ALMA MATER STUDIORUM - UNIVERSITY OF BOLOGNA

Department of Statistical Sciences "Paolo Fortunati"

Bachelor in Statistical Sciences ~ Curriculum Stats&Maths

Stationarity and Outlier Detection for Multivariate Time Series: A study of Brain Scans

Academic Year
2021/22

Supervisor
Christian Martin Hennig

Student
Ernesto Bocini

Acknowledgments

I would like to express my deep gratitude to Professor Christian Martin Hennig, my thesis supervisor, his attention to detail and his valuable critiques have brought out the best in me. Although I initially set myself too high and unrealistic goals, this project allowed me to apply many of the concepts studied during the three years and also learn some new ones, and I am extremely satisfied with the work done together.

I would also like to thank Professor Alessandra Luati for bringing forth helpful suggestions and for providing the data sets analyzed during the development of this paper.

My grateful thanks are also extended to my parents, my brothers and my sister, for their love and encouragement throughout my studies, especially during those periods of fanatical study. Without them, I would never have enjoyed so many opportunities.

Finally, I wish to thank my friends who supported me in each and every part of this venture and enriched these years with conversations and experiences. I wouldn't be who I am today if it wasn't for those mistakes and adventures.

I'd also like to thank the University of Bologna, for giving me the possibility to forward my career in the right direction. Last year of high school I didn't know which path to take after the diploma yet, but when I knew about the *STATS & MATHS* curriculum I immediately realized where I belonged. Therefore I really thank all the people that took part in this amazing journey. Especially, I wish to thank Francesco, Gianluca, Margherita and Lorenzo, which between one exam and another have become much more than simple study partners.

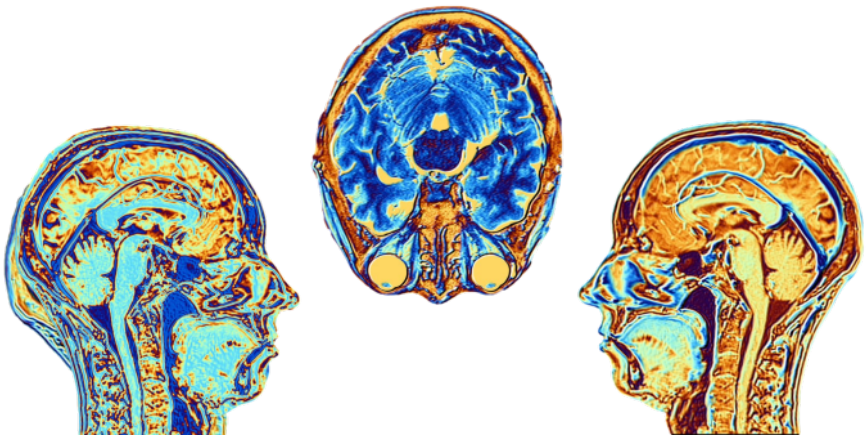
Contents

Abstract - Ever since I chose the STATS&MATHS curriculum, what I have loved the most about statistics is its fusion of solid mathematics with typical everyday situations. On top of that, advances in technology have enabled a variety of data management mechanisms. Mainly, what I found most interesting, is the time data analysis. Business, economic, engineering, and environmental data are often collected in roughly equally spaced time intervals, for example, hour, week, month, or quarter. In many problems (such as the one we are dealing with in this project), such time series data may be available on several related variables of interest, and multivariate statistical methods are needed to accurately diagnose them. This, along with my regular endeavor to undertake challenges of multi-disciplinary data sets, brought me to the stimulating health issues framework related to the field of neuroscience. This paper deals with a higher dimensional array and discusses two different methods for comparing time series data. The first method involves stationarity analysis, while the second involves outlier detection in the multivariate time series domain.

Keywords and phrases : Stationarity analysis , outlier detection, multivariate time series.

---

1 Introduction



1.1 Project overview and background

Our state of consciousness changes significantly during stages of sleep and rest. The general belief is that brain activity declines when we are resting. That is why, for many years, brain studies focused on examining changes in the brain that occur when people are engaged in an activity. Much less interest was shown towards studying what happens to the brain while people are doing very little. However, more recently,

scientists have decided to investigate brain activity during sleep using magnetic resonance imaging, or MRI, a medical imaging technique used in radiology to form pictures of the anatomy and the physiological processes of the body (from Wikipedia: MRI). These images provide important information in addition to EEGs. This project aims to study whether subjects with different traits - e.g., abilities and neuropsychiatric diseases - have different behaviors in their blood oxygenation level dependent (BOLD) at rest. This kind of study has been possible thanks to modern functional magnetic resonance imaging (fMRI) technologies. *BOLD - fMRI* is a non-invasive technique that has become increasingly popular in neurosciences. It measures the proportion of oxygenated hemoglobin in specific areas of the brain, mirroring blood flow (Wikipedia: BOLD-fMRI). Another critical point that we try to exploit here is the detection of spontaneous activations of the brain during resting state, and again how this may depend on the mentioned variables.

There are several benefits from this kind of study. Theoretical and experimental results suggest that brain activity measured during specific tasks may be described in terms of generic properties of brain activity at rest. Therefore understanding how those latter can be influenced by age and types of diagnosis is fascinating.

1.2 Project Objectives

Several objectives were set to enable this project to achieve its aim.

The initial objective relates to the preparation of the data so that it can be used to proceed with more advanced analysis. This will ensure a correct scanning of missing quantities and an understanding of which data sets we really want to analyze. In the case of this project, this means ensuring that the missing values can be removed without major hazards and that we have a clean mapping between the BOLD time series and the subjects traits. This is vitally important; otherwise the analysis couldn't proceed.

The second objective will be to define a statistic that is able to capture differences among subjects based on their BOLD time series, and that can be somehow explained by the subjects' traits. In this project, we will first try a straightforward approach based on the stationarity of the time series describing each subject. Then we will focus on outlier detection. The first step to achieve this will be to research and choose the outlier detection algorithm that most suits our data. Then we translate the algorithm into the software language, and we check whether it produces an effective distinction among subjects.

The just mentioned outlier detection procedure introduces two further objectives. Dimensionality reduction for multivariate time series and identification + fitting of a multivariate time series.

Unfortunately, to build a significant analysis, we would need many more subjects than the ones at hand. For this reason, the reader should take the paper as a walk-through of the methods to carry out to examine and compare different subjects BOLD series and only see the final results as an example of application.

1.3 Structure of the Report

The structure of this document will match the progression of the project itself. We start from data exploration and cleaning in section §2. Then we present the methods

and the results of the analysis of the data based on the stationarity investigation in section §3. Section §4 will contain the core of the project, namely the dimensionality reduction for multivariate time series, the model identification and fitting, and finally, the outlier detection procedure from a data science perspective and the results obtained when applied to the cleaned data. Some final conclusions and thoughts are contained in the very last section, §5.

2 Datasets

The multimodal imaging dataset comes from a pilot study of the Enhanced Nathan Kline Institute-Rockland Sample project (NKI1). The National Institute of Mental Health's strategic plan for advancing psychiatric neuroscience calls for increased discovery science initiatives to delineate developmental trajectories for risk and resilience across the lifespan. A detailed description of the project, scopes, and technical aspects can be found on the Enhanced Nathan Kline Institute website [1].

The pilot NKI1 study comprises multimodal imaging data and subject-specific covariates for n = 24 subjects. Detailed information can be found on the dedicated page of Nathan Kline Institute [2]. These data include information on Structural Networks, Functional Networks, and Dynamic Functional activity. The first two, measuring respectively the anatomical interconnections among brain regions (made by white matter fibers) and the synchronization in brain activity for each pair of brain regions, are not of interest in this project. Our focus will be on the Dynamic functional activity, which contains measures of changes in the blood-oxygen-level-dependent (BOLD) signal during resting-state fMRI sessions—more details in the following subsection.

In order to answer the query proposed in the introduction, we also need some subject-specific information. In fact, for each subject in the study, data on age, handedness, and psychological traits are available in an additional dataset.

These two datasets are further described in the following subsections.

2.1 Dynamical functional activity

Dynamic functional activity — monitored for each brain region — is obtained from resting-state fMRI (R-fMRI). This imaging technology monitors brain activity in different regions via dynamic changes in blood flow, creating a low-frequency blood-oxygen-level-dependent (BOLD) signal – while the subject is not performing an explicit task during the imaging session. In the present NKI1 study, the subjects are asked to stay awake with their eyes open. See Lee et al. (2013) [3] for additional details about R-fMRI.

The raw R-fMRI scans are pre-processed to derive the time-series data for each brain region. In this case, the C-PAC software has been used. For citation and details on the time-series extraction, refer to the documentation at Child Mind Institute (2017) [4] and [5], respectively.

Dynamic functional activity — based on the Dynamic functional activity information — based on the Desikan atlas parcellation — is contained in an array Y of dimensions 70×404×24×2 comprising the 70×404 multivariate time series activity data collected for the 24 subjects in each of the two scan-rescan imaging sessions. In particular, if we take the third dimension fixed at i and the fourth dimension fixed at k, $Y_{.,.,i,k}$ is a 70 × 404 matrix whose rows contain the dynamic activity data of the brain regions, collected at T = 404 equally spaced times, for subject i, monitored during scan k, for every i = 1, ... , 24 and k = 1, 2; see e.g., Figure 1.

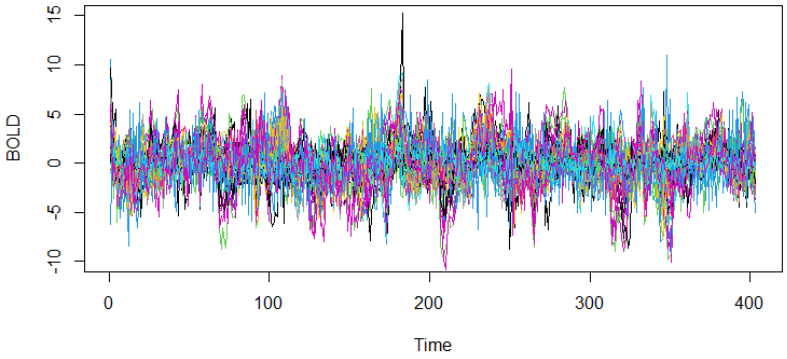


Figure 1: Dynamic activity time series of each brain region for the second subject.

Based on this description, the element $Y_{v,t,i,k} \in \Re$ is a real number measuring the activity of brain region v at time t, for subject i, during scan k. According to the data collection protocol, the time lag between consecutive time observations is 1400 ms. **Dynamic functional activity** data for two subjects are not available, whereas, for other 10 subjects, only the first scan is observed. Therefore, their corresponding entries in Y are empty.

2.2 Information on subjects

For each subject in the study, we know her/his ID - to match the statistical units to their corresponding imaging data - her/his age at the first scan; whether she/he is left-handed, right-handed, or ambidextrous; and if the subject has been diagnosed a current and/or lifetime mental disorder. If present, the type of disorders are also described and separated by '|' . Refer to Table 1.

The description is based on the International Classification of Diseases. Refer to this link for more information [6].

These data clearly motivate several interesting directions of research. For example, it is of relevant interest to develop flexible models which are able to infer global and local group differences in brain activity and connectivity across subjects with different traits - e.g., presence or absence of mental disease - or other subjects' traits. Finally, an appealing aspect of this dataset is that it provides replicated and multi-domain data on both brain functional activity. Hence, time series models investigating how different traits facilitate synchronization in dynamic brain activity, could have an important impact on the statistical and neuroscience literature. The only complication is the small sample size, which will represent a stumbling block in further analysis and eventual regressions.

