beta}\right)^{\mathcal{H}} \Gamma^{-1} \left(\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\beta}\right).$$

Here, recall the structure of the matrix $\Gamma = \begin{pmatrix} K & J \\ J^* & K \end{pmatrix}$, we can substitute in the following,

$$|\Gamma| = K^2 - |J|^2, \qquad \Gamma^{-1} = \frac{1}{K^2 - |J|^2} \begin{pmatrix} K & -J \\ -J^* & K \end{pmatrix}.$$

Along with these we require inspecting the augmented vectors and matrices to simplify the following problem,

$$\underset{\sim}{z}_i - \underset{\sim}{x}_i \underset{\sim}{\beta} = \begin{pmatrix} z_i \\ z_i^* \end{pmatrix} - \begin{pmatrix} 1 & x_{i1} & \dots & x_{ip} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & x_{i1}^* & \dots & x_{ip}^* \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \\ \beta_0^* \\ \vdots \\ \beta_p^* \end{pmatrix},$$

$$= \begin{pmatrix} z_i - x_i^T \boldsymbol{\beta} \\ \left(z_i - x_i^T \boldsymbol{\beta}\right)^* \end{pmatrix}.$$

We can then use these finding and substitute them into log-likelihood,

$$\log \mathcal{L}\left(\boldsymbol{\theta}\right) = -n\log\pi - \frac{n}{2}\log\left(K^2 - |J|^2\right)$$

$$-\frac{1}{2}\sum_{i=1}^{n}\begin{pmatrix} z_i - x_i^T\boldsymbol{\beta} \\ \left(z_i - x_i^T\boldsymbol{\beta}\right)^* \end{pmatrix}^{\mathcal{H}}\begin{pmatrix} \frac{K}{K^2-|J|^2} & \frac{-J}{K^2-|J|^2} \\ \frac{-J^*}{K^2-|J|^2} & \frac{K}{K^2-|J|^2} \end{pmatrix}\begin{pmatrix} z_i - x_i^T\boldsymbol{\beta} \\ \left(z_i - x_i^T\boldsymbol{\beta}\right)^* \end{pmatrix},$$

$$= -n\log\pi - \frac{n}{2}\log\left(K^2 - |J|^2\right)$$

$$-\frac{1}{2}\sum_{i=1}^{n}\left(\begin{pmatrix}\left(z_i - x_i^T\boldsymbol{\beta}\right)^*\\ z_i - x_i^T\boldsymbol{\beta}\end{pmatrix}\right)^{\mathbb{T}}\begin{pmatrix}\frac{K}{K^2-|J|^2}\left(z_i - x_i^T\boldsymbol{\beta}\right) + \frac{-J}{K^2-|J|^2}\left(z_i - x_i^T\boldsymbol{\beta}\right)^*\\ \frac{-J^*}{K^2-|J|^2}\left(z_i - x_i^T\boldsymbol{\beta}\right) + \frac{K}{K^2-|J|^2}\left(z_i - x_i^T\boldsymbol{\beta}\right)^*\end{pmatrix},$$

$$= -n\log\pi - \frac{n}{2}\log\left(K^2 - |J|^2\right) - \frac{1}{2}\cdot\frac{K}{K^2-|J|^2}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)^*\left(z_i - x_i^T\boldsymbol{\beta}\right)$$

$$+\frac{1}{2}\cdot\frac{J}{K^2-|J|^2}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)^{*2} + \frac{1}{2}\cdot\frac{J^*}{K^2-|J|^2}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)^2$$

$$-\frac{1}{2}\cdot\frac{K}{K^2-|J|^2}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)\left(z_i - x_i^T\boldsymbol{\beta}\right)^*,$$

$$= -n\log\pi - \frac{n}{2}\log\left(K^2 - |J|^2\right) - \frac{K}{K^2-|J|^2}\sum_{i=1}^{n}|z_i - x_i^T\boldsymbol{\beta}|^2 \qquad (3.2.4)$$

$$-\frac{1}{2}\cdot\frac{-J}{K^2-|J|^2}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)^{*2} - \frac{1}{2}\cdot\frac{-J^*}{K^2-|J|^2}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)^2.$$

From here we derive with respect to the components of Γ^{-1} which are, $\frac{K}{K^2-|J|^2}$, $\frac{-J}{K^2-|J|^2}$ and $\frac{-J^*}{K^2-|J|^2}$[1]. Before going right into it, we differentiate the second log term with respect these three parameters,

$$\frac{\partial}{\partial\frac{K}{K^2-|J|^2}}\left(-\log\left(K^2-|J|^2\right)\right) = \frac{\partial}{\partial\frac{K}{K^2-|J|^2}}\left(\log\left(\frac{K^2-|J|^2}{(K^2-|J|^2)^2}\right)\right),$$

$$= \frac{\partial}{\partial\frac{K}{K^2-|J|^2}}\left(\log\left(\left(\frac{K}{K^2-|J|^2}\right)^2 - \frac{-J}{K^2-|J|^2}\cdot\frac{-J^*}{K^2-|J|^2}\right)\right),$$

$$= 2\frac{K}{K^2-|J|^2}\left(\left(\frac{K}{K^2-|J|^2}\right)^2 - \frac{-J}{K^2-|J|^2}\cdot\frac{-J^*}{K^2-|J|^2}\right)^{-1},$$

$$= 2\frac{K}{K^2-|J|^2}\left(\frac{K^2-|J|^2}{(K^2-|J|^2)^2}\right)^{-1},$$

$$= 2K.$$

Similarly,

---

[1]Insightful method by Fred

$$\frac{\partial}{\partial \frac{-J}{K^2-|J|^2}} \left(-\log\left(K^2 - |J|^2\right)\right) = \frac{\partial}{\partial \frac{-J}{K^2-|J|^2}} \left(\log\left(\frac{K^2 - |J|^2}{(K^2 - |J|^2)^2}\right)\right),$$

$$= \frac{\partial}{\partial \frac{-J}{K^2-|J|^2}} \left(\log\left(\left(\frac{K}{K^2 - |J|^2}\right)^2 - \frac{-J}{K^2 - |J|^2} \cdot \frac{-J^*}{K^2 - |J|^2}\right)\right),$$

$$= -\frac{-J^*}{K^2 - |J|^2} \left(\left(\frac{K}{K^2 - |J|^2}\right)^2 - \frac{-J}{K^2 - |J|^2} \cdot \frac{-J^*}{K^2 - |J|^2}\right)^{-1},$$

$$= -\frac{-J^*}{K^2 - |J|^2} \left(\frac{K^2 - |J|^2}{(K^2 - |J|^2)^2}\right)^{-1},$$

$$= J^*.$$

Lastly,

$$\frac{\partial}{\partial \frac{-J^*}{K^2-|J|^2}} \left(-\log\left(K^2 - |J|^2\right)\right) = \frac{\partial}{\partial \frac{-J^*}{K^2-|J|^2}} \left(\log\left(\frac{K^2 - |J|^2}{(K^2 - |J|^2)^2}\right)\right),$$

$$= \frac{\partial}{\partial \frac{-J^*}{K^2-|J|^2}} \left(\log\left(\left(\frac{K}{K^2 - |J|^2}\right)^2 - \frac{-J}{K^2 - |J|^2} \cdot \frac{-J^*}{K^2 - |J|^2}\right)\right),$$

$$= -\frac{-J}{K^2 - |J|^2} \left(\left(\frac{K}{K^2 - |J|^2}\right)^2 - \frac{-J}{K^2 - |J|^2} \cdot \frac{-J^*}{K^2 - |J|^2}\right)^{-1},$$

$$= -\frac{-J}{K^2 - |J|^2} \left(\frac{K^2 - |J|^2}{(K^2 - |J|^2)^2}\right)^{-1},$$

$$= J.$$

Using these results will shorten the working out involved in differentiating the whole log-likelihood, without rewriting all of 3.2.4,

$$\frac{\partial}{\partial \frac{K}{K^2-|J|^2}} \log \mathcal{L}\left(\boldsymbol{\theta}\right) = \frac{n}{2}2K - \sum_{i=1}^{n}|z_i - x_i^T\boldsymbol{\beta}|^2,$$

$$\Rightarrow \widehat{K} = \frac{1}{n}\sum_{i=1}^{n}|z_i - x_i^T\boldsymbol{\beta}|^2. \tag{3.2.5}$$

$$\frac{\partial}{\partial \frac{-J}{K^2-|J|^2}} \log \mathcal{L}\left(\boldsymbol{\theta}\right) = \frac{n}{2}J^* - \frac{1}{2}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)^{*2},$$

$$\Rightarrow \widehat{J^*} = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)^{*2}. \tag{3.2.6}$$

$$\frac{\partial}{\partial \frac{-J^*}{K^2-|J|^2}} \log \mathcal{L}\left(\boldsymbol{\theta}\right) = \frac{n}{2}J - \frac{1}{2}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)^{2},$$

$$\Rightarrow \widehat{J} = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - x_i^T\boldsymbol{\beta}\right)^{2}. \tag{3.2.7}$$

These seem pretty intuitive and putting these together should net the maximum likelihood estimator of the covariance,

$$\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\left(\underset{\sim}{z}_i - \underset{\sim}{x}_i\underset{\sim}{\boldsymbol{\beta}}\right)\left(\underset{\sim}{z}_i - \underset{\sim}{x}_i\underset{\sim}{\boldsymbol{\beta}}\right)^{\mathcal{H}}. \tag{3.2.8}$$

In practice we would like to know how significant our results are so we would like to know some distributional properties of our estimators.

