On Two-Sample Tests For Time Series

Abeer Alzahrani

A Thesis
in
The Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Degree of
Master of Science (Mathematics) at
Concordia University
Montréal, Québec, Canada

September 2017

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Abeer Alzahrani**
Entitled: **On Two-Sample Tests For Time Series**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.
Signed by the Final Examining Committee:

_____ Chair
Dr. Yogendra Chaubey

_____ Examiner
Dr. Yogendra Chaubey

_____ Examiner
Dr. Arusharka Sen

_____ Supervisor
Dr. Wei Sun

Approved by _____
Chair of Department or Graduate Program Director

_____ 2017 _____
Dean of Faculty

Abstract

On Two-Sample Tests For Time Series

Abeer Alzahrani

In this thesis, we consider the two-sample problem of time series. Given two time series data $x_1, ..., x_n$ and $y_1, ..., y_m$, we would like to test whether they follow the same time series model. First, we develop a unified procedure for this testing problem. The procedure consists of three steps: testing stationarity, comparing correlation structures and comparing residual distributions. Then, we apply the established procedure to analyze real data. We also propose a modification to a nonparametric two-sample test, which can be applied to high dimensional data with equal means and variances.

Acknowledgments

I would like to express my sincerest gratitude to Professor Wei Sun, whose expertise and guidance have supported me during my whole graduate experience and inspired me on my path towards my master's degree. A thank you is also extended to my mother and my father, who inspired me to strive to bring out the best in me and to my brothers and my sisters, who always remind me to be the best role model for them.

Contents

List of Figures

List of Tables

Introduction

In this thesis, we consider the two-sample problem of time series. The problem is stated as follows:

Given two time series data $x_1, ..., x_n$ and $y_1, ..., y_m$ with the models M_1 and M_2, respectively, we want to test whether $M_1 = M_2$ or $M_1 \neq M_2$.

This problem has been widely considered and has a broad range of applications. For example, in microarray analysis, suppose there are two time series observations, one is considered as a control set and the other is exposed to an infection. People are intrested in testing if both samples share the same expression or if the infection has altered the exposed set. This problem has been studied in [20] by Stegle et al. where they purposed a robust Bayesian two-sample test to detect the difference of the two time series.

In marine ecosystems studies, time series are viewed as documentations and records of different species over long periods of time. It is important to compare the time series on different periods and to see relations between different species. Smith et al. [19] studied the relation between anchovy biomass and its primary production in Southern California Bight.

Two sample tests have been used to compare the performances of different methods. In [12], people discussed which method is more efficient in forecasting macroeconomic time series and further estimate the difference. Zhang et al. [21] compared four time series methods by using infectious disease data.

In the literature, different approaches have been applied to the two-sample problem of time series. The test proposed by Stegle et al. is efficient for gene problems with small sample sizes. Smith et al. [19] introduced a comparison method based on their data. In addition, Montero-Manso et al. [13] proposed a test for time series based on pairwise distance. In order to obtain a better performance in some cases, they required a large sample

size. They also needed the time series to have the same length. Some other papers considered hypothesis testing for time series. In these papers the time series model is assumed to be the $AR(1)$ model and the noise term is assumed to be Gaussian [18] or exponential [15].

In this thesis, we discuss the two-sample problem of time series in a more general perspective. Our analysis completely depends on the samples only, without prior information on the noise distributions and with one sample from each model. We consider this problem by following a procedure that shows where the differences between the time series samples are. We will detect if the differences are from the correlation structures of the models or from the noise distributions.

Our first step is to test stationarity of the two samples. If one sample is stationary and the other is not, then there is no need to continue the study. We conclude that the samples are from different models. If both of them are stationary or both are non-stationary, we continue our study. For the case that both of them are non-stationary we find a suitable transformation to make one sample stationary and then apply the same transformation to the other sample. If the second transformed sample is also stationary we continue the study. Chapter 1 presents tests that will be used to check the stationarity of the samples.

Our second step is to investigate the two stationary time series or the two stationary transformed time series. We will compare their correlation structures. This will be done by following the method preposed in [16], where the χ^2-test is used to test correlation based on residuals of a fitted autoregressive model. Our second step will be presented in Chapter 2.