2.3 Study of missing values

One thing that should be done very early on is to decide on the treatment of missing values. In order to do this, we need to find out how many they are and how they are

	age	hand	c_d	l_d	type_c_d	type_l_d
1	57	R	NO	NO	-	-
2	52	R			-	-
3	32	R			-	-
4	36	R			-	-
5	22	R			-	-
6	27	R	NO	NO	-	-
7	60	R	NO	YES	-	Alcohol Abuse | Cannabis Abuse
8	21	R	YES	YES	Major Depressive Disorder Single Episode Unspecified | Cannabis Abuse	Eating Disorder NOS
9	21	L	NO	NO	-	-
10	30	R	NO	NO	-	-
11	27	A	NO	NO	-	-
12	48	A	NO	NO	-	-
13	22	R	NO	YES	-	Major Depressive Disorder Single Episode Full Remission
14	19	R	NO	NO	-	-
15	57	R	NO	NO	-	-
16	25	R	NO	NO	-	-
17	38	R	YES	YES	Major Depressive Disorder Recurrent In partial remission	Alcohol Abuse
18	46	R	NO	NO	-	-
19	22	R	NO	YES	-	Alcohol Abuse | Cannabis Dependence | Depressive Disorder NOS
20	32	L	YES	YES	Generalized Anxiety	Alcohol Dependence unspecified
21	22	R	NO	NO	-	-
22	42	R	YES	YES	Major Depressive Disorder Recurrent Mild | Social Phobia	Alcohol Abuse | Cannabis Abuse | Cocaine Dependence | Opioid Dependence | ADHD Combined Type
23	31	L	NO	NO	-	-
24	36	R	NO	NO	-	-

Table 1: Information collected for subjects under study. NOS: not otherwise specified; ADHD: attention-deficit/hyperactivity disorder.

distributed over the individuals and over the 70 different regions. If they concentrate somewhere, then maybe it is better to a take region out or to take a person out. If the number is small and the position is not problematic, some rather simple techniques could be implemented, such as simple imputation or removing the missing unit itself.

The following table (Table 2) reports the absolute and relative frequencies for missing quantities for both the scan and the re-scan, subject-wise. As anticipated in the previous section, looking at the table, we can observe two major facts:

- Dynamic functional activity data for subjects 1 and 21 are not available.
- For other 10 subjects, only the first scan is observed. Therefore, their corresponding entries in the rescan are empty.

In addition, another fact is claiming for attention: subjects 2, 4, 5, 14, and 20 have 70 missing values, which corresponds exactly to the number of regions under study. It turns out that each of these subjects' regions has the last observation (at time index 404) missing. This simplifies the analysis: being the last observation of a time series, which by definition is ordered in time, means that removing the unit will only have the effect of slightly narrowing our time frame length without major damage to modeling. As a result, this project will be developed on the first scan only, after removing both subjects 1 and 21 and decreasing the length of all the time series to 403.

subject	missing.1	missing.2	missing.rate.1	missing.rate.2
subject_1	28280	28280	100 %	100 %
subject_2	70	70	0.2475 %	0.2475 %
subject_3	0	70	0 %	0.2475 %
subject_4	70	28280	0.2475 %	100 %
subject_5	70	0	0.2475 %	0 %
subject_6	0	70	0 %	0.2475 %
subject_7	0	70	0 %	0.2475 %
subject_8	0	70	0 %	0.2475 %
subject_9	0	0	0 %	0 %
subject_10	0	28280	0 %	100 %
subject_11	0	28280	0 %	100 %
subject_12	0	28280	0 %	100 %
subject_13	0	28280	0 %	100 %
subject_14	70	28280	0.2475 %	100 %
subject_15	0	28280	0 %	100 %
subject_16	0	28280	0 %	100 %
subject_17	0	28280	0 %	100 %
subject_18	0	28280	0 %	100 %
subject_19	0	70	0 %	0.2475 %
subject_20	70	28280	0.2475 %	100 %
subject_21	28280	28280	100 %	100 %
subject_22	0	70	0 %	0.2475 %
subject_23	0	0	0 %	0 %
subject_24	0	70	0 %	0.2475 %

Table 2: Distribution of missing values over the 24 subjects.

3 Analysis using repeated Dickey-Fuller tests

Before entering the analysis, let's briefly introduce the tools that we are going to use.

3.1 Unit Root Tests

Unit root tests are tests for stationarity in a time series. In general, we can say that a time series is non-stationary if it contains a unit root

$$unit\ root\ \Rightarrow\ non\ -stationarity\ ,$$

the reverse is not true.
While analyzing a time series, the existence of unit root(s) can cause several problems because many results of traditional statistical theory, such as the law of large numbers and the central limit theory, do not hold in these situations. For the unit root process, we need to apply the ARIMA model; that is, we take the difference (maybe several times) before applying the ARMA model. This is why it's important to spend some time understanding how they work and the mathematics behind them.

Let's start with the simplest situation: the first-order auto-regressive process $X_t \sim AR(1)$.

$$X_t = \phi_0 + \phi_1 X_{t-1} + \epsilon_t\ ,$$

where ϵ_t may or may not be white noise. We assume it to be zero-mean stationary ARMA process. The model can also be expressed with all X's on one side as

$$X_t - \phi_1 X_{t-1} = \phi_0 + \epsilon_t\ ,$$

using the backshift operator $BX_t = X_{t-1}$, we can re-express the model compactly as $X_t - \phi_1 BX_t = \phi_0 + \epsilon_t$ or equivalently,

$$(1 - \phi_1 B)X_t = \phi_0 + \epsilon_t\ .$$

The characteristic polynomial is $1-\phi_1 x$. This has a (unique) root at x = $\frac{1}{\phi_1}$. Therefore this process has a unit root if $\phi_1 = 1$, and in that case, the series converges to

$$X_t = \phi_0 t + X_0 + (\epsilon_1 + \epsilon_2 + ... + \epsilon_t).$$

The $\phi_0 t$ term implies that the series will be trending if $\phi_0 \neq 0$, the series is not mean-reverting and the mean is free to change over time. Also the variance is non-constant ($var(X_t) = var(\epsilon_1 + \epsilon_2 + ...\epsilon_t)$), which is a function of t. In short, the unit root process is not stationary. If, instead, we had $|\phi_1| > 1$, the series would be called an explosive non-stationary AR(1) process. In order to get a better idea, look at the following example.

Take two AR(1) processes:

- Process 1: $x_t = 0.5x_{t-1} + \epsilon_t$,
- Process 2: $x_t = x_{t-1} + \epsilon_t$.

Where $\epsilon_t \sim N(0,1)$ white noise. Process 1 has no unit root, while process 2 has one. We can verify this by calculating the characteristic polynomials as before.
Now imagine we start both processes off at zero, i.e. $x_1 = 0$, and that we have a drift of positive epsilons, and both processes get at $t = 10$ to the value of $x_{10} = 2$. What happens next? According to the distribution of ϵ_t we are expecting the future $\epsilon_{11}, \epsilon_{12}, ...$ to be equal to zero. Therefore:

- Process 1: $x_{11} = 1,\ x_{12} = 0.5,\ x_{13} = 0.25,\ x_{14} = 0.125,\ ...$, converging eventually to the initial value, 0.
- Process 2: $x_{11} = 2,\ x_{12} = 2,\ x_{13} = 2,\ ...$, stabilizing on this new position.

This proves that a process with a unit root sometimes stabilizes around a new position (not-mean reverting) due to historical good or bad luck, and it will shift around randomly, but there is nothing forcing it back to the initial trend. On the other hand, when there is no unit root the process doesn't blow up, and there actually is a force that will make the process drift back to the old position, although the random noise will still knock it around a bit.
Now, in order to generalize what we said to a higher-order process, let's take an autoregressive process of order p:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \epsilon_t.$$

Again, ϵ_t may or may not be white noise and we assume it to be a zero-mean stationary ARMA process. For convenience, we also assume that $X_0 = 0$. In order to find all the roots of the characteristic equation, we need to solve what follows:

$$z^p - z^{p-1}\phi_1 - z^{p-2}\phi_2 - .... - \phi_p = 0.$$

If $z = 1$ is one solution of the above equation with multiplicity 1, then the process is said to be integrated of order 1, and it has a (unique) unit root. If $z = 1$ is a root of multiplicity n, the process is said to be integrated of order n.
This was a brief introduction to unit roots processes. Further details are beyond the scope of this project and can be found in chapter 6.3 of Brockwell and Davis (1991) [7].

Now that we understand the importance of spotting unit root(s) for a time series, we need to find some detection procedure that allows us to test for their presence. There are many different tests that can be implemented to tackle this problem. Two of the most popular are the *Dickey-Fuller (DF) Unit Root Test* and its augmented version, the *Augmented Dickey-Fuller (ADF) Unit Root Test*. In the next subsections, we're giving a glimpse of these two methods. For more detailed information, refer to chapter 10 of Fuller (1996) [8].

3.1.1 Dickey-Fuller (DF) Unit Root Test

The Dickey-Fuller (DF) test is the most popular test for unit root. It is used to test for the presence of unitary roots in a process, and since $unit\ roots \Rightarrow non-stationarity$, what we are actually testing as null hypothesis is that the time series under study is non-stationary. Let's start considering the simplest case, i.e. we fit the zero-mean

AR(1) regression $X_t = \phi X_{t-1} + \epsilon_t$, with $\epsilon_t \sim NID(0, \sigma^2)$. Now subtract X_{t-1} from both sides obtaining what follows:

$$X_t \; - \; X_{t-1} \; = \; (\phi - 1) X_{t-1} \, \epsilon_t \; , \tag{1}$$

which becomes, in a little bit more neat way:

$$\Delta X_t \; = \; \delta X_{t-1} \; + \; \epsilon_t \; .$$

Where $\delta = \phi - 1$. The reason we did this transformation is to make the left-hand side stationary under the null hypothesis of $\delta = 0$. Moreover, note that testing for the presence of a unit root is equivalent to testing the null hypothesis:

$$H_0 : \phi = 1 \qquad or \;\; equivalently \qquad H_0 : \delta = 0 \; .$$

The alternative hypothesis under the DF test framework is

$$H_1 : \phi < 1 \qquad or \;\; equivalently \qquad H_1 : \delta < 0 \; .$$

Note that this is a one-sided alternative since we are interested in the alternative of stationarity or trend stationarity, and not in time series with explosive behavior, which are non-stationary in a way that cannot be rectified by differencing. Because of that, we need to reject the null hypothesis when the t-test is less than the critical value and we don't compute the absolute value of the t-test. The t-statistics can be computed as follows:

$$\hat{\tau} = \frac{\hat{\phi} - 1}{\hat{\sigma} (\sum_{t=2}^{n} X_{t-1}^2)^{-\frac{1}{2}}} \; , \tag{2}$$

with $\hat{\phi}$ and $\hat{\sigma}^2$ are the least squares estimates of ϕ and σ^2:

$$\hat{\phi} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2},$$

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^{n} (X_t - \hat{\phi} X_{t-1})^2}{n-1}.$$

If X_t is stationary (i.e. under the alternative hypothesis) and we set $T = \textit{length of time series}$, then it can be shown that (c.f. Hamilton (1994) pg. 216 [9])