3.3   Properties of Estimates

3.3.1   Distributional Properties of the Regression Co-efficients

To look at how the regression co-efficients are distributed, we first outline a property of the complex normal distribution found in [8] and [10]. Take a complex normal vector $\mathbf{Z} \sim \mathcal{CN}(\boldsymbol{\mu}, V)$ length n, a complex matrix $A \in \mathbb{C}^{m \times n}$ and a constant complex vector $\mathbf{b} \in \mathbb{C}^m$,

$$\mathbf{Z} \sim \mathcal{CN}(\boldsymbol{\mu}, V),$$
$$A\mathbf{Z} \sim \mathcal{CN}(A\boldsymbol{\mu}, AVA^{\mathcal{H}}),$$
$$A\mathbf{Z} + \mathbf{b} \sim \mathcal{CN}(A\boldsymbol{\mu} + \mathbf{b}, AVA^{\mathcal{H}}). \tag{3.3.1}$$

To find the distribution of our $\underset{\sim}{\boldsymbol{\beta}}$, we begin with our augmented $2n$ dimensional complex normal vector $\underset{\sim}{Z}$ from 3.1.2 such that $\underset{\sim}{Z} \sim \underset{\sim}{\mathcal{CN}}_{2n}\left(\underset{\sim}{X}\underset{\sim}{\boldsymbol{\beta}}, \Gamma_n\right)$ (Definition 4 [8]). Note here that this is a known covariance matrix. We then take the full data matrix $\underset{\sim}{X}$, and notice the following things are true when we take Γ_n to be in the form of $\mathbb{1}_n \otimes \Gamma$.

$$\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{Z} = \sum_{i=1}^{n} \underline{x}_i^{\mathcal{H}} \Gamma^{-1} \underline{z}_i,$$

$$\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} = \sum_{i=1}^{n} \underline{x}_i^{\mathcal{H}} \Gamma^{-1} \underline{x}_i.$$

So when we take the linear transformation of $\underline{Z}$ by a constant and due to it's properties shown in [8],

$$\left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \right) \underline{Z} \sim \underline{\mathcal{CN}} \left(\left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \right) \underline{X} \underline{\beta}, \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \right) \Gamma_n \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \right)^{\mathcal{H}} \right),$$

$$\left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{Z} \right) \sim \underline{\mathcal{CN}} \left(\left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right) \underline{\beta}, \underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right).$$

Then once more,

$$\left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right)^{-1} \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{Z} \right) \sim \underline{\mathcal{CN}} \Big(\left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right)^{-1} \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right) \underline{\beta}$$

$$, \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right)^{-1} \underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right)^{-1\mathcal{H}} \Big),$$

$$\sim \underline{\mathcal{CN}} \left(\underline{\beta}, \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right)^{-1\mathcal{H}} \right),$$

$$\left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right)^{-1} \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{Z} \right) \sim \underline{\mathcal{CN}} \left(\underline{\beta}, \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right)^{-1} \right).$$

However since,

$$\left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{X} \right)^{-1} \left(\underline{X}^{\mathcal{H}} \Gamma_n^{-1} \underline{Z} \right) = \left(\sum_{i=1}^{n} \underline{x}_i^{\mathcal{H}} \Gamma^{-1} \underline{x}_i \right)^{-1} \left(\sum_{i=1}^{n} \underline{x}_i^{\mathcal{H}} \Gamma^{-1} \underline{z}_i \right),$$

$$\left(\sum_{i=1}^{n} \underline{x}_i^{\mathcal{H}} \Gamma^{-1} \underline{x}_i \right)^{-1} \left(\sum_{i=1}^{n} \underline{x}_i^{\mathcal{H}} \Gamma^{-1} \underline{z}_i \right) \sim \underline{\mathcal{CN}} \left(\underline{\beta}, \left(\sum_{i=1}^{n} \underline{x}_i^{\mathcal{H}} \Gamma^{-1} \underline{x}_i \right)^{-1} \right).$$

And inspecting the left hand side, it is exactly the augmented estimator for the regression co-efficient, $\widehat{\underline{\beta}}$, hence it has the distribution,

$$\widehat{\underline{\beta}} \sim \underline{\mathcal{CN}} \left(\underline{\beta}, \left(\sum_{i=1}^{n} \underline{x}_i^{\mathcal{H}} \Gamma^{-1} \underline{x}_i \right)^{-1} \right). \tag{3.3.2}$$

3.3.2 Consistency in the Gamma Estimator

It is a simple and useful exercise to show the consistency of the Γ estimate $\widehat{\Gamma}$ in 3.2.8, this means that when the sample sizes are large enough, our maximum likelihood estimate $\widehat{\Gamma}$ will converge to the constant covariance matrix Γ,

$$\widehat{\Gamma} = \frac{1}{n} \sum_{i=1}^{n} \left(z_i - x_i \underset{\sim}{\boldsymbol{\beta}} \right) \left(z_i - x_i \underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}},$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(x_i \underset{\sim}{\boldsymbol{\beta}} + \underset{\sim}{\epsilon}_i - x_i \underset{\sim}{\boldsymbol{\beta}} \right) \left(x_i \underset{\sim}{\boldsymbol{\beta}} + \underset{\sim}{\epsilon}_i - x_i \underset{\sim}{\boldsymbol{\beta}} \right)^{\mathcal{H}},$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\underset{\sim}{\epsilon}_i \right) \left(\underset{\sim}{\epsilon}_i \right)^{\mathcal{H}},$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\underset{\sim}{\epsilon}_i \underset{\sim}{\epsilon}_i \right)^{\mathcal{H}}.$$

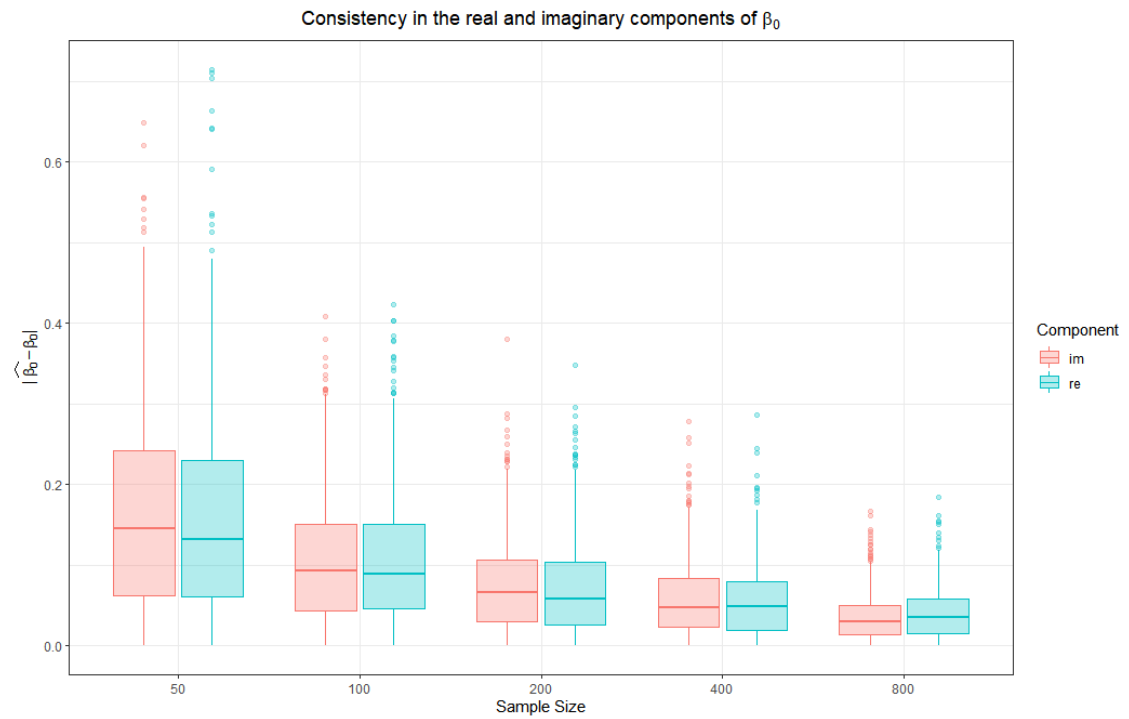From here we can employ the weak law of large numbers, turning this into the expectation seen just before 3.1.3,

$$\widehat{\Gamma} \xrightarrow{\mathcal{P}} \mathbb{E}\left[\underset{\sim}{\epsilon}_i \underset{\sim}{\epsilon}_i^{\mathcal{H}} \right] = \Gamma. \tag{3.3.3}$$

$$\underset{\sim}{\widehat{\boldsymbol{\beta}}} = \left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} x_i \right)^{-1} \left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} z_i \right),$$

$$= \left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} x_i \right)^{-1} \left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} \left(x_i \underset{\sim}{\boldsymbol{\beta}} + \epsilon_i \right) \right),$$

$$= \left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} x_i \right)^{-1} \left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} x_i \right) \underset{\sim}{\boldsymbol{\beta}} + \left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} x_i \right)^{-1} \left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} \epsilon_i \right),$$

$$= \underset{\sim}{\boldsymbol{\beta}} + \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} x_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} \epsilon_i \right).$$