If the two samples share the same correlation structure, we then apply a two-sample test on the residuals of each sample. Given that we have only one sample from each model, an applicable test to use would be the Bayesian two-sample test proposed by Holmes et al. [7], which uses the Pólya tree process as a prior distribution. This is step three that will be introduced in Chapter 3. By following these three steps we can decide whether the two time series share the same model. If not, we identify whether the difference is caused by the correlation structures or the noise distributions.

If there are at least four samples from each model we also consider other tests. These include the Biswas and Ghosh test (BG-test) [2] and its slight modification, the BG^3-test.

The BG-test is an inter-point test that is based on the Baringhaus and Franz test (BF-test) [1]. Different from the BF-test, the BG-test is also applicable to dependent data. A main assumption of the BG-test is that samples should have different means or different variances. To overcome this difficulty, we suggest the BG^3- test which will also be introduced in Chapter 3.

In Chapter 4, we apply the established procedure to analyze real data. The data is obtained from Environment and Climate Change Canada, which monitors the national long-term water quality. We also use illustrative examples to show the advantage of using the BG^3-test. In the appendix, we give the R codes, which might be useful for interested readers.

Chapter 1

Stationarity of Time Series

Roughly speaking, a time series is stationary if the mean, variance and autocorrelation structure do not change over time. Given a time series $\{X_t, t \in \mathbb{Z}\}$ where $E(X_t^2) < \infty$. Denote $\mu_X(t) = E(X_t)$, and $\gamma_X(r,s) = Cov(X_r, X_s),\ \forall\, t, r$ and $s \in \mathbb{Z}$. $\{X_t\}$ is said to be stationary (weakly stationary) if (cf. [4]):

(1) $\mu_X(t)$ is independent of t.

(2) $\gamma_X(t+h,t)$ is independent of t for each $h \in \mathbb{Z}$.

We say that $\{X_t\}$ is strictly stationary if the joint distributions of $(X_1, ..., X_n)$ and $(X_{1+h}, ..., X_{n+h})$ do not change for all $h \in \mathbb{Z}$ and $n > 0$. Throughout this thesis when we use the term stationary we mean weakly stationary.

Given a time series data, the basic method to test its stationarity is to consider the data plot and the sample autocorrelation function (ACF) plot.

Definition 1.1 *([4])*
Let $x_1, ..., x_n$ be observations of a time series. The sample mean of $x_1, ..., x_n$ is:

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

The sample autocovariance function is:

$$\hat{\gamma}(h) = n^{-1}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n.$$

The sample autocorrelation function is:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n.$$

If the $|\hat{\rho}(h)|$ plot shows a slow decay as h increases, this indicates that the sample has a trend component. If the $|\hat{\rho}(h)|$ plot exhibits a periodic behavior, then it indicates that the sample has a seasonal component. Hence, the sample ACF can be used to check stationarity of the observed data. We should note that the lags (h) of the sample ACF should not be close to the sample size and should not be larger than a third of the sample size for it to be a good estimate of the ACF. If the time series data is stationary, we will try to fit it with an $AR(p)$ model.

To further analyze the stationarity of the time series data, we consider the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS-test) [10] and the augmented Dickey-Fuller test (ADF-test) [17]. The KPSS-test and the ADF-test check stationarity and non-stationarity through analyzing the existence of unit roots. Suppose a time series $\{y_t\}$ follows an $AR(p)$ model:

$$y_t - \phi_1 \, y_{t-1} - \phi_2 \, y_{t-2} - \cdots - \phi_p \, y_{t-p} = \epsilon_t.$$

By rewriting the model with the backward shift operator B, we get:

$$(1 - \phi_1 \, B - \phi_2 \, B^2 - \cdots - \phi_p \, B^p) y_t = \epsilon_t, \tag{1}$$

where $By_t = y_{t-1}$ and $B^j(y_t) = y_{t-j}, \; j \geq 1$. We check for unit roots by considering the characteristic equation. To this end, we replace the backward shift operator B in (1) by a variable x and find the zeros of the equation:

$$1 - \phi_1 \, x - \phi_2 \, x^2 - \cdots - \phi_p \, x^p = 0.$$

$\{y_t\}$ is regarded as stationary if all solutions satisfy $|x| \neq 1, \; \forall x \in \mathbb{C}$. Here $\mathbb{C}$ denotes the set of all complex numbers. $\{y_t\}$ is non-stationary if a unit root $|x| = 1$ exists. The non-stationarity in the case of unit roots is overcome by differencing.