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0, (1 - \phi^2)) \; ,$$

or

$$\hat{\phi} \overset{A}{\sim} N(\phi, \frac{1}{T}(1 - \phi^2))$$

and it follows that $t_{\phi=1} \overset{A}{\sim} N(0,1)$. However, under the null hypothesis of non-stationarity, the above results give

$$\hat{\phi} \overset{A}{\sim} N(1, 0),$$

which is undefined because when the standard deviation is zero, the Gaussian PDF turns into the Dirac delta function. This happens because under H_0, the process X_{t-1} on the

right-hand side of our model is non-stationary. Therefore it is highly persistent (non-ergodic) and the autocorrelation decays to zero very slowly. This implies that the usual sample moments do not converge to fixed constants. Instead, Phillips (1987) showed that the sample moments of X_t converge to random function of Brownian motion.

$$\hat{t} \xrightarrow{d} \frac{\int W\, dW}{\left(\int W^2 dW\right)^{1/2}}.$$

W stands for Brownian motion, and the numerator is a stochastic integral. The mathematics behind this result stands completely beyond the scope of this project, and we leave the most curious readers with the following two references: lecture notes from Washington University (2016) [10] and Phillips (1987) [11].
Following the references, the test statistic $t_{\phi=1}$ follows the distribution called Dickey-Fuller (DF) distribution and doesn't have a closed form representation. Consequently, quantiles of the distribution must be computed by numerical approximation or by computer simulation: at the significance levels 0.01,0.05,0.1, the critical values are -2.58 , -1.95 , -1.62 , respectively.
At the end of the day, we only need to compute the same old t-statistic that we are used to and compare its value with the DF critical value. Then we either find that this t-statistic is less than the critical value or that it is greater than the critical value. In the first scenario, we reject H_0, which in real terms means that we reject that the process has a unit root saying that the time series is, in fact, stationary. On the other hand, following the second scenario, we do not reject H_0, and we do not have enough evidence to say that the time series is non-stationary. This is how we decide for a simple AR(1) model whether the series is or is not stationary.
Now, of course real time series can and will be more complicated than AR(1). So how do we extend this to something more complicated? That's where the ADF test comes in. See next subsection.

3.1.2 Augmented Dickey-Fuller (ADF) Unit Root Test

The DF unit root test described above has a very narrow scope, since it is only valid if the time series is well characterized by an AR(1) process with white noise errors. In most of the cases, however, we are in the presence of higher-order processes and processes with more complicated dynamic structures. What if the error is correlated, for instance?
In order to augment the basic autoregressive unit root test to accommodate general ARMA(p,q) models, Said and Dickey (1984) [12] created the Augmented Dickey-Fuller (ADF) test.
The ADF test is based on estimating the test regression

$$\Delta X_t = \alpha + \beta t + (\phi - 1)X_{t-1} + \sum_{i=1}^{k} \psi_i \Delta X_{t-i} + \epsilon_t \tag{3}$$

where α and βt represent deterministic terms, the constant term and the trend term, respectively, and βt can be dropped if the series is not trending. The k lagged difference terms, ΔX_{t-i}, are used to approximate the ARMA structure of the errors,

and the value k is set so that the error ϵ is serially uncorrelated. The term ψ_i represents the polynomial of the moving average representation of an ARMA(p,q) process:

The null and the alternative hypothesis for the ADF are actually the exact same as for the DF test. We're again testing whether $\phi = 1$ versus $\phi < 1$, and the procedure is the same. We go ahead and calculate the t-statistic for ϕ and compare it with the same Dickey-Fuller distribution as before and go to the same conclusion process.

There is one last question to be addressed, how many lags are *enough*?
Choosing the number of lags for the ADF test is an important practical issue for the implementation of the ADF test. If k is too small, the remaining correlation in the errors will bias the test. If k is too large, then the power of the test will suffer.
Luckily, differently from X_{t-1}, the lagged terms are stationary, and therefore we can directly test their coefficients (ψ_i , $i = 1, 2, ..., k$), computing the t-statistic for each of them and compare these statistics against critical values in the typical t-distribution, and then we go ahead and make conclusions about whether each of these is significant. Obviously, we can also test all of them together using an F-test. In addition to this, there is an iterative procedure that allows us to find a good trade-off between low correlation in the errors and good power of the test. This procedure was introduced by Ng and Perron (1995) [13] and is the following:

1. Set an upper bound k_{max} for k,
2. Estimate the ADF regression with $k = k_{max}$
3. If the absolute value of the t-statistic for testing the significance of the last lagged difference is greater than 1.6, set $k = k_{max}$ and proceed with the unit root test. Otherwise remove the last lagged difference and repeat.

A useful rule of thumb for determining k_{max} is :

$$k_{max} = \left[12 \left(\frac{T}{100} \right)^{\frac{1}{4}} \right] \tag{4}$$

Alternatively, a solution that may be less *ad hoc* but definitely faster is to choose the following number of lagged differences:

$$k = (T - 1)^{\frac{1}{3}}$$

Our future analysis is developed using R software and is based on this choice of k

3.2 Analysis of brain scans data using repeated Dickey-Fuller test

Now that we have introduced all the necessary tools and notations, we can proceed with our inquiry into the neurological dataset.
As already mentioned in the introduction, the goal of the analysis is to study whether subjects with different traits - e.g., handedness, age, and neurobehavioural diseases - exhibit different levels of stability in their BOLD series at rest.

In this first approach, we are basing the analysis on a series of Augmented Dickey-Fuller tests. The idea is to run an ADF test for all the brain regions (70) of each subject (22) and respectively store the number of rejected tests. In this way, we obtain a statistic that is easily comparable across subjects and that can be modeled as a function of the individuals' traits, an example is shown in Figure 2.

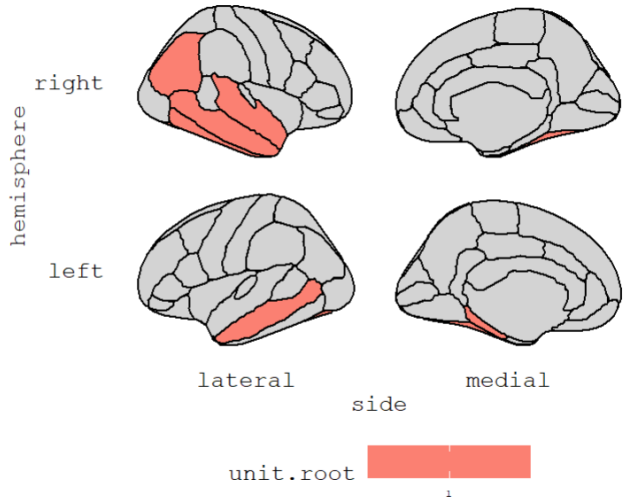


Figure 2: Brain regions for Subject 9 accordingly to the Desikan atlas parcellation into 70 regions.

Chances are that individuals with different characteristics, like age or presence of some type of diagnosis, result in having a different number of brain regions rejecting the unit root significantly, and that a properly specified model can detect the influence of each trait. Now, before proceeding, it should be mentioned that in order to perform any regression, the general rule of thumb is that *if you expect to be able to detect effects with reasonable power, you need 10-20 observations (subjects) per parameter (covariate) estimated* (Frank Harrell (2015) [14]). Yet, we only have 22 subjects, among which 18 only have complete information about current or lifetime diagnosis. Hence, we don't have enough individuals to find a wealth of potentially interesting patterns, which would require far more individuals. That is why we decided to restrict our research on the effect of the age and of the presence/absence of any type of diagnosis only. In this way, we model the "number of series rejecting the unit root significantly" ($y = stationary$) only as a function of a quantitative variable ($x_1 = age$) and a categorical dichotomous variable ($x_2 = diagnosis$). Where *diagnosis* for subject i is $TRUE$ if the subject shows at least one of the two types of diagnosis (current/lifetime), and it is $FALSE$ otherwise. The reader should keep in mind, nonetheless, that our small sample of subjects may as well not be representative of the real behavior of the underlying population.
In order to proceed with the regression we need to decide on the type of function that we want to fit. On one hand, it is true that our dependent is a count of events, and that count data usually make the Poisson distribution one of the options to take into account. On the other hand, with so many observations at right at the maximum 70, our dependent is actually different from a Poisson distribution. Therefore, a scatter plot could give us some insights on which type of regression is best suited for our dependent, Figure 3. From the scatter plot we immediately realize that something in the way we

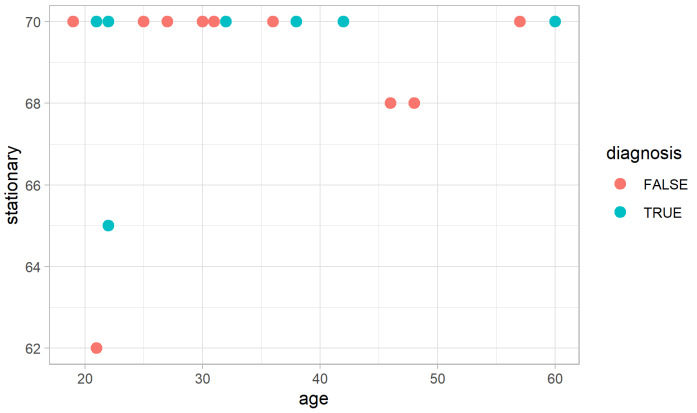


Figure 3: Scatterplot for the dependent variable *stationary* against the independent *age*, colored accordingly to presence (*TRUE*) or absence (*FALSE*) of *diagnosis*.

defined our dependent variable is malfunctioning, as there are too many observations taking the same value (70). There are only four observations that don't reach the maximum for the *stationary* variable, among which three are healthy subjects without any kind of diagnosis, and one has current or lifetime diagnosis. Therefore, points appear to follow a bimodal distribution and we could summarize the situation by dividing subjects into two classes: those who reach 70 and those who don't. However, since the great majority of points take value 70 out of 70 for the *stationary* dependent variable (Y), it makes no sense to design any GLS regression, which assumes a distribution around Y. Thus, problem in this case lies somewhere outside model specification, and is related to how the dependent variable is specified. It could be that the number of series rejecting the unit root significantly cannot really show a substantial and data analytically clear difference between the time series. In order to investigate this point further, and therefore understand whether we should focus on something different from stationarity, let's start by looking at the following graph, Figure 4:

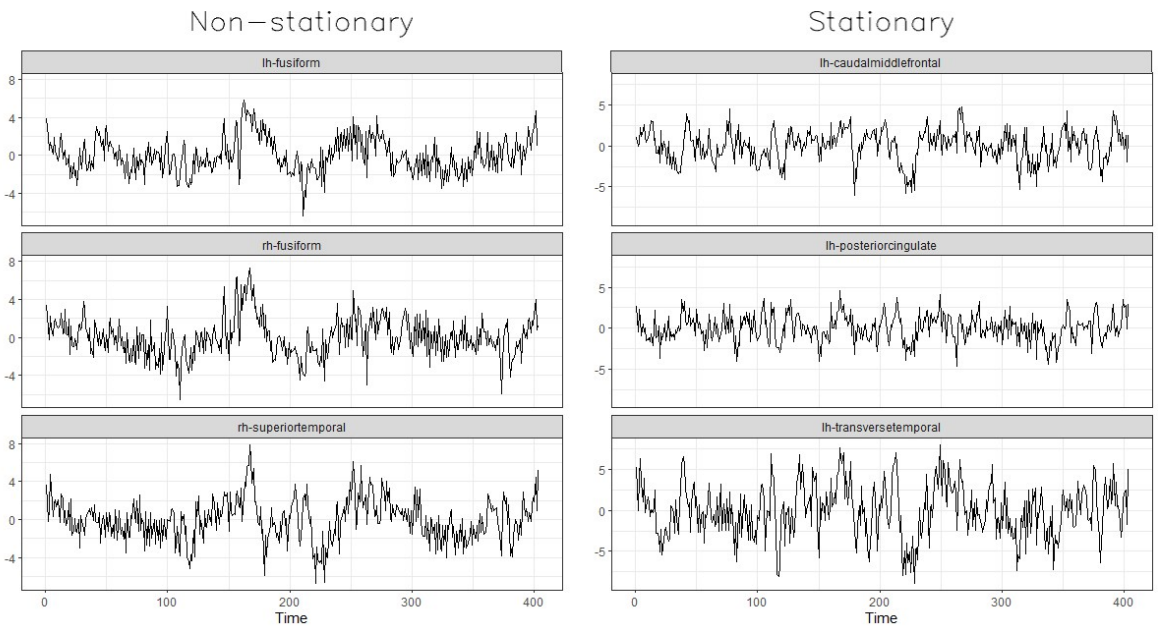


Figure 4: Subject 9. On the left: 3 randomly selected series that resulted non-stationary to the ADF test; on the right: three randomly selected series that resulted stationary to the ADF test.