Again, with law of large numbers we have,

$$= \underset{\sim}{\boldsymbol{\beta}} + \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} x_i \right)^{-1} \left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} \mathbb{E}\left[\epsilon_i \right] \right),$$

$$= \underset{\sim}{\boldsymbol{\beta}},$$

$$\underset{\sim}{\widehat{\boldsymbol{\beta}}} \xrightarrow{\mathcal{P}} \underset{\sim}{\boldsymbol{\beta}} \qquad \text{as } n \to \infty.$$

Just like before in 2.2.2, we can numerically show the fact that these MLE estimates are consistent, by showing that each of the real and imaginary components of $\beta_0, \beta 1$ and Γ converge to their true values as sample sizes increase, this time including violin plots because they are aesthetically pleasing. More on how the data is generated will be discussed in 4.1.1.

Consistency in the real and imaginary components of β_0

Consistency in the real and imaginary components of β_0

Consistency in the real and imaginary components of β_1

Consistency in each of the components of Γ

Consistency in each of the components of Γ
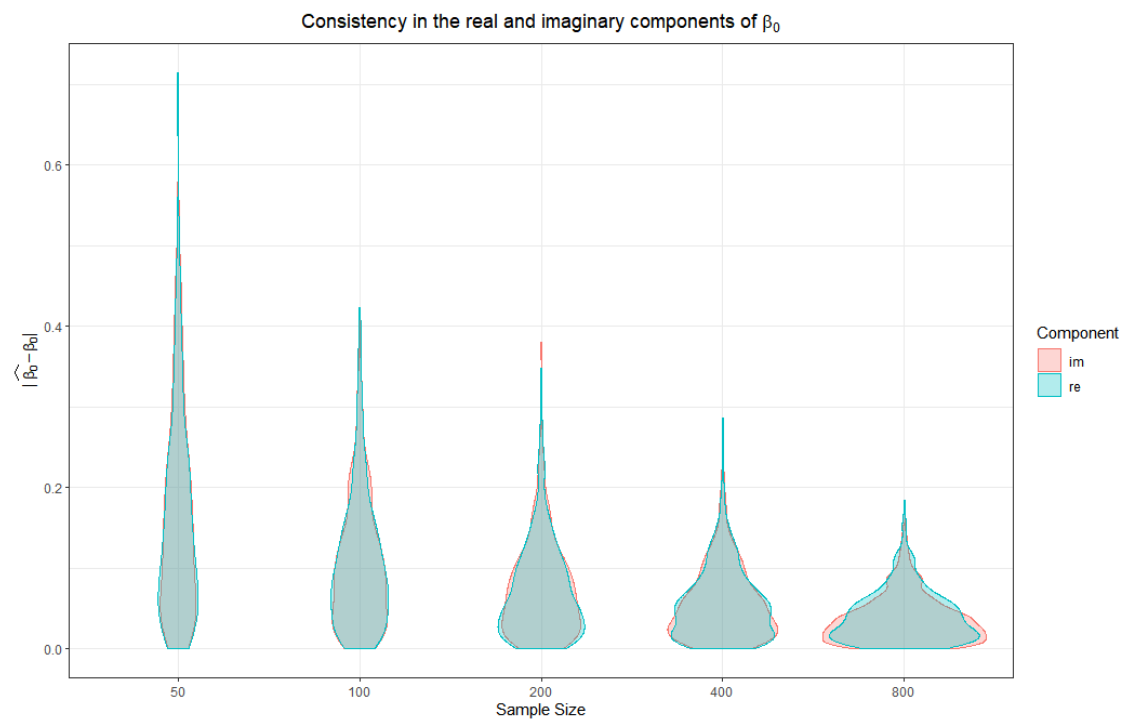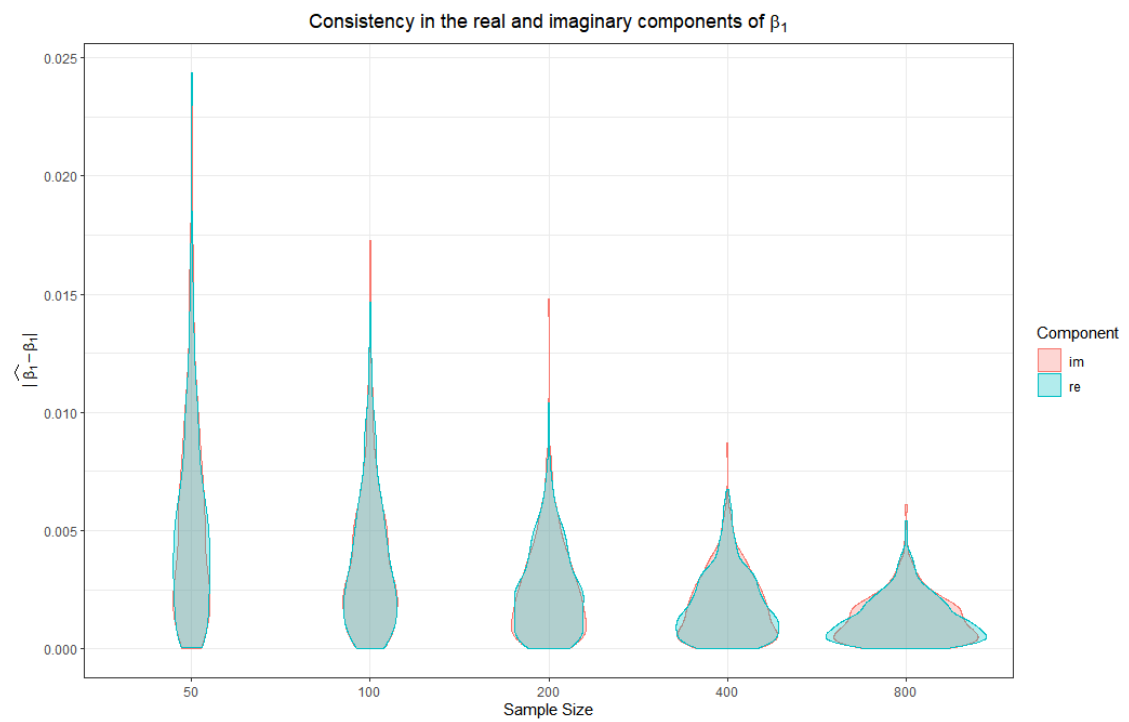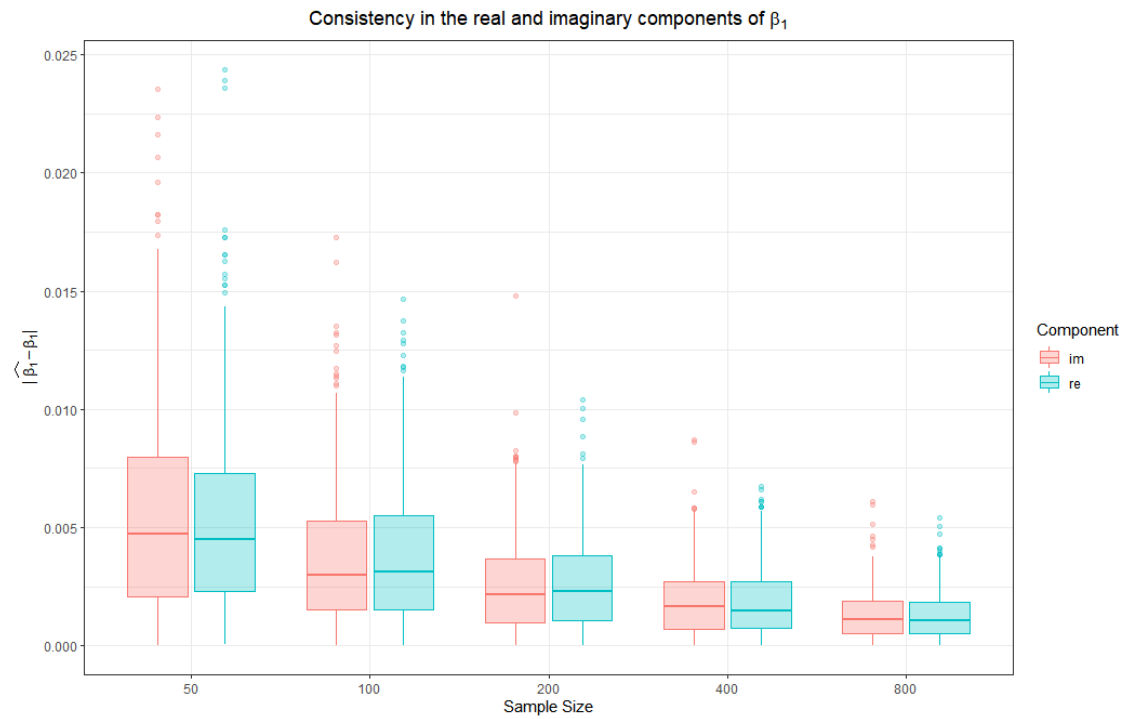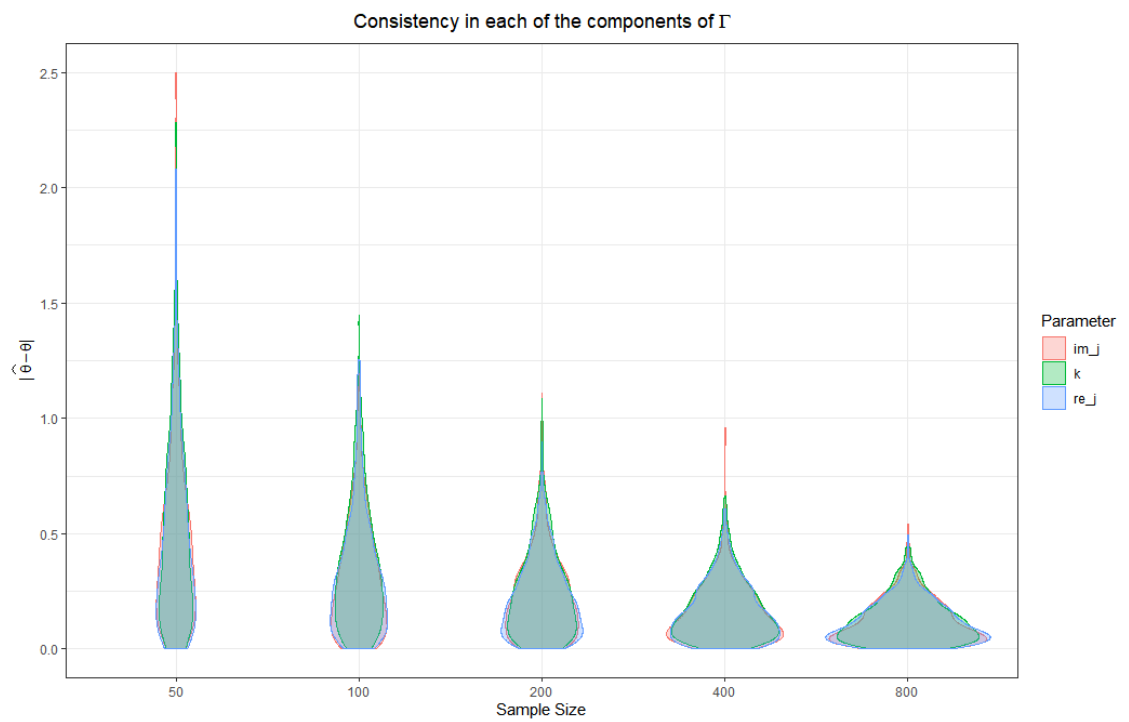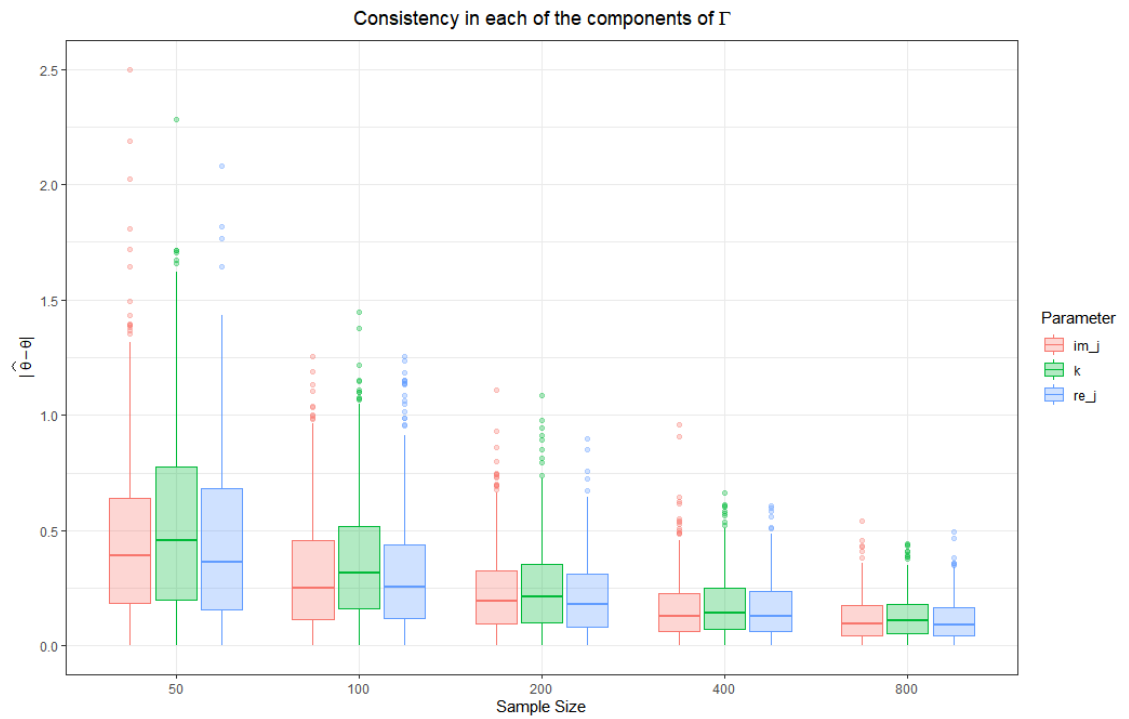
CHAPTER 4