The KPSS-test tests the null hypothesis of trend-stationarity against the unit root existence. By trend-stationarity we mean that the removal of the deterministic trend will result in a stationary time series. The test assumes that a time series could be decomposed in the

following form:

$$y_t = \xi t + r_t + \epsilon_t, \tag{2}$$

where ξt is the deterministic trend, $r_t = r_{t-1} + u_t$ is a random walk with $u_t \sim iid\ (0, \sigma_u^2)$, r_0 is fixed, and ϵ_t is a stationary error. Based on this decomposition, the null hypothesis is equivalent to testing $\sigma_u^2 = 0$ which results in a trend-stationary time series. To test this hypothesis, the Lagrange multiplier (LM) statistic is used. Based on the upper-tail LM statistic test we accept or reject H_0, under the assumptions: u_t are normal and $\epsilon_t \sim^{iid} N(0, \sigma_\epsilon^2)$. The test hypotheses are:

$$H_0 :\ \sigma_u^2 = 0 \quad vs. \quad H_1 :\ \sigma_u^2 > 0.$$

Denote the residuals from (2) by e_t and the estimate of the error variance of the residuals by $\hat{\sigma_\epsilon^2}$, with $S_t = \sum_{i=1}^{t} e_i,\ t = 1, ..., T$. The LM statistic is calculated as ([10]):

$$LM = \sum_{t=1}^{T} S_t^2 / \hat{\sigma_\epsilon}^2.$$

This statistic was derived as a special case from the one developed by Nabeya and Tanaka [14]. Considering the decomposition in (2), if $\xi = 0$ then it becomes a level-stationary problem around the value of r_0 rather than a trend-stationary problem. The test statistic in this case is calculated by replacing e_t with $y_t - \bar{y}$.

As for the ADF-test, the null hypothesis is that the time series has a unit root, hence is not stationary. Assume we have an AR(p) model $\{y_t\}$:

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p} = \epsilon_t.$$

By rewriting the time series, we get:

$$y_t - y_{t-1} = (\sum_{i=1}^{p} \phi_i - 1) y_{t-1} - (\sum_{i=2}^{p} \phi_i)(y_{t-2} - y_{t-1}) - \cdots - \phi_p (y_{t-p} - y_{t-(p-1)}) + \epsilon_t$$

$$\Delta y_t = (\sum_{i=1}^{p} \phi_i - 1) y_{t-1} - (\sum_{i=2}^{p} \phi_i) \Delta y_{t-2} - \cdots - \phi_p \Delta y_{t-p} + \epsilon_t.$$

Hence to test for a unit root we check if $1 - \sum_{i=1}^{p} \phi_i = 0$. So the ADF-test statistic is given as:

$$\hat{t}_{\Phi=1} = \frac{\hat{\Phi} - 1}{SE(\hat{\Phi})},$$

where $\Phi = \sum_{i=1}^{p} \phi_i$, $\hat{\Phi}$ is the least-squares estimator and $SE(\hat{\Phi})$ is the standard error of the estimator. The ADF-test statistic is similar to the t-statistic, with the difference that under the null hypothesis $\hat{t}_{\Phi=1}$ follows the Dickey-Fuller distribution.

We note that the above two tests compliment to each other. Since they have alternating hypotheses, if the KPSS-test accepts the null hypothesis and the ADF-test rejects it, we conclude that the sample is trend-stationary. However if the KPSS-test rejects H_0 and the ADF-test accepts it, we conclude that the sample has a unite root hence is non-stationary. The existence of a trend component can be decided based on the original data plot and the ACF plot.

Returning to our main problem, if we find that the observed data $X = (x_1, ..., x_n)$ and $Y = (y_1, ..., y_m)$ are both non-stationary, we need to perform some transformations to get stationary samples. We start with X and apply a suitable transformation, say $t(\cdot)$, to obtain a stationary time series $t(X)$. Then we apply the same transformation to Y. If $t(Y)$ is a stationary series, we proceed to the second step of our procedure. However, if $t(Y)$ remains non-stationary then this is a clear indicator that the two samples have different models.