Despite being on a slightly different scale, looking at the figure we can realize how the test for stationarity of the series does not really pierce our research purpose, from the moment that series labelled as non-stationary and stationary don't look very different. This actually holds not only for subject 9, but for all the subjects in our analysis. To strengthen the concept, we can also see how subjects with very different traits scored the same, while subjects with similar traits scored differently. Take for instance subject 9, 22 and 23, Table 3, next page.

By comparing subject 9 with subject 22 we observe that different traits correspond to

sub_id	age	handedness	current_diagnosis	lifetime_diagnosis	stationary
9	21	L	NO	NO	62
22	42	R	YES	YES	70
23	31	L	NO	NO	70

Table 3: Comparing stationarity number for subject 9, 22 and 23.

different stationary number. But if we compare subject 22 with subject 23, or subject 9 with subject 23, we can see what we said before, i.e. the *stationarity* statistic is not able to capture the effect of the covariates of interest.

3.2.1 Conclusions

This method is flawed; all the estimated coefficients are non-significant and even if they were, taking a closer look at the coefficients estimates made us realize that they are not really meaningful from the moment that the dependent variable is not well suited for our final goal. We should proceed with a different plan of action.

4 Analysis using outlier detection

Analyzing stationarity region-wise for all the subjects was not a successful idea, mainly because the time series representing the blood oxygenation level dependent (BOLD) at rest are predominantly stationary, regardless of the traits of the subject. On that account, instead of focusing on the global stationarity of the time series, we decided to concentrate on the detection of spontaneous activations of the brain. Hopefully, this will bring more differentiation among subjects with different characteristics. In order to do so, we assume that spontaneous activations correspond to extreme values of the underlying distribution of our BOLD time series. This procedure can be pinned down to an outlier detection procedure.

Outlier detection for temporal data is a broad and complex field. While a detailed exposition is beyond the scope of this project, we will provide an overview of the key ideas behind the approach we are going to implement, especially from a computer science (algorithm) standpoint.
Temporal outliers can lead to model misspecification, biased parameter estimation, and poor forecasts. However, our aim is not to solve any of these issues, but mere detecting the outliers that can be classified as unsolicited mind activations. Therefore our focus is not a classical data cleaning, instead, we are interested in analyzing the outlier itself (Fig. 5).

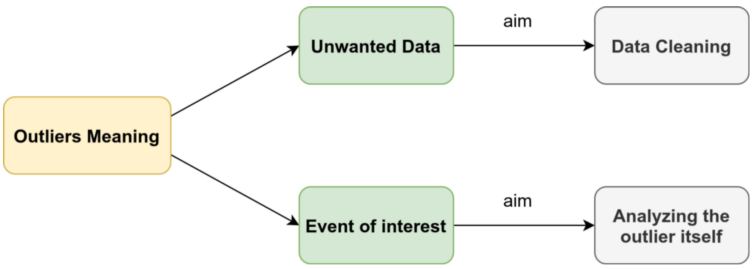


Figure 5: Types of interpretation of outliers.

In the statistics community, outlier detection for time series data has been studied for decades, pioneered by Box and Tiao(1975)in their quest to solve the Los Angeles pollution problem and continued by several others. In order to produce the results at the end of this section, we have referred to Manish Gupta *et al.* (2013) [15] and Ruey S. Tsay (2000)*et al.* [16] for outlier detection in multivariate time series; F. Matarise *et al.* (2012) [17] for types of outliers; J.F.MacGregor (1995) [18] and Daniel Pena (2006) [19] for process control of multivariate processes. In the following subsections, we are giving a brief overview of the main concepts we are going to implement in the analysis, starting from the types of temporal outliers.

4.1 Four types of temporal outliers

Given the dynamic structure of time series, there are various definitions of outliers. In this subsection, we introduce four types of outliers, which, accordingly to the all-encompassing paper by F. Matarise (2012)[17] are (AO) Additive Outlier, (IO) Innovational Outlier, (TC) Temporary Change, and (LS) Level Shift. Their effects can be

observed in the following charts, which are taken from F. Matarise, who obtained them through simulation, Figure 6:

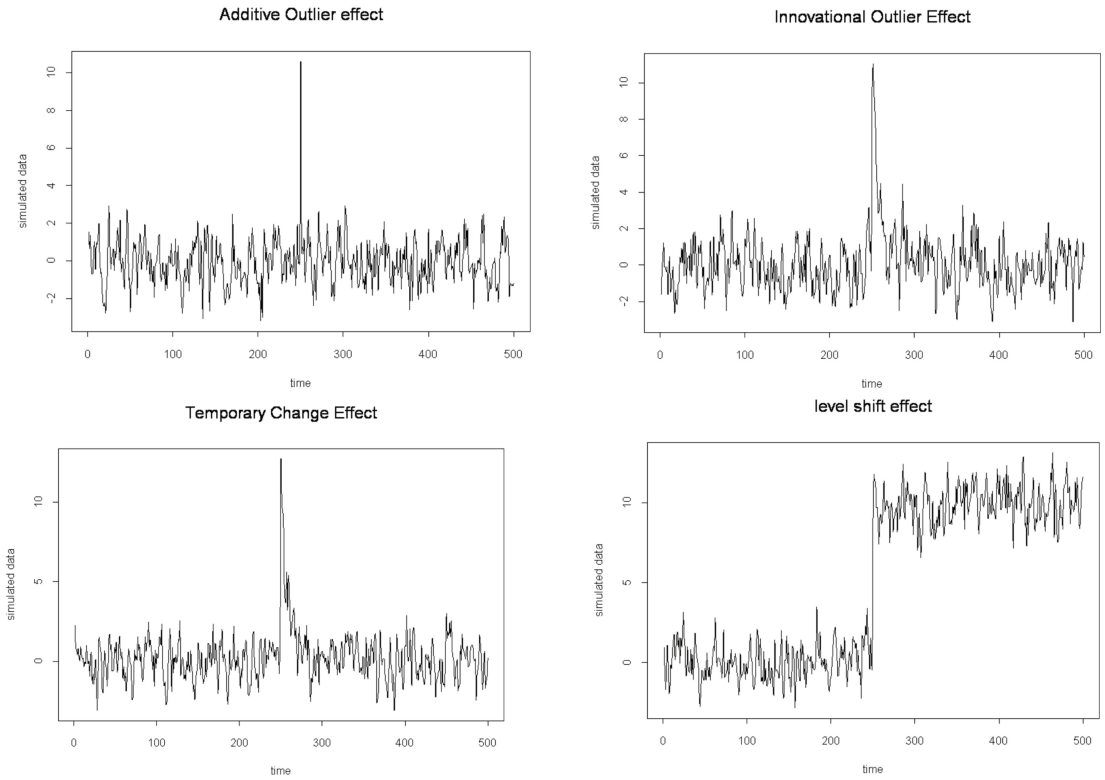


Figure 6: Effects of the four types of outliers.

The Additive Outlier (AO) is usually referred to as a gross error affecting the $t-th$ observation as shown in the figure; the Level Shift (LS) is an abrupt but permanent shift in the series caused by an intervention; the Temporary Change (TC) is not a one time effect like the Additive Outlier but results in an effect at time t and dies out gradually; the Innovation Outlier (IO) is similar in effect to the TC when dealing with weakly dependent time series, but in case of highly persistent time series, IO may produce a level shift for various types of ARIMA models. Weakly dependent mathematically means that $corr(x_t, x_{t+h}) \to 0$ as $h \to \infty$, where h is the time lag. This means that x_t is becoming less and less correlated with its future values as we move further into the future. In this case, the stochastic process is said to be asymptotically uncorrelated. The significance of this assumption is that it replaces the assumption of random sampling through the use of The Law of Large Numbers and The Central Limit Theorem.

In order to describe them in real terms, let's introduce some needful notation.
Let $t = 1, ..., T$ and $x_t = (x_{1t}, \ ... \ , x_{kt})'$ be a k-dimensional time series that can be modelled as an multivariate (vector) autoregressive integrated moving-average model (VARIMA):

$$\Phi(L)x_t = c + \Theta(B)\epsilon_t, \tag{5}$$

where

$$\Phi(L) = I - \Phi_1 L - .... - \Phi_p B^p, \ \ \Theta(L) = I - \Theta_1 L - .... - \Theta_q L^q$$

are $k \times k$ matrix polynomials of finite degrees p and q, L is the lag operator (also known as back-shift operator) such that $Lx_t = x_{t-1}$, c is a k-dimensional constant vector, and $\{\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{kt})'\}$ is a sequence of *iid* Gaussian random vectors with zero mean and positive-definite covariance matrix Σ.
It's also useful to define the autoregressive representation and the moving average representation of our model in (5) :

- AR representation : $\Pi(L)x_t = c_0 + \epsilon_t$,
  where $c_0 = \{\Theta(1)\}^{-1}c$ and $\Pi(L) = \{\Theta(L)\}^{-1}\Phi(L)$,
- MA representation : $x_t = c_* + \Psi(L)\epsilon_t$,
  where $c_* = \{\Phi(1)\}^{-1}c$ and $\Psi(L) = \{\Phi(L)\}^{-1}\Theta(L)$.

In addition, let $\zeta_t^{(h)}$ be an indicator variable for time index $h = 1, ..., T$, such that $\zeta_t^{(h)} = 1$ if $t = h$ and $\zeta_t^{(h)} = 0$ if $t \neq h$. Denote the observed time series by $y_t = (y_{1t}, ..., y_{kt})'$, and let $\omega = (\omega_1, ..., \omega_k)'$ be the size of the initial impact of an outlier on the series x_t. More details on ω are given under the outlier detection procedure section, §4.2.

Finally, the four types of outliers can be equated to:

$$y_t = x_t + \alpha(L)\omega\zeta_t^{(h)}, \tag{6}$$

- $\alpha(L) = \Psi(L)$ for Innovation Outlier,
- $\alpha(L) = I$ for Additive Outlier,
- $\alpha(L) = (1 - L)^{-1}I$ for Level Shift,
- $\alpha(L) = \{D(\delta)\}^{-1}$ for Temporary Change.