Simulations

4.1 Validating Distributional Properties

We've shown numerically the consistency of the MLE estimators in the complex regression model in the previous chapter because it lined up better in that section as it paralleled with what was written in 2.2.2. But now we have the opportunity to show numerically the distributional properties of the regression co-efficients, so that they can be extended for statistical testing.

4.1.1 Process of Simulating Observations

To look at how effective the estimators are and to judge their accuracy compared to other models we must first must know how to simulate from the improper complex normal distribution. Suppose we take the model from 3.1.1 and take the simplest dimensions where $p = 1$, it becomes a simple complex numbered regression problem, in long form the model will be,

$$Z_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

Where $\epsilon_i \sim \mathcal{CN}(0, \Gamma)$. The overarching goal is to set some values of β_0 and β_1, randomly generate some observed data x_i and draw some $\epsilon_i \sim \mathcal{CN}(0, \Gamma)$ to obtain some Z_i. Then we use the estimators obtained earlier that require the observations Z_i and the observed data x_i to estimate a β_0 and β_1 eg $\widehat{\beta_0}$ and $\widehat{\beta_1}$, comparing this with their true values we set at the start.

Getting into more detail, after you have set some regression parameters and have created some observed data, you have to draw some errors from the improper complex normal distribution. To do this we require Definition 1 from Ducharme [8], where if $\boldsymbol{\chi} = (\chi_1, \chi_2)^T \sim \mathcal{N}_2(\mu_\chi, \Sigma_\chi)$ and Σ_χ is positive definite, then $Z = \chi_1 + j\chi_2$ is complex normally distributed. From here, we can take two approaches to formulate some complex numbered error. Recall the structure of Γ,

$$\Gamma = \begin{pmatrix} K & J \\ J^* & K \end{pmatrix}.$$

We can choose some K and J such that Γ is positive definite and that K is positive because of how it is defined. Or we could choose some parameters for Σ_χ that allow it to be positive definite and that it's diagonals are positive, these two are equivalent and can be determined by one another. For example if Γ was defined by the following, then,

$$\Gamma = \begin{pmatrix} K & \mathrm{Re}(J) + j\mathrm{Im}(J) \\ \mathrm{Re}(J) - j\mathrm{Im}(J) & K \end{pmatrix}, \qquad \Sigma_\chi = \begin{pmatrix} \frac{K+\mathrm{Re}(J)}{2} & -\frac{1}{2}\mathrm{Im}(J) \\ -\frac{1}{2}\mathrm{Im}(J) & \frac{K-\mathrm{Re}(J)}{2} \end{pmatrix}.$$

After those parameters are checked, meaning they will create a positive definite matrix, we then use the "mvtnorm" function in R to simulate a pair of observations to serve as the real and imaginary parts of the error, repeating this for however many samples are needed, then combining it all together to create a number of observations that would be distributed $Z_i \sim \mathcal{CN}(\beta_0 + \beta_1 x_i, \Gamma)$. This is done in our function "generate_eps" in "simulation_headers.R", if you want to check it out.