Chapter 2

Correlation Comparison

In this chapter we use the approach proposed by Quenouilles [16], which gives an efficient method to compare the correlation structures of two lengths of time series. This method is to be used on two different time series or different periods of the same time series. Given two sets of observations $x_1, ..., x_n$ and $y_1, ..., y_m$, we calculate their ACFs r_s and $r'_s,\ s = 1, ..., p$, respectively. Then we test the consistency of the two sets of estimated ACFs with a correlation structure that depends on the pooled data. A main assumption of this method is that $\sum_{i=q+1}^{\infty} \phi_i^2$ may be neglected for $q << (n + m)/p$: if the samples $x_1, ..., x_n$ have a true fit as:

$$x_t = \sum_{i=1}^{\infty} \phi_i \, x_{t-i} + \epsilon_t,$$

we estimate a fitted model as an $AR(p)$ model instead of the true model $AR(\infty)$.

Now let us describe the correlation comparison method. Assume that we have two time series observations $x_1, ..., x_n$ and $y_1, ..., y_m$.

(1) We start by pooling both observations and fitting them with an autoregressive model. For fitting the combined time series $x_1, ..., x_n, y_1, ..., y_m$ we use the Yule-Walker $AR(p)$ model:

Definition 2.1 *([4])*
The fitted Yule-Walker $AR(p)$ *sample model is:*
$X_t - \hat{\phi}_{p1} X_{t-1} - \cdots - \hat{\phi}_{pp} X_{t-p} = Z_t, \quad \{Z_t\} \sim WN(0, \hat{\sigma}_p),$
where

$\hat{\boldsymbol{\phi}}_p = (\hat{\phi}_{p1}, ..., \hat{\phi}_{pp})' = \hat{\boldsymbol{R}}_p^{-1}\hat{\boldsymbol{\rho}}_p,\ \hat{\sigma}_p = \hat{\gamma}(0)\left[1 - \hat{\boldsymbol{\rho}}_p'\hat{\boldsymbol{R}}_p^{-1}\hat{\boldsymbol{\rho}}_p\right],$
and
$\hat{\boldsymbol{\rho}}_p = (\hat{\rho}(1), ..., \hat{\rho}(p))' = \hat{\gamma}_p/\hat{\gamma}(0),\ \hat{\boldsymbol{R}}_p = [\hat{\rho}(i-j)]_{i,j=1}^p\,.$

(2) By using the fitted model, we calculate the partial ACFs (PACFs) v_s, v_s' of the residuals of each sample separately, and for any two samples we assume independence.

(3) Calculate $\chi^2_{(p)} = \sum_{s=1}^p (v_s - v_s')^2/(\frac{1}{n-s} + \frac{1}{m-s})$, and use the upper-tailed χ^2- test to accept or reject similar correlation structures.

This method is justified by the following fact. If $\{Z_t\} \sim iid(0, \sigma^2)$ in the Yule-Walker model, then the sample partial correlations are normally distributed and independent with $var(v_s) = 1/(n-s),\ 1 \le s \le p$. Hence the square of the difference of their distributions will follow a χ^2- distribution. This justifies the last step of the method that measures the goodness of fit, which has been verified by the works of Daniels et al. [5] and Jenkins et al. [8].

To further explain the method, we apply it to the following examples:

(1) $M_1:\ X_t = 0.1X_{t-1} + e_t, \quad e_t \sim N(0,1).$

(2) $M_2:\ X_t = 0.1X_{t-1} + w_t, \quad w_t \sim N(1,1).$

(3) $M_3:\ X_t = 0.5X_{t-1} + u_t, \quad u_t \sim N(1,1).$

(4) $M_4:\ X_t = 0.5X_{t-1} + o_t, \quad o_t \sim Exp(1).$

Each time series is of length 100 and we assume $X_0 = 0$. These models will be compared as follows:

(1) $M_1\ vs.\ M_2$ (2) $M_2\ vs.\ M_3$ (3) $M_2\ vs.\ M_4$ (4) $M_3\ vs.\ M_4$.

After generating the samples we start our illustration by combining both samples in each case and using Yule-Walker in R to find their fitted model $\{y_t\}$. In the following, $\{w_t\}$ denotes white noise.