In the last equation $\{D(\delta)\}$ is a k x k diagonal matrix with diagonal elements $\{(1 - \delta L), ..., (1 - \delta L)\}$ and $0 < \delta < 1$

4.2 Outlier detection procedure

Given a time series database, there are sundry methods aimed at identifying point outliers or subsequence outliers. These method ranges from basic statistical decomposition and can work up to autoencoders.
The procedure we are going to implement comes from Tsay *et al.* (2000) [16] idea of building a vector ARIMA model to identify the types of outliers. It can be summarised as follows:

1. Build a VARIMA model for the series under study, assuming no outliers and let $\hat{a}_t$ be the estimated residuals and $\hat{\Pi}_i$ the estimated coefficients of the autoregressive representation described before.
2. Estimate the outlier size by using $\hat{\omega}_{i,h} \quad i = I, A, L, T$ depending on the type of outlier: innovational, additive, level shift or temporary change. More details will follow.

3. Test the null hypothesis $H_0 : \omega = 0$ against the alternative $H_1 : \omega \neq 0$. To do so, two statistics are used :
   Joint test statistic:
   $$J_{i,h} = \hat{\omega}'_{i,h} \Sigma^{-1}_{i,h} \hat{\omega}_{i,h},$$
   Component test statistic:
   $$C_{i,h} = \max_{1 \leq j \leq k} \frac{|\hat{\omega}_{j,i,h}|}{\sqrt{\sigma_{j,i,h}}},$$
   where $i = I,\ A,\ L,\ or\ T$; $\Sigma_{i,h}$ is the covariance matrix of the estimator of $\omega_{i,h}$ and $\sigma_{i,j,h}$ is the $j-th$ diagonal element of $\Sigma_{i,h}$. We can also define the overall test statistics as:
   $$J_{max}(i, h_i) = \max_h J_{i,h}, \quad C_{max}(i, h_i^*) = \max_h C_{i,h} \quad (i = I, A, L, T), \tag{7}$$
   h_i denotes the time index when the maximum of the test statistic $J_{i,h}$ occurs and h_i^* denotes the time index when the maximum of $C_{i,h}$ occurs. These statistics can be compared to their empirical critical values at the end of the paper.

4. Once an outlier is identified, its impact on the underlying time series is removed accordingly to equation (6). The adjusted series is treated as the new dataset and the detecting procedure is iterated. We terminate the detection procedure when no significant outlier is detected.

For a multivariate innovational outlier at time index h, we can easily estimate the outlier size by using $\hat{\omega}_{I,h} = \hat{a_h}$, where the subscript I indicates innovational outlier. For the other types of outliers, the same estimation idea applies, and we shall give details for the case of a multivariate additive outlier only, from the moment that it is the only other type that will be useful in the final analysis, the reason will be clear in a moment. In this case, we have:

$$\hat{a}_t = \left(I - \sum_{i=1}^{\infty} \hat{\Pi}_i L^i \right) \zeta_t^{(h)} \omega + \epsilon_t = \left(\zeta_t^{(h)} - \sum_{i=1}^{\infty} \hat{\Pi}_i \zeta_{t-i}^{(h)} \right) \omega + \epsilon_t,$$

with $\epsilon_t \sim N(0, \Sigma)$, and the estimator of ω_A is

$$\hat{\omega}_{A,h} = - \left(\sum_{i=0}^{n-h} \hat{\Pi}'_i \Sigma^{-1} \hat{\Pi}_i \right)^{-1} \sum_{i=0}^{n-h} \hat{\Pi}'_i \Sigma^{-1} \hat{a}_{h+i} \quad (\Pi_0 = -I). \tag{8}$$

This can be interpreted as a generalized least squares estimator. Besides, the covariance matrix of the estimator is $\Sigma_{A,h} = \left(\sum_{i=0}^{n-h} \hat{\Pi}'_i \Sigma^{-1} \hat{\Pi}_i \right)^{-1}$.

On top of this, under the null hypothesis of no outlier in the sample and if we assume that the model of x_t is known, $J_{max}(I, h_I)$ is the maximum of a random sample of size n from a chi-squared distribution with k degrees of freedom. Thus, the asymptotic distribution of $J_{max}(I, h_I)$ can be obtained using the extreme value distribution. Unfortunately, the same doesn't hold for the other joint test statistic $J_{max}(A, h_A)$, which is a maximum of a dependent sample from a chi-squared distribution with k degrees of freedom. Its asymptotic distribution is therefore more complicated, since it depends

on the serial dependence of the test statistic. In order to find critical values, we need a simulation study. In addition, asymptotic distributions of $C_{max}(i, h_i^*)$ also depend on serial correlations of $\{C_{i,h}\}$. Therefore, we use the already simulated values to obtain finite sample critical values for the two test statistics that can be consulted in section §6.
Now, if a single joint statistic $J_{max}(i, h_0)$ is significant at time index h_0, we identify a multivariate outlier of type i at h_0, where $i = I, A$. In the case of multiple significant joint test statistics, we identify the type based on the test that has the smallest p-value. When both the joint statistics are non-significant at a given level, we use the component statistics $C_{max}(i, h_i^*)$ to check for additional outliers.

4.3 Identification of the VARMA Model and Estimation of Multivariate Time Series

The first step of the mentioned detection procedure expects us to specify and fit a vector ARMA (VARMA) model. Under the univariate framework, this was done by simply sticking to the Box-Jenkins procedure. With our multivariate structure, however, this requires a little bit more attention. Let's spend a few words on this.

4.3.1 Identification of the VARMA(p,q) model

In order to identify the correct p and q for our vector ARMA model, we adopt the Ljung-Box (LB) statistics Q. The LB test is a way to test for the absence of serial autocorrelation up to a specified lag h. The LB test is used to investigate the correlation in a time series' residuals, and therefore it gives insights on whether the model we are fitting is taking into account all the correlation structure of the true (unknown) generating process, or whether some of this correlation is captured by the error terms. Essentially, it is a test of lack of fit: if the autocorrelations of the residuals are not significantly different from zero, we conclude that the model does not show 'significant lack of fit'. The hypothesis system under study is the following:

- H_0 : residuals are independently distributed, i.e. our model *does not* show lack of fit;
- H_1 : residuals are not independently distributed; therefore they exhibit serial correlation, and the model *does* show a lack of fit.

Given a fitted model VARMA(p,q), we can compute the LB statistic on its residuals as:

$$Q = T(T+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{T-k} \tag{9}$$

where T is the time series length, $\hat{\rho}_k^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested. Under H_0 the statistic Q asymptotically follows a $\chi^2_{(h)}$, therefore, for a given significance level α, the critical rejection region of the hypothesis of randomness is : $Q > \chi^2_{1-\alpha,h}$, where $\chi^2_{1-\alpha,h}$ is the $(1-\alpha)$-quantile of the chi-squared distribution with h degrees of freedom.

For model specification, we need to find a compromise between goodness of fit and parsimony, therefore We search for the combination of minimum p and q in the

VARIMA(p,q) model such that the Ljung-Box statistics is non-significant, i.e. there is not enough evidence to reject the independence of the residuals.

In the analysis, we used the *R* function *Eccm*() inside the *MTS* package. The function takes as argument a data matrix (T-by-k) of a vector time series, where T is the sample size and k is the dimension, and returns a two-way table of the p-values of Ljung-Box test statistic computed for different p and q combinations.

4.3.2 Estimation of the VARIMA model

In order to understand the estimation of a VARIMA model, a great reference is Box and Tiao's publication *Modeling multiple time series with applications* [20]. As described in the paper, the estimation of the model is achieved by performing a repeated pseudo regression procedure.

In the analysis, we used the *R* function *marima*() inside the *marima* package. The function takes as argument the time series matrix, and returns all the necessary elements for performing the outlier detection procedure, among which both AR and MA estimates, covariance matrix of residuals, covariance matrix of (all) input data and so forth.

4.4 Dimensionality Reduction

The estimation that we just described is unfortunately often unfeasible because of the computational complexity that characterizes it. In actual fact, especially for high-dimensional time series, many of the existing tools for dealing with VARMA models face the challenge of complexity in their structures. This difficulty occurs because the number of parameters expands enormously fast as the dimension increases. Therefore reducing the size of the series becomes critical to manage such data. In this section, we aim to give the readers a brief introduction to dimensionality reduction in multivariate time series and how we decided to proceed.

Many approaches have been proposed in the literature for dimension reduction, like canonical analysis, scalar component models, reduced rank models, and factor models. Another approach of interest is the principal component analysis (PCA). PCA seeks dimension reduction by retaining a small number of (principal) components that are linear combinations of the original variables. PCA is, in fact, a commonly used technique to perform dimension reduction for static and independent multivariate data. However, because of the dynamic nature of multivariate time series data, the classical PCA technique could be problematic, and some of the techniques mentioned few lines above are likely to perform better. Nevertheless, without being strictly dogmatic, we can analyze the following result from Daniel Pena (2006) [19].
Let x_t be defined as at the beginning of §5.1 , i.e. a $k \times 1$ vector, and assume it represents a stationary process with mean μ. We derive that its unconditional joint probability distribution does not change when shifted in time. Consequently, parameters such as mean and variance also do not change over time, and we can define the covariance matrix as:

$$\Gamma_x(l) = E\left[(x_{t-l} - \mu)(x_t - \mu)'\right], \tag{10}$$

where l is the time lag between observations, and, since we are assuming stationarity, $\Gamma_x(0)$ represents the variance of the dataset. We are interested in finding linear combi-

nations $y_{1t} = m'x_t$ of the vector of time series with maximum variance. Now recall the Wold Representation Theorem:

> ***Wold Representation Theorem****: Let x_t be a (weakly) stationary process. Then x_t can be written as the sum of a linear combination of lags of white noise plus a deterministic trend $x_t = \Psi(L)\epsilon_t = \sum_{j=0}^{L\to\infty} \psi_j \epsilon_{t-j} + dt$; where $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ and $\psi_0 = 1$; $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$.*
> *Assuming x_t has zero mean, the first and second moments of x_t are $E(x_t) = 0$ and $Var(x_t) = \sigma_\epsilon^2 \times \sum_{j=0}^{\infty} \psi_j^2 < \infty$.*

The last two identities are easily demonstrated as follows:
let x_t be a zero mean (weakly) stationary process, by the Wold representation, we can write it as $x_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, with the deterministic trend equal to zero. Then

$$E(x_t) = E\left(\sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}\right) = \sum_{j=0}^{\infty} \psi_j E(\epsilon_{t-j}) = 0 \ ;$$

$$Var(x_t) = E(x_t^2) = E\left((\epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + ...)(\epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + ...)\right) =$$

$$\sigma_\epsilon^2 + \psi_1^2\sigma_\epsilon^2 + \psi_2^2\sigma_\epsilon^2 = \sigma_\epsilon^2 \times \sum_{j=0}^{\infty} \psi_j^2 < \infty \ .$$

Going back to PCA, since we are assuming stationarity, we can define the model for the linear combination y_{1t} as $y_{it} = \psi(L)u_t$; then $Var(x_{it}) = \sigma_u^2 \sum \psi_i^2$. Now, linear combinations which are white noise will be associated to a small variance (since they are independent and their covariance between time points t and $t-k$ will be zero and the variance small); and linear combinations close to non-stationarity will be associated to a large variance (since they have non-zero covariance between time points t and $t-k$, and changes at point $t-k$ may have a long lasting impact). This association suggests looking for linear combinations of large variance or small variance, it is well known that they will be given by the eigenvectors m_i in

$$\Gamma_x(0)m_i = \lambda_i m_i,$$

and the corresponding eigenvalues, λ_i, will be the variances of the linear combinations. If we remove the assumption of stationarity, this approach can be extended to the non-stationary case. Assume that x_t is a non-stationary integrated process $I(d)$, then equation (10) becomes:

$$C(l) = \frac{1}{T^{2d}} \sum (x_{t-l} - \bar{x})(x_t - \bar{x})', \tag{11}$$

where $\bar{x} = T^{-1} \sum x_t$. From here, we can get the final solutions by solving the equation:

$$C(0)m_i = \lambda_i m_i.$$

The reasons behind this results will not be explained here, for more information, please refer to Daniel Pena (2006) [19]. Although we realize there may be some methodologies that perform a better job, we decided to proceed in our analysis using PCA. Searching

for the best methodology to be applied would deviate from the original aim of this project, and it would be so demanding and laborious that it could represent a project on its own. However, we leave to the interested readers the following references: *Jolliffe - chapt.12 of Principal Component Analysis* [21], *Alshammri - MDPCA* [22], *Daniel Pena - Dimension Reduction in Multivariate Time Series* [19].