4.1.2 Checking the Distribution of the Regression Co-efficients

We worked out in theory earlier that the regression co-efficients $\widehat{\underset{\sim}{\boldsymbol{\beta}}}$ are distributed complex normal 3.3.2,

$$\widehat{\underset{\sim}{\boldsymbol{\beta}}} \sim \underline{\mathcal{CN}}\left(\underset{\sim}{\boldsymbol{\beta}}, \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \right)^{-1} \right).$$

Using the complex normal property 3.3.1 we discussed earlier we can rearrange this as,

$$\left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \right)^{\frac{1}{2}} \left(\widehat{\underset{\sim}{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}} \right) \sim \underline{\mathcal{CN}}\left(\mathbf{0}, \mathbb{1}_{2p} \right).$$

We can then make an additional transformation into something that is multivariate normally distributed, if we take the matrix M from earlier and give it a $2p$ by $2p$ dimensions, we can do the following transformation. Suppose we have a complex normal random vector $\mathbf{Z} \sim \mathcal{CN}(\mathbf{0}, V_p)$, that it is length p then the augmentation $\underline{\mathbf{Z}}$ has dimension $2p$ and is distributed $\underline{\mathcal{CN}}(\mathbf{0}, V_{2p})$ then recall that $\boldsymbol{\mathcal{X}} = M\underline{\mathbf{Z}} \sim \mathcal{N}_{2p}(\mathbf{0}, \Sigma_\chi)$, where the multivariate covariance Σ_χ is related to V_{2p} by $\Sigma_\chi = MV_{2p}M^{\mathcal{H}}$ from 3.1.5. So if do the following,

$$M \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \right)^{\frac{1}{2}} \left(\widehat{\underset{\sim}{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}} \right) \sim \mathcal{N}_{2p}\left(\mathbf{0}, M\mathbb{1}_{2p}M^{\mathcal{H}} \right),$$

$$\sim \mathcal{N}_{2p}(\mathbf{0}, \frac{1}{2}MM^{-1}),$$

$$\sim \mathcal{N}_{2p}(\mathbf{0}, \frac{1}{2}\mathbb{1}_{2p}),$$

$$\sqrt{2}M \left(\sum_{i=1}^{n} \underset{\sim}{x}_i^{\mathcal{H}} \Gamma^{-1} \underset{\sim}{x}_i \right)^{\frac{1}{2}} \left(\widehat{\underset{\sim}{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}} \right) \sim \mathcal{N}_{2p}(\mathbf{0}, \mathbb{1}_{2p}).$$

This is great, because the covariance of the distribution is identity, meaning that each of the terms on the left are independent standard normal. Each component on the left is a standardised real or imaginary component of beta, but the arrangement might be confusing, the first p components are the real components of $\boldsymbol{\beta}$ and the last p are the imaginary components. So we can simulate the whole thing and easily separate the components such that each is standard normal, we'll just show this working with $p = 1$ so a complex model with just intercept β_0 and β_1 and 1000 samples each with 1000 data points, using the method described earlier. It must be noted that the computation of the square root of the variance $\left(\sum_{i=1}^{n} x_i^{\mathcal{H}} \Gamma^{-1} x_i\right)^{\frac{1}{2}}$ must be done with singular value decomposition or by a Cholesky decomposition, which is quite taxing on computation, but nevertheless, here is the simulation,

Standard normal density overlaying histogram of simulated real and imaginary componenets of beta
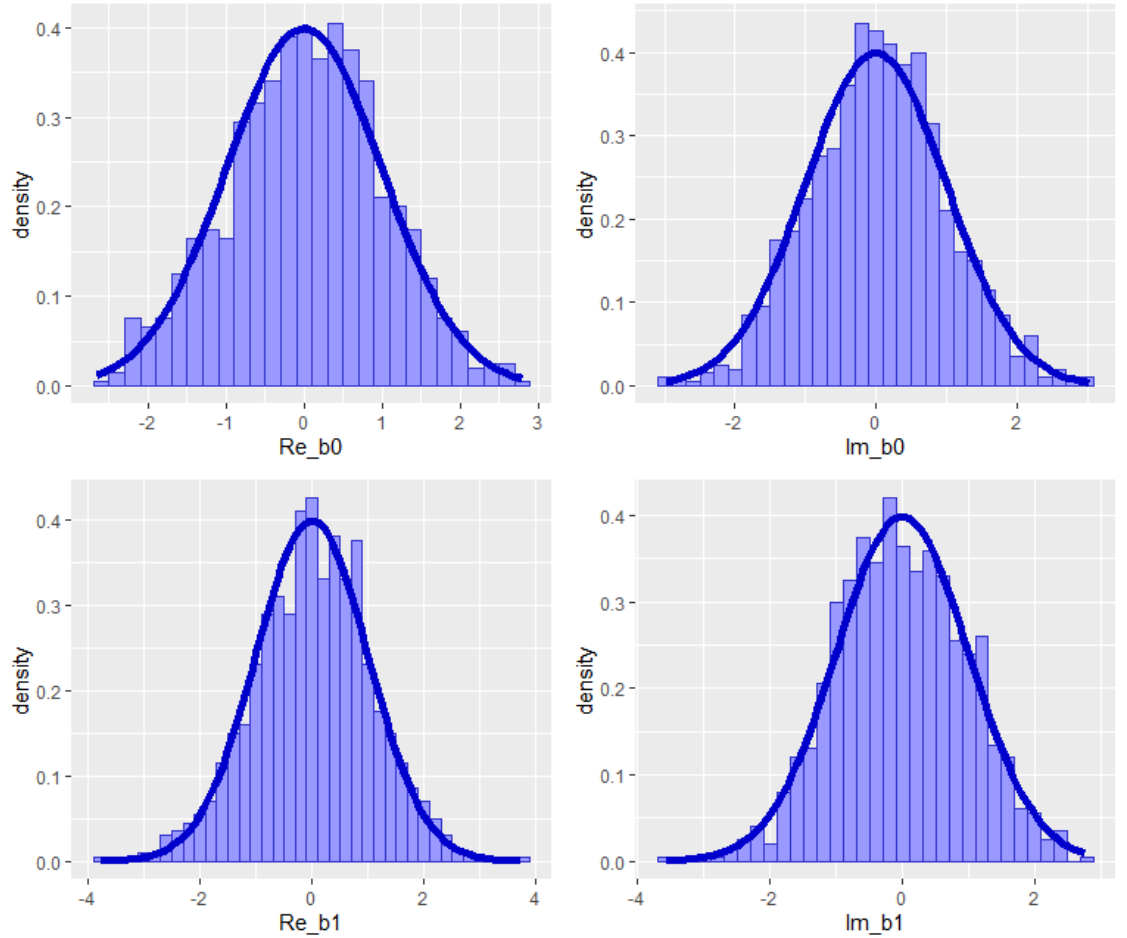
It looks pretty good, they all seem to follow the standard normal density function pretty nicely and looking at the table of means and standard deviations for each of the components, it looks to confirm our expectations,

	$\text{Re}(\beta_0)$	$\text{Im}(\beta_0)$	$\text{Re}(\beta_1)$	$\text{Im}(\beta_1)$
mean	0.02960567	0.03459596	0.01854765	-0.02643420
sd	0.9898272	0.9694767	1.0489810	0.9838040

However in practice you would not know the true gamma, you would have to estimate these. We have shown earlier that our gamma estimate converges to the constant true gamma 3.3.3, demonstrating this further, using the same random

seed, we can replace Γ with $\widehat{\Gamma}$ in the former simulation of the densities of the components of $\boldsymbol{\beta}$ and net the following results,