- $M_1\ vs.\ M_2$:

$$w_t = y_t - 0.0566\ y_{t-1} - 0.0503\ y_{t-2} - 0.0735\ y_{t-3} - 0.0695\ y_{t-4} - 0.0728\ y_{t-5} - 0.0533\ y_{t-6} - 0.1412\ y_{t-7} - 0.1131\ y_{t-8} - 0.1253\ y_{t-9}.$$

- M_2 *vs.* M_3:

$$w_t = y_t - 0.1822\, y_{t-1}.$$

- M_2 *vs.* M_4:

$$w_t = y_t - 0.2424\, y_{t-1}.$$

- M_3 *vs.* M_4:

$$w_t = y_t - 0.368\, y_{t-1}.$$

Following the second step of the method, we calculate the PACFs of the residuals of each time series by virtue of the fitted model in each case. By using the obtained v_s, v'_s, we get $\chi^2_p = \sum_{i=1}^{p} (v_i - v'_i)^2 / (\frac{2}{100-i})$:

	$\chi^2_{(p)}$	$\chi^2_{0.05,\ p}$	**Result**
$\mathbf{M_1}$ **vs.** $\mathbf{M_2}$:	$\chi^2_{(9)} = 5.495457$	$\chi^2_{0.05,\ 9} = 16.29$	accept H_0
$\mathbf{M_2}$ **vs.** $\mathbf{M_3}$:	$\chi^2_{(1)} = 5.295579$	$\chi^2_{0.05,\ 1} = 3.84$	reject H_0
$\mathbf{M_2}$ **vs.** $\mathbf{M_4}$:	$\chi^2_{(1)} = 10.28117$	$\chi^2_{0.05,\ 1} = 3.84$	reject H_0
$\mathbf{M_3}$ **vs.** $\mathbf{M_4}$:	$\chi^2_{(1)} = 0.8266143$	$\chi^2_{0.05,\ 1} = 3.84$	accept H_0

Table 2.1: Correlation comparison: simulation study

Based on the values $\chi^2_{0.05,\ 9} = 16.29$ and $\chi^2_{0.05,\ 1} = 3.84$ and the upper-tailed χ^2-test, we get the results in Table 2.1. Here accepting H_0 means both samples share the same correlation structure and rejection means that they have different models. All results of the test are compatible with the original sample correlations, and this illustration shows that even for different noise distributions we can still obtain the expected results.

Chapter 3

Residual Comparison

If the two samples pass the first two steps, then we only need to compare the residual terms to test our main hypothesis. To test the distributions of the residuals, which are obtained from the fitted model via the correlation comparison test, we perform a two-sample test. Assume that $e_1, ..., e_n \sim F$ and $w_1, ..., w_m \sim G$, we want to test whether they have the same underlying distribution:

$$H_0 : \ F = G,$$
$$H_1 : \ F \neq G.$$

For this kind of problem there are a number of available nonparametric tests. In this thesis we will consider three tests, where each test has its own strength.

Considering the fact that we have one sample from each distribution and usually the residuals are iid or weakly dependent, we choose to use the Bayesian two-sample test [7]. The test depends mainly on the Bayes factor. Assume that the data follow one of the hypotheses H_0 and H_1. Based on the prior probabilities $Pr(H_0)$ and $Pr(H_1) = 1 - Pr(H_0)$, we get the posterior probabilities $Pr(H_0|data)$ and $Pr(H_1|data)$. Applying the Bayes rule, we get:

$$Pr(H_k|data) = \frac{Pr(data|H_k)Pr(H_k)}{Pr(data|H_0)Pr(H_0) + Pr(data|H_1)Pr(H_1)}, \ k = 0, 1.$$

Calculating the ratio of posterior of both hypotheses, we get:

$$\frac{Pr(H_0|data)}{Pr(H_1|data)} = \frac{Pr(data|H_0)}{Pr(data|H_1)} \cdot \frac{Pr(H_0)}{Pr(H_1)},$$

where the Bayes factor is:

$$B = \frac{Pr(data|H_0)}{Pr(data|H_1)}.$$

The Bayes factor calculates the probability of both hypotheses given the data. Depending on this ratio, we accept or reject the null hypothesis. The main issue with Bayesian test is to provide an explicit expression for the Bayes factor. In the parametric case, supposing our hypotheses are around a parameter θ_k, we use integration to calculate the density:

$$Pr(data|H_k) = \int Pr(data|\theta_k, H_k)\pi(\theta_k|H_k)d\theta_k.$$

However, in the nonparametric case this is more challenging since there is no exact parameter to integrate around. Hence, this mainly depends on the prior used for the data and its integration. The two-sample Bayesian test in the study by Holmes et al. [7] uses the Pólya tree distribution as its prior. The authors provided a closed formula of the Bayes factor.