To culminate the brief description of the dimensionality reduction procedure, we should notice that, once again, the final tables containing empirical quantiles (extrapolated from Tsay *et al.* (2000) [16]) are putting some restrictions on our analysis. In fact, the maximum number of dimensions that the tables support is three, delineating an upper bound for the number of principal components that we are able to retain.

4.5 Analysis of brain scans data using Outlier Detection procedure

Finally, we can apply all the methods we just mentioned to our brain scans study. What follows (§4.5.1) is an illustration of the steps performed on subject 6, which is an exemplification of what has been done on all the subjects at hand. After that, in (§4.5.2), we display and interpret the results obtained for all the subjects, and then we close this paper with an epilogue containing the main take-home concepts (§4.5.3).

Note that: all the computations have been performed using *R software*, and only the main functions will be briefly described. For major references on this, see the appendix when indicated.

4.5.1 Outlier detection on subject 6

The instance we discuss is developed on the fifth subject, which we recall is 27 y.o. , right-handed and has no record of current nor lifetime disease.
First and foremost, we start by taking a quick glimpse at the structure of subject 6, Fig. 7:

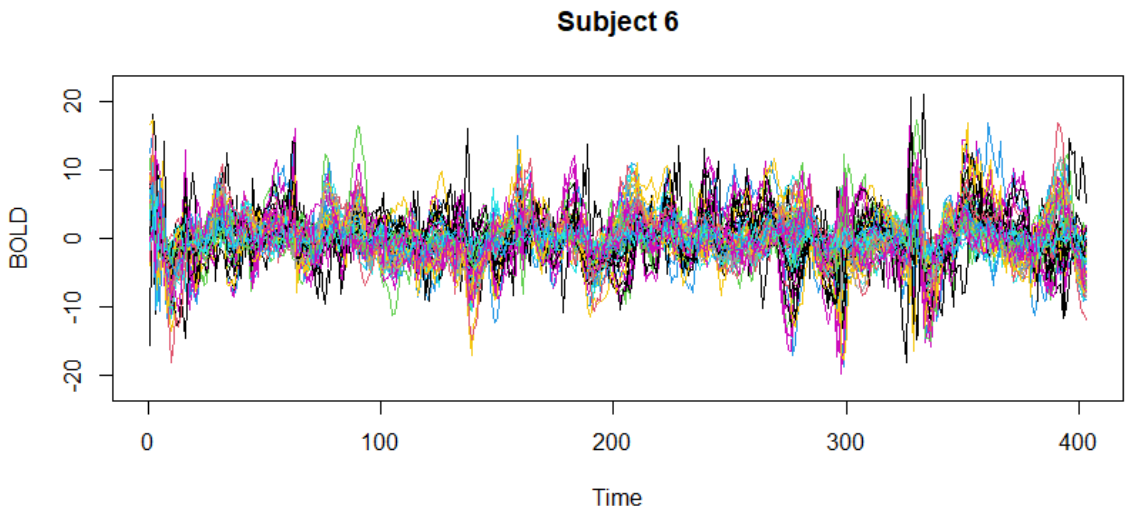


Figure 7: Time series plot for subject 6. 70 brain regions measured over 403 points in time.

It is fairly difficult to extrapolate valid information from this plot, and the reason is that subject 6 is highly dimensional. As we said, this complicates things when it comes to its analysis, especially if we are planning to specify a time model for it. Thus, before going any further, we need to proceed with a dimensionality reduction. From the results obtained in §3.2, more specifically in Table **??**, we know that 70 out of the 70 time series that compose subject 6 are stationary accordingly to the ADF test. Because of that, we can implement a classical PCA as mentioned in the previous subsection §4.4 without worrying about the differentiating order of our vector time series. We use the built-in *R* function *princomp()*, which performs a principal components analysis on the given numeric data matrix.

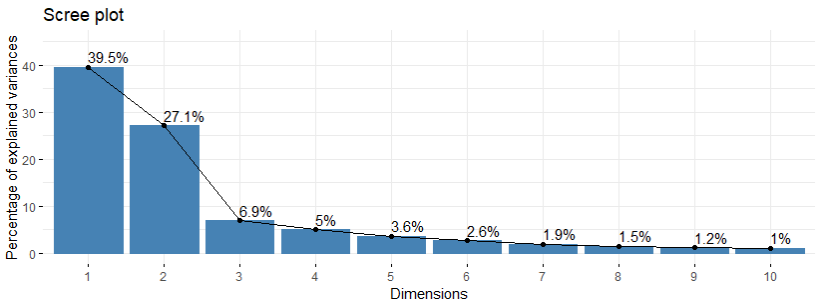


Figure 8: Screeplot PCA(Subject 6)

Figure 8 represents a screeplot, which depicts the variances against the number of the principal component. Doing little math computations, we can calculate the cumulative percentage of explained variance, which is already around 73.5% taking only the first 3 components, while it goes beyond the usual threshold of 80% with 5 components. As already mentioned when introducing the tools for the analysis, the maximum number of components we can retain is 3, which in any case preserves the structure of the

data fairly well, by visually comparing the principal components in 9 with the original components.

Figure 9: Principal components for subject 3

In the top part of Figure 9, we can observe how easier is now to visualize subject 6 with respect to Figure 7.
Now that we have a smaller number of dimensions, we can fit a VARIMA(p,d,q) model on subject 6. Since all the three components are stationary (ADF test), the order of integration d of our $VARIMA(p, d, q)$ model is 0, and we can identify the optimal order of p and q by means of Ljung-Box test statistic as described in §4.3.1. This is done by calling *Eccm* function in R that returns the following table of p-values, Table 4.

	0	1	2	3
0	0.00	0.00	0.00	0.00
1	0.00	0.00	0.00	0.18
2	0.00	0.00	0.27	0.61
3	0.19	0.84	0.97	0.97

Table 4: P-values of the Ljung-Box statistics for subject 3. Column: MA order, q. Row: AR order, p.

As already explained, in the table we shall seek for minimum p and q order corresponding to non-significant p-values. Even though, in other situations, one may prefer to select $p = 2$ and $q = 2$ from the moment that it gives a larger p-value with the same level of parsimony, we go for $p = 1$ and $q = 3$. This was done because the tables containing the empirical quantiles at the end of the paper were computed using vector AR(1) and

AR(6) models. Hence, in order to obtain meaningful comparisons, smaller AR order has been preferred at the cost of a higher MA order.

Thus, for subject 2 we identify a $VARMA(1,3)$.

Before entering the core of the detection procedure, let's take a closer look at Figure 9. The components are characterized by some noise that knocks the series around their mean. Yet, all of them seem to drift back to the old position after each shock. Therefore, the only anomalies that we could be interested in detecting are innovational outliers, temporary changes, and additive outliers. Anyhow, since the detection of IO and TC is very similar, we will only focus on AO and IO.

Applying the proposed detection procedure and using 10% critical values, we summarise the detection results in the following table 5 . Ten outliers have been detected by the procedure. Once an outlier was detected, we removed its effects on the data and re-estimated the trivariate $VARMA(1,3)$ model. By exploiting the results obtained in §4.1 we know that removing the effect of an additive outlier at time h is relatively trouble-free since it affects the h-th observation only and has no effect on its neighbors. Therefore we just need to solve the already described equation $y_t = x_t + \alpha(L)\omega\zeta_t^{(h)}$ for x_t with $\alpha(L) = I$ and $\zeta_h^{(h)} = 1$ while $\zeta_t^{(h)} = 0$ if $t \neq h$. We obtain:

$$x_h = y_h - I \times \omega_{I,h}.$$

Things complicate if we consider removing the effect of an innovation outlier. An innovational outlier(IO) is an extraordinary shock at time h influencing y_h, y_{h+1}.... To dwell the results obtained when discussing the four types of outliers, we know that the model $y_t = x_t + \alpha(L)\omega\zeta_t^{(h)}$ has $\alpha(L) = \Psi(L)$, where $\Psi(L) = \{\Phi(L)\}^{-1}\Theta(L)$ is the polynomial of coefficients in the moving average representation, which is an infinite polynomial of the form $\{\Psi(L) = (I + \Psi_1 L + \Psi_2 L^2 + ... + \Psi_k L^k + ...)\}$. Therefore we have:

$$\begin{aligned} y_t &= x_t + \alpha(L)\omega\zeta_t^{(h)}, \\ y_t &= x_t + \Psi(L)\omega\zeta_t^{(h)}, \\ y_t &= x_t + (I + \Psi_1 L + \Psi_2 L^2 + ...)\omega\zeta_t^{(h)}, \end{aligned}$$

Now note that the lag operator will lag the ζ_t back in time, but will not affect the (h) which is still the point that we detected to be IO. This means

$$\begin{aligned} y_h &= x_h + \omega\zeta_h^{(h)} + \Psi_1\omega\zeta_{h-1}^{(h)} + \Psi_2\omega\zeta_{h-2}^{(h)} + ... \\ y_{h+1} &= x_{h+1} + \omega\zeta_{h+1}^{(h)} + \Psi_1\omega\zeta_{h+1-1}^{(h)} + \Psi_2\omega\zeta_{h+1-2}^{(h)} + ... \\ y_{h+2} &= x_{h+2} + \omega\zeta_{h+2}^{(h)} + \Psi_1\omega\zeta_{h+2-1}^{(h)} + \Psi_2\omega\zeta_{h+2-2}^{(h)} + ... \\ &\vdots \\ y_{h+k} &= x_{h+k} + \omega\zeta_{h+k}^{(h)} + \Psi_1\omega\zeta_{h+k-1}^{(h)} + ... + \Psi_k\omega\zeta_{h+k-k}^{(h)} + ... \\ &\vdots \end{aligned}$$

Since $\zeta_h^{(h)} = 1$ while $\zeta_t^{(h)} = 0$ if $t \neq h$. We obtain

$$y_h = x_h + \omega\zeta_h^{(h)}$$

$$y_{h+1} = x_{h+1} + \Psi_1\omega\zeta_h^{(h)}$$

$$y_{h+2} = x_{h+2} + \Psi_2\omega\zeta_h^{(h)}$$

$$\vdots$$

$$y_{h+k} = x_{h+k} + \Psi_k\omega\zeta_h^{(h)}$$

Potentially $h + k$ could achieve T, i.e. the length of the time series. However, in the implementation of the analysis, we decided to keep k at the fixed level of 5, which we believe accounts for the major effects of the innovational outlier.