Standard normal density overlaying histogram of simulated real and imaginary componenets of beta
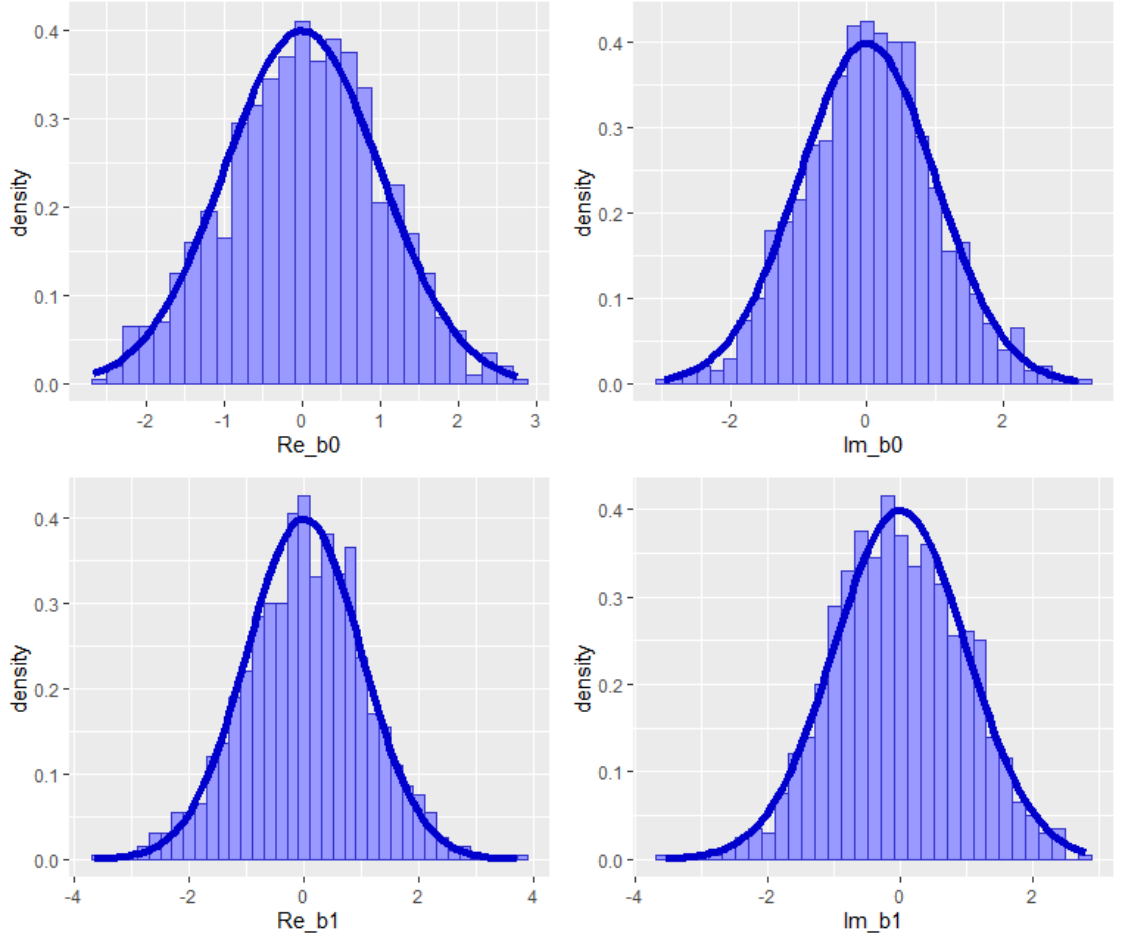
They look so alike! Which is great news as in the upcoming statistical test, we do not have to indulge ourselves in finding $\widehat{\Gamma}$'s awful distribution to test estimates of $\boldsymbol{\beta}$. Along with the graphs, we have a similar table,

	$\text{Re}(\beta_0)$	$\text{Im}(\beta_0)$	$\text{Re}(\beta_1)$	$\text{Im}(\beta_1)$
mean	0.02911989	0.03661271	0.01876298	-0.02692299
sd	0.9893894	0.9714519	1.0527006	0.9864588

When finding the distribution of $\boldsymbol{\beta}$ without the known Γ, but $\widehat{\Gamma}$ we can emplore Slutsky's Theorem, since $\widehat{\underset{\sim}{\boldsymbol{\beta}}}$ converges in distribution and $\widehat{\Gamma}$ converges in probability we can say,

$$\sqrt{2}M\left(\sum_{i=1}^{n}x_i^{\mathcal{H}}\widehat{\Gamma}^{-1}x_i\right)^{\frac{1}{2}}\left(\widehat{\underset{\sim}{\boldsymbol{\beta}}}-\underset{\sim}{\boldsymbol{\beta}}\right)\xrightarrow{d}\sqrt{2}M\left(\sum_{i=1}^{n}x_i^{\mathcal{H}}\Gamma^{-1}x_i\right)^{\frac{1}{2}}\left(\widehat{\underset{\sim}{\boldsymbol{\beta}}}-\underset{\sim}{\boldsymbol{\beta}}\right),$$

where the d denotes in distribution. So we can simply perform z tests on each of the real and imaginary components of $\boldsymbol{\beta}$.

If you note back at 3.2.3 and 3.2.5 - 3.2.8, the estimate for $\boldsymbol{\beta}$ and Γ require each other, so to obtain these estimates we must use some numerical method or solve the system of equations. However solving the equations or doing them numerically is extremely convoluted due to large inversion of the Γ matrix, so we've come up with a simple yet effective iterative method to compute these inspired by the FGLS method [11].

The method involves calculating the ordinary least squares estimator for $\boldsymbol{\beta}$, then passing it into $\widehat{\Gamma}$, then passing that into the estimator for $\widehat{\underset{\sim}{\boldsymbol{\beta}}}$ so on and so forth until it converges, depending on the tolerance you set, this converges very quickly, eg at a tolerance 0.0000001 this generally happens in 3 steps. Because each step requires taking the inverse of a $2n$ by $2n$ block diagonal matrix with 2 by 2 Γ blocks, this does become very costly in computation.

4.2 Comparisons

We must check how good our estimator is compared to other linear regression methods on complex numbers, to establish some benefit and then explain why the benefit exists. The two methods we will compare the complex linear regression model with are the method of taking the modulus of the data and running a linear model on that and on a multiple linear regression where the real and imaginary parts are stacked, then the data is regressed on it, however with no specifications on the covariance structure.

Random complex data is going to be generated 1000 times for differing sample sizes and then the three methods will estimate a value for the regression co-efficient $\boldsymbol{\beta}$ and their estimates will be compared to the randomly chosen true value.

4.2.1 Individual Results

So in the linear model that regresses the modulus of the data, the comparison to the true beta is done by taking the absolute value of the true beta and taking the difference, giving us the following results on the estimates for β_0 and β_1.
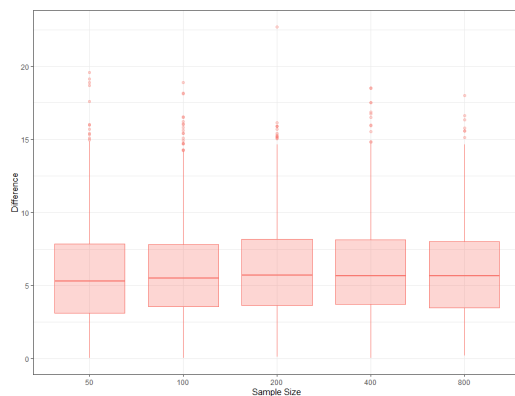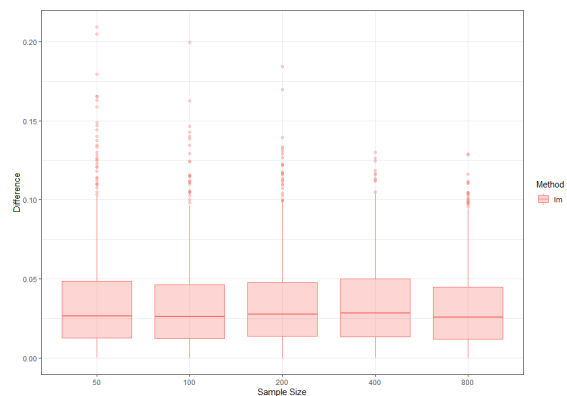

Figure 4.1: Modulus method with β_0 Figure 4.2: Modulus method with β_1

We see that taking the modulus does not appear to give accurate estimates of $\boldsymbol{\beta}$ when it dealing with complex data generated from the complex normal density.

The β_1 estimate appears at least close to the true value but neither estimates demonstrate any form of consistency as sample sizes increase.

We move on the to multiple linear regression with no specified covariance structure, using the base function "lm" in R and regressing the real and imaginary parts of the covariates onto the observations, where the observations are stack real on-top of imaginary.
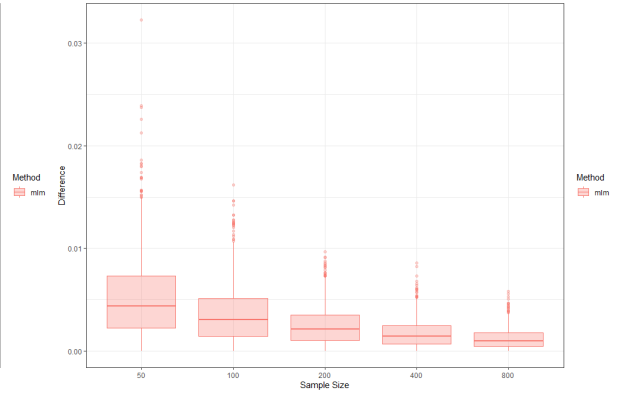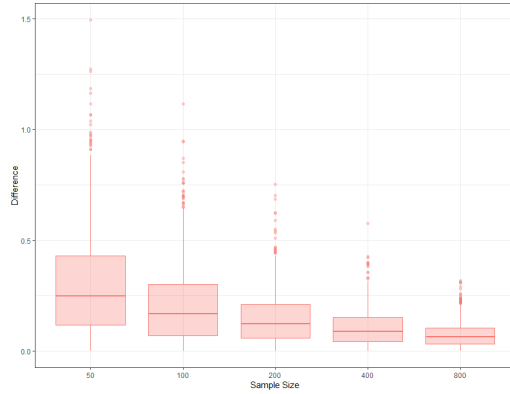