Other tests can be used even if the observations are not iid, but they require a sample size greater than 3 for it to perform well with a level of significance $\alpha = 0.05$. Considering the time series samples as high dimensional vectors, where the number of time points indecates the dimension of the vector, we will deal with the two-sample problem using the Biswas and Ghosh test (BG-test) [2]. This test is an improved version of the Baringhaus and Franz test (BF-test) [1]. The BF-test is based on the inter-point distance. It assumes that the observations are independent and preforms very well in location problems, but has a fairly poor performance in high dimensional low sample size settings (HDLSS) when dealing with a scale or location-scale problem. Based on the BF-test, Biswas and Ghosh [2] proposed a new inter-point test that overcomes this problem. To study the behavior of their test in high dimension, they made some assumptions which ensure that the test can be used for weakly dependent samples, hence it could be applied to the time series samples directly. The BG-test assumes that samples have different means and variances. So it fails in cases of different distributions that share the same mean and variance such as the $N(1, 1)$ and $Exp(1)$. We will propose an adjustment to the existing BG-test that allows it to apply

to more general situations.

Although the two-sample Bayesian test is a $1-$dimensional test, it has some advantages over the previously mentioned nonparametric tests. It performs well in all mentioned problems, including the case of equal means and variances. Choosing which test is more suitable depends on the samples we want to test. Since in the correlation structure chapter we chose the Yule-Walker method to model our observations, we can say that the residuals are white noise (might not iid). To assume that the residuals are iid we need to consider how they will be used, their relation to other information and the how the data was collected.

If the residuals are assumed to be iid, it is sufficient to use the BG-test if our sample sizes n, m are at least 4 for a 0.05 level of significance. We can also apply the two-sample Bayesian test for any size of samples, since the residuals are iid we can pool the residuals of each group of time series and test them as one sample from each group. However, if the residuals are only white noise (WN), the only test applicable will be the BG-test, and in this case we need a sample size that is 4 or greater to get more accurate results. If samples pass the BG-test in the WN case, this implies that the residuals have the same distribution. But since we have no information on the variances or means of the residual samples, this could lead to a false conclusion. Hence we need a modified method to get a more precise result. In the end of this thesis, we will introduce the so called BG$^3-$test.

3.1 Two-sample Bayesian test

Holmes et al. [7] proposed a Bayesian approach to the two-sample problem. Given two sets of samples $\boldsymbol{y}^{(1)} \sim^{iid} F^{(1)}$ and $\boldsymbol{y}^{(2)} \sim^{iid} F^{(2)}$, the Bayesian test depends on evaluating the evidence of the null hypothesis against the alternative, where $F^{(1)}$ and $F^{(2)}$ are unknown distributions.

Since the underlying distributions of the data are unknown, to apply the Bayesian approach, we need to define priors over the data. Holmes et al. [7] used Pólya tree priors (PT priors):

$$\text{under } H_0 : F^{(1)} = F^{(2)},$$
$$\text{under } H_1 : F^{(1)} \neq F^{(2)},$$

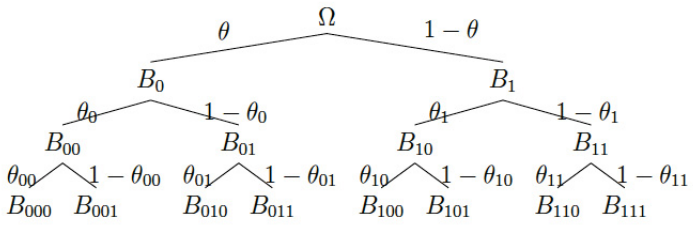


Figure 3.1: A Pólya tree distribution till level $k = 3$

where $F^{(1)}$ and $F^{(2)}$ are independent draws from the PT prior, and the test evaluates the Bayes factor for the two models.