.

(a) Joint test statistics

iterations	$J_{max}(I, h_I)$	$J_{max}(A, h_A)$	**Time**	**Type**
1	45.53 (188)	70.66 (188)	188	AO
2	40.83 (133)	39.26 (133)	133	IO
3	40.04 (355)	38.99 (321)	355	IO
4	38.38 (321)	37.08 (321)	321	IO
5	36.06 (128)	39.99 (331)	331	AO
6	37.04 (128)	35.05 (356)	128	IO
7	37.16 (190)	33.49 (356)	190	IO
8	29.78 (356)	34.29 (356)	356	AO
9	20.42 (90)	25.60 (132)	-	-
Crit.	32.24	32.66		

(b) Component test statistics

iterations	$C_{max}(I, h_I^*)$	$C_{max}(A, h_A^*)$	**Time**	**Type**
9	4.15 (90)	4.53 (322)	90	IO
10	3.97 (320)	4.56 (322)	320	IO
11	3.82 (25)	4.50 (322)	-	-
Crit.	3.90	4.56		

Table 5: width=.7 Results of multivariate detection for subject 6. J stands for *Joint* statistics, while C stands for *component* statistics. Appendix I indicates *Innovation* outliers, while A indicates *additive* outliers. *Crit.* represents the critical values for our test statistics at a 10% significance level; its values are taken from the statistical tables in section §6 using AR(1), trivariate case, and sample size 400.

What we can conclude is that for subject 6, being 27 y.o., right-handed and with

no current nor lifetime diagnosis, we have 10 brain activations at rest.

4.5.2 Global results

The procedure carried out on subject 6 is now performed on all the subjects, and the number of outliers is stored. Let's explore the results of the outlier detection algorithm in relation with the same variables that we defined during the analysis with repeated ADF tests, namely *age* and *diagnosis*.
Looking at Figure 10.a we can clearly see that the great majority of blue points, which represent diagnosed subjects, lay above the red ones, which instead represent healthy individuals. Therefore it is evident that the presence of any type of diagnosis has a positive effect on the number of outliers. The plot also suggests some negative dependency on age, especially for the diagnosed group.
Similar conclusions can be drawn from the dot plots in Figure 10.b, where we categorized the quantitative variable "age" into two classes: *older* > 30 and *younger* ≤ 30. For the *diagnosis* variable is still evident that the diagnosed population has on average a higher number of outliers, since the dots are completely shifted to the right. Less evident, but still detectable, is the effect of the age class, for which we can see that the younger population has a longer right tail.
To briefly recap, and keeping in mind that we are relating outliers to spontaneous brain activations, we can say that:

- **age** : slightly negative effect on the number of spontaneous brain activations.
- **current and lifetime diagnosis**: positive effect on the number of spontaneous brain activations.

To get insights on how striking those results are, we can design a regression model. Since the dependent variable *outliers* represents count data and takes values far from the upper bound, a Poisson regression model may be of interest. Poisson regression explains the log of the expected number of outliers as a linear function of the independent variables as follows:

$$log(E(Y|x)) = \alpha + \beta' x.$$

In the model, $Y \in \mathfrak{R}^n$ is the dependent variable, $x \in \mathfrak{R}^n$ is a vector of independent variables (covariates), $\alpha \in \mathfrak{R}$ is the intercept and $\beta \in \mathfrak{R}^n$ is the vector of coefficients. The results are shown in Table 6 below.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.43	0.37	3.89	0.00
age	-0.03	0.01	-2.72	0.01
diagnosisTRUE	1.71	0.25	6.73	0.00

Table 6: Estimated coefficients using Poisson regression model.

As we were expecting, $diagnosis = TRUE$ has a relatively large positive effect while age has sightly negative effect on the number of spontaneous brain activations. Both

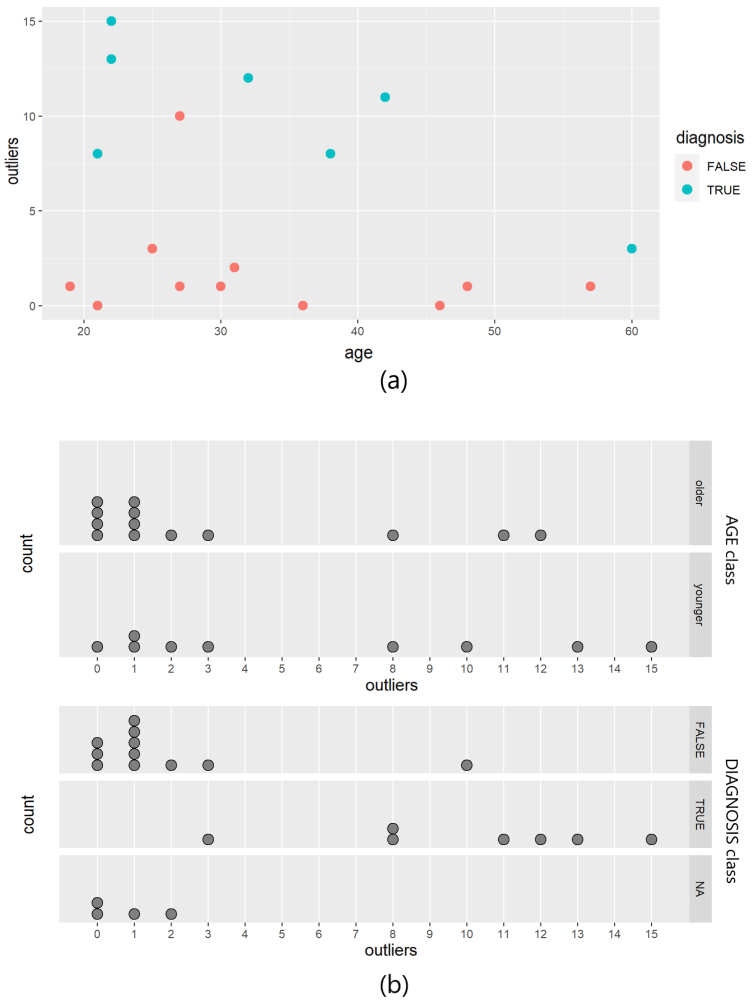


Figure 10: (a) Scatterplot age vs. outliers with coloring by diagnosis; (b) dot plots for number of outliers classified on *diagnosis* and *age*, where age has been categorized into two classes: *older* $>$ 30 and *younger* $\leq$ 30.

the effects appear to be strongly significant, a result that was perplexing at the beginning. Considering our small sample size we were expecting less outstanding p-values. Especially for the *age* variable, Figure 10 doesn't suggest anything that significant. However, upon closer inspection, it seems that the strong significancy for *age* comes from two crucial observations, the one with largest *age* and the one with largest *number of outliers* among $diagnosis = FALSE$. As a matter of fact, if we remove these two observations, the p-value associated to the *age* coefficient becomes larger than 0.1, loosing all its significancy. On top of that, let's recall that the conditional variance should be equal to the conditional mean according to the Poisson regression assumptions. Then, one possible cause that forced the negative coefficient for *age* into significance is that the model couldn't consider a constant age effect, because the two observations that we mentioned would have looked out of order.

In addition to this, one may argue that looking at Figure 10.a the variables *age*

and *diagnosis* may interact, and therefore their interaction should be considered when fitting the Poisson model. However, we decided not to include the interaction for the following two points:

- The impression of interaction is mainly caused by a single observation (the one with oldest age).
- We have few data, and adding another regressor to our model may be risky. We would end up in a situation in which the estimators would not be precise enough to tell significantly what exactly is going on.

To conclude, we can say that, on average, we expect subjects suffering from any current or lifetime diagnosis (keeping the other covariates fixed) more likely to have unstable BOLD (blood oxygenation level dependent) and consequently more spontaneous neuronal activations. On the contrary, older subjects seems to be slightly more likely to have fewer number of activations, even though the significance of this effect perches on two crucial observations as described above.
Once again, we would like to remind the reader that these results are based on 22 individuals only, among which only 18 have complete information about the diagnosis status. Therefore these results may as well not be representative. As an example, take subject 6, the red dot with the largest *number of outliers* in Figure 10.a . This subject has a number of spontaneous activations that can't be explained based on our previous interpretation. Despite being young, and therefore more apt to have a less stable BOLD, his/her 10 activations are hard to come only from this, and considering that he/she doesn't have any lifetime nor current diagnosis, this draws precariousness on our model. Still, it may also be that our sample is truly representative, and subject 6 is an actual outlier. We can ask rheumatological experts and cognitive neuroscientists to comment on the findings to check for this. Otherwise, the only alternative is to collect a larger sample. This method, nonetheless, seems already much more suited than the ADF analysis to investigate our query.

Below you can find Table 7, which reports the total number of additive and innovational outliers for subjects in the study.

Subject	IO	AO	Subject	IO	AO
2	0	0	13	5	10
3	0	1	14	0	1
4	0	0	15	1	0
5	1	1	16	2	1
6	7	3	17	4	4
7	2	1	18	0	0
8	2	6	19	7	6
9	0	0	20	5	7
10	1	0	22	5	6
11	0	1	23	2	0
12	1	0	24	0	0

Table 7: Number of outliers for all the subjects under study.

5 Conclusions

The project gives a detailed walkthrough of the procedure that has been implemented in order to compare brain activities for different subjects. Two main methods have been proposed. One based on the stationarity analysis of each brain region, subject by subject. The other one based on an outlier detection algorithm. When describing the methods, in both cases, we introduced all the necessary tools to perform the final analysis. The level of details was dictated by our willingness to give the reader a crystal clear picture of what is going on and why some methods are problematic and some are not.
Despite being unhelpful for the final goal, the analysis using ADF repeated tests shows the difficulties that can be encountered when dealing with this type of tasks, namely how difficult it is to come out with an effective statistic and how hard is the task of a statistician when dealing with few (but complex) information. This, together with the fact that it also represents a good way to get to know the data, is the reason why we decided to keep it inside the report.
Quite noticeably, the core of the project is the outlier detection procedure and its implementation. Behind this section there was intense studying and research. In order to accomplish the final task of outlier detection, we successfully dealt with the problem of dimensionality reduction for multivariate time series, we solved the issue of VARIMA model specification, for which there is very scarce literature, and finally we created an R code that was able to identify and store all the interesting outliers accordingly to the mentioned algorithm.

At the end of the day, this thesis has brought me studying advanced topics under the framework of time series analysis, which I am passionate about. Moreover, it gave me the possibility to play in the backyard of neuroscience, which pushed forward my interest towards data-science applications in health-related issues.

With this being said, it is indubitably true that there is still the possibility for future work on the analysis using outlier detection algorithm:

- One way to improve this analysis would be of course the one of enlarging our sample size. This will result in more accurate estimates of the contribution of each trait to the brain activations while at rest.
- Also of interest could be to perform our own simulation study on the fitted models in order to create even more precise empirical critical values. This could be rather simple to do.
- Another option would be combining the temporal analysis that we have performed with some type of spatial data, obtain a more sensible spatio-temporal analysis. This could be done, for instance, by also considering the correlation structure among the different brain regions(ROIs), taking into account both the geometric distance between pairs of ROIs and their anatomic distance. Specifically by defining spatial weights that are directly proportional to the number of white fibers and inversely proportional to the Euclidean distance.