Figure 4.3: Multivariate method with β_0 Figure 4.4: Multivariate method with β_1

This is more of what we want to see, closeness to the true beta and more accuracy as n increases, similarly we will look at the complex linear regression model.
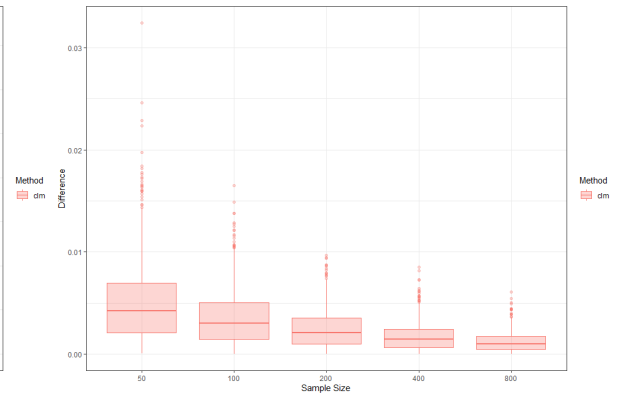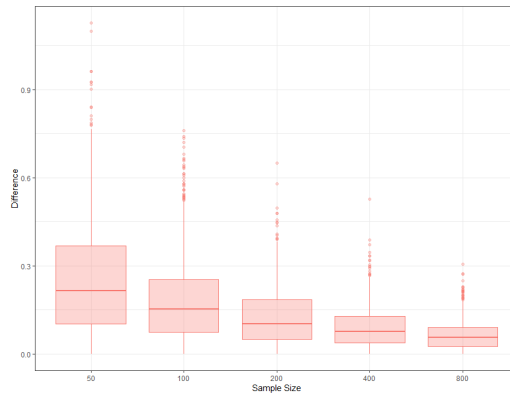

Figure 4.5: Complex method with β_0 Figure 4.6: Complex method with β_1

This seems quite similar in property to the multivariate linear model, so we'll check their closeness in the same graph. We will ignore the results of the modulus method, because they were clearly too far from the true values.
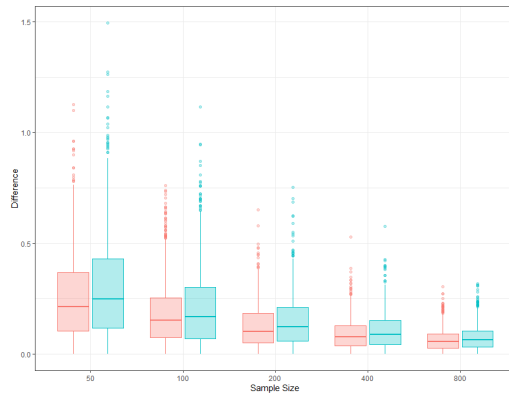
4.2.2 Side by Side Comparisons
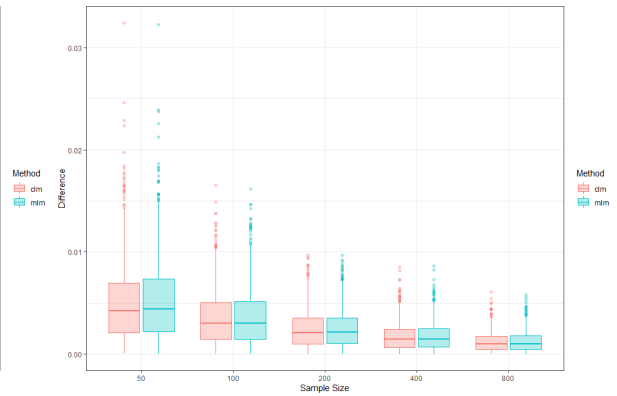

Figure 4.7: Comparison for β_0

Figure 4.8: Comparison for β_1

It is a little hard to see the difference, but complex regression method does better, but not by much especially when considering the β_1 estimate, the complex method is coloured in red (on the left) and the multivariate one in green (on the right).

These differences between the two methods really comes down to the specification of covariance structures, because the complex regression model is able to organically describe the variances and covariances of complex normal error by being a specific case of the bi-variate normal distribution and we have taken no specific structure to the covariance structure of the multivariate linear regression method, so it is out performed by the complex one. To get to the same level of accuracy as the complex regression model, a bi-variate covariance structure just be symmetric with different values on the diagonal, none like this was found in the "gls" function from the "nlme" package [1].

---

[1]https://stat.ethz.ch/R-manual/R-devel/library/nlme/html/corClasses.html

CHAPTER 5

Application - Wind

We test how effective the method and implementation of complex numbered regression is by applying it to a relatively simple scenario. Here we have wind data, that is a single point in time having wind speed and direction, acting as the modulus and argument of a complex data point. Using past winds (from a day ago) as covariates and current wind as observations. Later on we also introduce time of day and time of year as complex covariates, where the time of day can be measured as a fraction of the total amount of time passed, becoming the argument of a complex number with modulus 1. This makes sense if you imagine a 24 hour clock 00:00 should be right next to 23:59, we do a similar process for time of year.

5.1 Introduction to the Problem

We seek to create value for people who want to model wind for predictive purposes by providing complex regression co-efficients that could have beneficial interpretation and an accurate model for linear regression that can directly intake complex numbered data. The source of the data was introduced earlier, however not the process of acquisition, it was requested from the National Oceanic and Atmosphere Administration and was data from a weather station in Kansas near a town called Coffeyville, the raw data is clean and made usable with perl, all scripts are contained in the github link in the Appendix.

5.1.1 Data Cleaning

The data of course isn't in a perfect state for number crunching immediately most of the work is done by the perl script, "wind_script.pl", extracting the important information from the raw data. Then to create usable complex data, true bearing must be changed into argument, done by the function "true_arg" in "simulaton_headers.R" and the time covariates are created by passing them as epoch time minus the start of the year. As well as manipulating the data frames to have wind from one day ago or from 30 minutes ago.

5.2 Results

So in a simple case when we regress the wind from the day before hand with the current wind (file "18_simulate.R"), taking a sample of 500, we can look at the output of our summary function and at the regression co-efficients.

```
"        Summary output for Complex Linear Model        "
"-----------------------------------------------------------"
"        Complex Regression Co-efficient Estimates        "
"-----------------------------------------------------------"
"Parameter                       Estimate"
"Beta0         -0.811892239241352+0.2866818843721189i"
"Beta1         0.251925003288227-0.019890386676097i"
"-----------------------------------------------------------"
"Complex Regression Co-efficient Estimates in Polar Form"
"-----------------------------------------------------------"
"Parameter         Modulus                 Argument"
"Beta0         0.861020040979027         2.80215582051333"
"Beta1         0.252708992249777         -0.0787901568714544"
```

Figure 5.1: Summary output for the regressing day before wind on current wind

The output gives two ways to view the regression co-efficients giving more angles to view the problem. These regression co-efficients can be viewed as a sort of intercept and gradient like in the real case, but the interpretation requires more processing. For example "Beta0" can be viewed as the current expected wind given that there was no wind the day beforehand, however suppose the wind the day before hand was 10 kilometres an hour northerly, then the current wind would be the vector "Beta0" plus the vector "Beta1" times $10j$, so the impact of the wind the day before on today's wind is roughly a quarter of the speed it used to be and rotated slightly clockwise.

We also give in the summary output, a printout of the p values of each of the β and their real and imaginary components,

```
"-----------------------------------------------------------"
"        Z-Scores of each component of beta        "
"-----------------------------------------------------------"
"Component                       Score"
"Real Beta0         170.723570368807"
"Imaginary Beta0    83.2831633769332"
"Real Beta1         528.252767176408"
"Imaginary Beta1    65.1608447953815"
"-----------------------------------------------------------"
"        p-values of each component of beta        "
"-----------------------------------------------------------"
"Component                       p-value"
"Real Beta0         0"
"Imaginary Beta0    0"
"Real Beta1         0"
"Imaginary Beta1    0"
"-----------------------------------------------------------"
"                 p-values of each beta                 "
"-----------------------------------------------------------"
"Beta                 p-value"
"Beta0         0"
"Beta1         0"
```

Figure 5.2: P-values for the regressing day before wind on current wind

We see because the scores are so large the p-values are going to be 0, due to the limitations in the software, which just means that the wind the day before is significant in forecasting the wind currently.

We also introduce the time covariates and see their impact on the model. Practically this can be used to describe winds changing depending on both the time of day and time of year, which in practice does happen, due to winds being changes in pressure and pressure being built from the heating and cooling of the air. So very generally this could help predict where winds would be generally coming from, for a fixed location winds could form a pattern over time due to factors like unique geography, distance from mountain ranges and the sea.

```
"---------------------------------------------------------"
"        Complex Regression Co-efficient Estimates        "
"---------------------------------------------------------"
"Parameter                  Estimate"
"Beta0         -0.803317402519914+0.266411568178035i"
"Beta1         0.247545749423208-0.0173898897053358i"
"Beta2         0.433953767318267-0.2538207733945991i"
"Beta3         0.242421955681137+0.1656768514275578i"
"---------------------------------------------------------"
"Complex Regression Co-efficient Estimates in Polar Form"
"---------------------------------------------------------"
"Parameter        Modulus                Argument"
"Beta0         0.846341522584365       2.82136756253111"
"Beta1         0.248155810573644      -0.0701339775560138"
"Beta2         0.502733365861664      -0.529244581539847"
"Beta3         0.29362769572236        0.599513919771605"
```

Figure 5.3: P-values for the regressing day before wind on current wind

In this model, "Beta0" and "Beta1" are the same regression co-efficients as before, however the interpretation of "Beta2" is the for the time of day and "Beta3" for the time of year, because the two time covariates always have modulus of 1, the interpretation of the model with no previous wind the day before changes. Now it is given a specific time of day and year the expected wind becomes the sum of three vectors, for example for the simplest case, take midnight January 1st, the expected current wind, given no wind the day before becomes the sum of the vectors "Beta0", "Beta2" and "Beta3".

This could present itself as a different way to run autonomous wind turbines, perhaps where real time methods don't work, the average wind speed and direction could be calculated with covariates like the ones presented in the above model.

CHAPTER 6

Conclusion

6.1 Summary

The largest take away from this, is the complex regression model being a transformation of the bivariate normal. So the results could potentially be achieved through specific and careful implementation of a bivariate normal model (discussed at the end of Chapter 4). However, we are providing a statistical tool that doesn't require the specific formulation of covariance structure, but naturally accounts for the complex normal noise. In beating out commonly used linear regression tools for complex data, the tool exists as a better method without the difficulty in creating a specific covariance structure.

There is the intrinsic value of finding out for completeness that the extension into complex numbers, gives very similar results to the estimators of generalised least squares, however also as the platform to further research, more complicated regression methods and the like. This complex linear regression method also provides a different interpretation into it's regression co-efficients, hence how complex data is related to one another.

6.2 Discussion

There were short comings in this investigation, some due to difficulty, such as the distribution of $\widehat{\Gamma}$ and others due to time, being unable to investigate the convergence of the iterative method 4.1.3 to calculate the regression parameters. That particular problem also is an avenue to find an exact solution to calculating the maximum likelihood estimates, it seems as though the number of variables and the number of equations match and a non-linear system of equations can be solved.

From the very start, a few different regression models could be used, these would potentially lead to different results, one such my co-supervisor had began work on and has suggested that his way has a nicer interpretation of the regression co-efficients. An alternative model was also suggested by my main supervisor, one that look directly at the modulus and argument by splitting the covariate and regression co-efficient into a polar form eg $|\beta_1||x_1|\exp(j\beta_2 x_1)$.

References

[1] Lai, S., Glover, G.H. Detection of BOLD fMRI signals using complex data, *Proceedings of the ISMRM* (1997), 1671

[2] Rowe, D.B. and Logan, B.R, A complex way to compute fMRI activation, *NeuroImage* **23** (2004),1078–1092

[3] Picinobo, B, Second-Order Complex Random Vectors and Normal Distributions., *IEEE Transactions on Signal Processing, Institute of Electrical and Electronics Engineers* **44** (1996), 2637–2640

[4] Adrián Barquero, Qiaochu Yuan and Math Stack Exchange community, Interesting results easily achieved using complex numbers, *Mathematics Stack Exchange* (2010)

[5] García-Martín, J., Gómez-Gil, J., Vázquez-Sánchez, E., Non-Destructive Techniques Based on Eddy Current Testing, *Sensors* **11** (2011), 2525–2565

[6] Yan, C. J., Zhang, S. G. and Ning, W., Estimations of the improper linear regression models with complex-valued data, *Journal of Graduate University of China Academy of Sciences* **29(2)** (2012), 145–153

[7] Mandic, D.P., Javidi, S., Goh, S.L., Kuh, A., Aihara, K., Complex-valued prediction of wind profile using augmented complex statistics, *Renewable Energy* **34** (2009), 196–201

[8] Ducharme, G.R., Lafaye de Micheaux, P. and Marchina, B., The complex multinormal distribution, quadratic forms in complex random vectors and an omnibus goodness-of-fit test for the complex normal distribution, *Ann Inst Stat Math* **68** (2016), 77–104

[9] Fischer, R. F. H., Precoding and Signal Shaping for Digital Transmission: Appendix A: Wirtinger Calculus, *Wiley Online Library* **A.3.1** 411, `https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471439002.app1`

[10] Wikipedia Contributors, Complex normal distribution — Wikipedia, The Free Encyclopedia, *Wikipedia* (2019), `https://en.wikipedia.org/w/index.php?title=Complex_normal_distribution&oldid=907253384`, [Online; accessed 13-November-2019]

[11] Wikipedia Contributors, Feasible generalized least squares — Wikipedia, The Free Encyclopedia, *Wikipedia* (2019), `https://en.wikipedia.org/wiki/Generalized_least_squares#Feasible_generalized_least_squares`, [Online; accessed 13-November-2019]

`https://github.com/jjlt/thesis_files`

A.1 File Descriptions

This file is also found in the github link, it contains a description of the contents of the repository.

```
# thesis_files
All files relevant to my Honours Thesis.
This README.md is the documentation to GitHub repository
######################################################################
#       Name            #                Description               #
######################################################################

01_simulate.R # For rudamentory testing of complex normal
# generation

02_simulate.R # Testing for cases where there is no
# psuedo-covariance

03_simulate.R # Testing covariate matrices size p > 1 work

04_simulate.R # Inspecting the results of the iterative estimation
# method

05_simulate.R # Initial comparison with testing taking the modulus
# of complex data

06_simulate.R # First important file. Simulating complex normal
# and checking for normality of the beta estimates.
# This file produced the graphics for the density
# histograms. Warning that large values of n, will
# take ages.

07_simulate.R # Another important file where graphics were made,
# this one created the numerical testing of the
# components of the regression parameters having
# the consistency, eg they converge to the true
# values.
```

08_simulate.R # Testing file for statistical tests.
Non-substaintial file.

09_simulate.R # Looking at the covariance of the beta estimate,
using true values of gamma and estimated values,
seeing if they line up with the theoretical amount.

10_simulate.R # Used to produce the graphics for showing the
the consistency of the real regression parameters
numerically.

11_simulate.R # Non-substaintial file inspecting the convergence
# of gamma. Turns out it does converge, just not as
fast.

12_simulate.R # Attempting to do statistical tests and transforming
the beta estimates. Seeing if they results looked
correct. Tried making insignificant data, does work.
p-values can be higher when there isn't a linear
relationship.

13_simulate.R # Trying out the wind data for the first time. 30
minute beforehand wind on current wind.

14_simulate.R # We investigate if we can actually non-linearly solve
the MLE, but it looks way too disgusting, because it
has to be split into real and imaginary parts as the
function cannot deal with them.

15_simulate.R # Interesting file. Comparisons with other linear
regression models on their predictive power on
wind data. lm on the modulus works pretty well,
actually better than clm, mlm and gls. I think it is
because the wind data is very simple and wind from
thirty minutes ago is always going to be pretty close.
gls worked okay, it was implemented in two ways, but
it surpisingly wasn't close to clm or mlm. This would
be the last moment gls was relevant.

16_simulate.R # We introduce the new time covariates and to tests
with clm.

17_simulate.R # Important file. We do comparisons with the other
linear regression methods, similar to 15_simulate,
but this time on simulated data and we test the
closeness of regression co-efficient beta. clm wins

```
# and gls is no where near, no worth putting into the
# graphics to be honest. Be warned when running this
# use smaller sample sizes for speed and compare the
# gls for yourself.

18_simulate.R # Changing the previous winds to 24 hours beforehand
# rather than 30 minutes beforehand. Makes it a more
# interesting story.

1912794.csv # Raw data from NOAA, had to do a lot of work finding
# access to data, and finding access to data in a csv
# form so I can easy clean it.

new_data.csv # Cleaned data without time variables

new_data_1.csv # Cleaned data with time.

simulation_headers.R# Important file. This file is used as a header
# containing all the useful functions including the
# whole clm function. Inside the code is very well
# documented, except for the last few functions,
# which are obvious in function. Careful the clm and
# fgls functions are sensitive to the correct imput.

wind_data.csv # Cleaned and rearranged data.

wind_data2.csv # See new_data_1.csv

wind_script.pl # The perl script which cleans the raw data.

/06_simulate/ # Contains graphics and R workspaces, so the very long
# processes don't have to be repeated. Works spaces with
# error in their name contained data generated with
# a bug, so don't use those.

/07_simulate/ # See above, has a a section of older graphics.
# again use working R workspace

/08_simulate/ # Just a workspace

/09_simulate/ # Just a workspace

/10_simulate/ # Graphics

/17_simulate/ # Workspace

/graphics/ # The final set of graphics used to compile the
```

thesis and presentation