6 Statistical tables

These statistical tables come from Tsay *et al.* (2000) [16] and are obtained through a simulation study. They obtained critical values and power of the test statistics in § 4.2 . They achieved this by employing two vector AR(1) and AR(6) models to obtain quantiles for $k = 2, 3$ number of dimensions and for sample sizes $n = 100, 200, 400$. For further details, please look at the reference.

Table 1. *Simulation study. Empirical quantiles of the* $J_{\max}(i, h_i)$ *statistics in* (6) *based on* 10 000 *realisations.*

Sample size	Test	Probability 50%	90%	95%	97·5%	99%	50%	90%	95%	97·5%	99%
		Vector AR(1)					Vector AR(6)				
		Bivariate case									
100	$J_{\max}(I, h_I)$	9·74	13·03	14·35	15·60	17·34	9·76	13·08	14·43	15·61	17·23
	$J_{\max}(A, h_A)$	9·70	13·07	14·32	15·57	16·96	9·74	13·63	15·08	16·63	18·59
	$J_{\max}(L, h_L)$	7·61	11·13	12·37	13·50	14·82	6·31	9·33	10·44	11·63	13·30
	$J_{\max}(T, h_T)$	9·58	12·95	14·27	15·43	17·05	9·33	12·92	14·34	15·65	17·58
200	$J_{\max}(I, h_I)$	11·20	14·66	16·01	17·47	19·06	10·80	14·24	15·54	16·82	18·41
	$J_{\max}(A, h_A)$	11·13	14·66	15·95	17·37	19·18	10·70	14·57	16·08	17·70	19·38
	$J_{\max}(L, h_L)$	8·37	12·18	13·49	14·81	16·40	6·20	9·15	10·36	11·53	13·17
	$J_{\max}(T, h_T)$	11·04	14·55	15·87	17·19	18·67	10·12	13·72	15·14	16·47	18·17
400	$J_{\max}(I, h_I)$	12·60	16·19	17·63	19·06	20·81	11·79	15·12	16·47	17·75	19·54
	$J_{\max}(A, h_A)$	12·56	16·21	17·64	18·86	20·81	11·47	15·26	16·71	18·25	20·16
	$J_{\max}(L, h_L)$	9·62	13·48	14·88	16·20	18·05	6·37	9·16	10·25	11·31	13·05
	$J_{\max}(T, h_T)$	12·57	16·13	17·53	18·96	20·83	10·58	14·18	15·53	17·02	18·80
		Trivariate case									
100	$J_{\max}(I, h_I)$	15·55	25·00	29·56	34·23	41·43	14·43	23·75	28·10	32·50	38·82
	$J_{\max}(A, h_A)$	15·50	25·08	20·49	34·00	41·81	14·66	25·10	30·14	34·98	42·23
	$J_{\max}(L, h_L)$	10·64	18·09	21·56	25·46	32·06	8·62	15·94	19·85	24·39	27·75
	$J_{\max}(T, h_T)$	15·48	25·14	30·05	34·81	42·18	14·03	24·14	29·24	36·61	41·12
200	$J_{\max}(I, h_I)$	19·20	28·45	32·10	36·72	42·10	16·44	26·94	31·85	36·77	43·16
	$J_{\max}(A, h_A)$	19·10	28·90	33·04	37·28	43·02	16·40	27·96	33·31	38·90	45·49
	$J_{\max}(L, h_L)$	12·12	19·99	23·24	26·43	31·51	8·71	16·22	20·28	24·13	29·81
	$J_{\max}(T, h_T)$	19·12	28·86	32·85	37·18	43·29	15·56	27·18	32·48	37·68	44·88
400	$J_{\max}(I, h_I)$	22·86	32·24	36·12	40·04	45·09	18·78	30·54	35·69	40·95	47·32
	$J_{\max}(A, h_A)$	22·96	32·66	36·49	40·32	45·77	18·11	30·70	36·11	41·61	48·89
	$J_{\max}(L, h_L)$	14·50	22·89	26·39	30·27	34·56	9·20	17·16	20·90	24·68	29·50
	$J_{\max}(T, h_T)$	23·10	32·79	36·67	40·67	45·47	17·11	30·01	35·08	40·66	48·01

Table 2. *Simulation study. Empirical quantiles of the statistics* $C_{\max}(i, h_i^*)$ *in* (6) *based on* 10 000 *realisations.*

Sample size	Test	Probability 50%	90%	95%	97·5%	99%	50%	90%	95%	97·5%	99%
		Vector AR(1)					Vector AR(6)				
		Bivariate case									
100	$C_{\max}(I, h_I^*)$	2·89	3·39	3·58	3·74	3·96	2·89	3·37	3·54	3·69	3·87
	$C_{\max}(A, h_A^*)$	2·89	3·39	3·57	3·73	3·94	2·99	3·58	3·78	3·98	4·22
	$C_{\max}(L, h_L^*)$	2·61	3·18	3·35	3·52	3·71	2·41	2·93	3·10	3·26	3·50
	$C_{\max}(T, h_T^*)$	2·87	3·37	3·55	3·74	3·95	2·76	3·29	3·47	3·64	3·86
200	$C_{\max}(I, h_I^*)$	3·11	3·60	3·78	3·95	4·15	3·05	3·52	3·69	3·84	4·03
	$C_{\max}(A, h_A^*)$	3·11	3·60	3·78	3·93	4·15	3·16	3·74	3·92	4·12	4·34
	$C_{\max}(L, h_L^*)$	2·74	3·33	3·50	3·68	3·88	2·38	2·89	3·06	3·22	3·46
	$C_{\max}(T, h_T^*)$	3·09	3·58	3·76	3·93	4·11	2·90	3·40	3·58	3·73	3·96
400	$C_{\max}(I, h_I^*)$	3·32	3·80	3·96	4·13	4·35	3·20	3·65	3·81	3·96	4·17
	$C_{\max}(A, h_A^*)$	3·31	3·80	3·97	4·12	4·32	3·30	3·83	4·01	4·21	4·43
	$C_{\max}(L, h_L^*)$	2·94	3·51	3·69	3·86	4·06	2·39	2·84	3·00	3·16	3·34
	$C_{\max}(T, h_T^*)$	3·31	3·78	3·95	4·12	4·34	3·00	3·48	3·66	3·84	4·05
		Trivariate case									
100	$C_{\max}(I, h_I^*)$	3·01	3·48	3·64	3·79	3·96	3·03	3·50	3·67	3·84	4·03
	$C_{\max}(A, h_A^*)$	3·24	3·93	4·18	4·44	4·74	3·35	4·28	4·66	5·04	5·52
	$C_{\max}(L, h_L^*)$	2·77	3·54	3·83	4·11	4·52	2·66	3·61	4·05	4·44	4·97
	$C_{\max}(T, h_T^*)$	3·23	3·94	4·18	4·45	4·78	3·18	4·15	4·54	4·91	5·41
200	$C_{\max}(I, h_I^*)$	3·22	3·69	3·85	4·02	4·20	3·19	3·66	3·82	3·98	4·15
	$C_{\max}(A, h_A^*)$	3·56	4·24	4·50	4·72	4·97	3·56	4·45	4·81	5·20	5·68
	$C_{\max}(L, h_L^*)$	2·90	3·67	3·93	4·18	4·49	2·64	3·63	4·05	4·40	4·87
	$C_{\max}(T, h_T^*)$	3·55	4·28	4·52	4·77	5·06	3·36	4·35	4·78	5·16	5·63
400	$C_{\max}(I, h_I^*)$	3·43	3·90	4·07	4·21	4·38	3·35	3·81	3·97	4·14	4·32
	$C_{\max}(A, h_A^*)$	3·86	4·56	4·80	5·03	5·34	3·76	4·72	5·07	5·40	5·81
	$C_{\max}(L, h_L^*)$	3·08	3·84	4·13	4·39	4·67	2·63	3·62	3·98	4·35	4·79
	$C_{\max}(T, h_T^*)$	3·87	4·64	4·89	5·12	5·41	3·57	4·60	4·97	5·32	5·75

References

[1] The NKI-RS Team. Enhanced Nathan Kline Institute - Rockland Sample. http://fcon_1000.projects.nitrc.org/indi/enhanced/, 2017. [Online; accessed 01-May-2022].

[2] Colcombe Milham. NKI - Nathan Kline Institute. http://fcon_1000.projects.nitrc.org/indi/CoRR/html/nki_1.html. [Online; accessed 01-May-2022].

[3] Smyser C. D. Lee, M. H. and J. S. Shimony. *Resting–state fMRI: a review of methods and clinical applications.* American Journal of Neuroradiology, 2013.

[4] Child Mind Institute. C-PAC: Configurable Pipeline for the Analysis of Connectomes . https://fcp-indi.github.io/, 2017. [Online; accessed 01-May-2022].

[5] Child Mind Institute. Pre and Post processing - Time series extraction . http://fcp-indi.github.io/docs/user/tse.html, 2017. [Online; accessed 01-May-2022].

[6] Alkaline Software. web's free ICD-9-CM medical coding reference . http://www.icd9data.com/2013/Volume1/290-319/default.htm, 2017. [Online; accessed 01-May-2022].

[7] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting.* Springer, 2002.

[8] Wayne A Fuller. *Introduction to statistical time series.* John Wiley & Sons, 2009.

[9] James Douglas Hamilton. *Time series analysis.* Princeton university press, 2020.

[10] Chang-Jin Kim ?? Unit root tests. *econ584*, 2016.

[11] Peter CB Phillips. Time series regression with a unit root. *Econometrica: Journal of the Econometric Society*, pages 277–301, 1987.

[12] Said E Said and David A Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

[13] Serena Ng and Pierre Perron. Unit root tests in arma models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90(429):268–281, 1995.

[14] Frank E Harrell et al. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*, volume 3. Springer, 2015.

[15] Manish Gupta, Jing Gao, Charu C Aggarwal, and Jiawei Han. Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and data Engineering*, 26(9):2250–2267, 2013.

[16] Ruey S Tsay, Daniel Pena, and Alan E Pankratz. Outliers in multivariate time series. *Biometrika*, 87(4):789–804, 2000.

[17] Florance Matarise, Lawrence Dhliwayo, Immaculate Susan Nduna, Innocent Maposa, and Lillian Siziba. Detecting level shifts, temporary changes and innovational outliers in intervention analysis. *International Journal of Statistics and Systems*, 7(3):241–254, 2012.

[18] John F MacGregor and Theodora Kourti. Statistical process control of multivariate processes. *Control engineering practice*, 3(3):403–414, 1995.

[19] Daniel Pena and Pilar Poncela. Dimension reduction in multivariate time series. In *Advances in distribution theory, order statistics, and inference*, pages 433–458. Springer, 2006.

[20] George C Tiao and George EP Box. Modeling multiple time series with applications. *journal of the American Statistical Association*, 76(376):802–816, 1981.

[21] Ian T Jolliffe. *Principal component analysis for special types of data*. Springer, 2002.

[22] Fayed Alshammri and Jiazhu Pan. Moving dynamic principal component analysis for non-stationary multivariate time series. *Computational Statistics*, 36(3):2247–2287, 2021.