Pólya tree priors are a class of distributions for random probability measures that partitions the domain Ω into disjoint binary partitions as seen in Figure 3.1. At level k we have:

$$\begin{aligned}
&\{B_j^{(k)}, j = 0, ..., 2^k - 1\}, \\
&B_i^{(k)} \cap B_j^{(k)} = \emptyset, \quad i \neq j, \\
&B_j^{(k)} = B_{2j}^{(k+1)} \cup B_{2j+1}^{(k+1)}, \quad k = 1, 2, ..., \ j = 0, ..., 2^k - 1.
\end{aligned}$$

By constructing a random measure on the sets B_j of the tree, a random measure is defined on Ω, where :

$$\begin{aligned}
\theta_j &: \text{probability of going left}, \\
1 - \theta_j &: \text{probability of going right},
\end{aligned}$$

where θ_j is a random variable with the distribution $\theta_j \sim \pi_j$ and the partitions are indexed as:

- $\boldsymbol{\epsilon}_k = \{\epsilon_{k1}, ..., \epsilon_{kk}\}$: sample path of an element till level (k).
- $\epsilon_{ki} \in \{0, 1\}$: indicates that at level (i) we either go left (0) or right (1).
- $B_{\boldsymbol{\epsilon}_k}$: the partition where the particle at level (k) falls in.
- $\theta_{\boldsymbol{\epsilon}_m} \sim Beta\,(\alpha_{\boldsymbol{\epsilon}_m 0}, \alpha_{\boldsymbol{\epsilon}_m 1})$.

From the PT Figure, we can see that the probability of the particle falling into $B_{\boldsymbol{\epsilon}_k}$ is:

$$Pr\left(B_{\boldsymbol{\epsilon}_k}\right) = \prod_{i=1}^{k} \left(\theta_{\boldsymbol{\epsilon}_i - 1}\right)^{1-\epsilon_{ii}} \left(1 - \theta_{\boldsymbol{\epsilon}_i - 1}\right)^{\epsilon_{ii}}.$$

The Pólya tree distribution was constructed by Lavine [11], where he gave the following definition:

Definition 3.1 *([11])*
Let $\Pi = \{B_0, B_1, B_{00}, ...\}$ *and* $\mathcal{A} = (\alpha_0, \alpha_1, \alpha_{00}, ...)$. *A random probability measure* F *on* Ω *is said to have a Pólya tree distribution, or Pólya tree prior, with parameters* $(\Pi, \mathcal{A})$, *written* $F \sim PT\left(\Pi, \mathcal{A}\right)$, *if:*
$\exists$ *non-negative numbers* $\mathcal{A} = (\alpha_0, \alpha_1, \alpha_{00}, ...)$ *and random variables* $\Theta = (\theta, \theta_0, \theta_1, \theta_{00}, ...)$ *such that:*

(1) θ_j *in* Θ *are mutually independent.*

(2) $\forall k = 1, 2, ...$ *and every* $\boldsymbol{\epsilon}_k \in \{0,1\}^k$, $\quad \theta_{\boldsymbol{\epsilon}_k} \sim Beta\left(\alpha_{\boldsymbol{\epsilon}_k 0}, \alpha_{\boldsymbol{\epsilon}_k 1}\right).$

(3) $\forall k = 1, 2, ...$ *and every* $\boldsymbol{\epsilon}_k \in \{0,1\}^k$,

$$F\left(B_{\boldsymbol{\epsilon}_k} | \Theta\right) = \prod_{i=1}^{k} \left(\theta_{\boldsymbol{\epsilon}_i - 1}\right)^{1-\epsilon_{ii}} \left(1 - \theta_{\boldsymbol{\epsilon}_i - 1}\right)^{\epsilon_{ii}}. \tag{3}$$

Hence by sampling the θ_j's in Θ, we obtain a realization of $F \sim PT(\Pi, \mathcal{A})$. The Pólya tree can be centered on some distribution G such that $E\left[F\right] = G$. Lavine [11] defined this probability measure using equation (3) as follow:

$$\begin{aligned} G(B_{\boldsymbol{\epsilon}_k}) =& E\left[\prod_{i=1}^{k} \left(\theta_{\boldsymbol{\epsilon}_i - 1}\right)^{1-\epsilon_{ii}} \left(1 - \theta_{\boldsymbol{\epsilon}_i - 1}\right)^{\epsilon_{ii}}\right] \\ =& \prod_{i=1}^{k} E\left[\left(\theta_{\boldsymbol{\epsilon}_i - 1}\right)^{1-\epsilon_{ii}}\right] E\left[\left(1 - \theta_{\boldsymbol{\epsilon}_i - 1}\right)^{\epsilon_{ii}}\right] \\ =& \prod_{i=1}^{k} \frac{\alpha_{\boldsymbol{\epsilon}_k 0}}{\alpha_{\boldsymbol{\epsilon}_k 0} + \alpha_{\boldsymbol{\epsilon}_k 1}} \frac{\alpha_{\boldsymbol{\epsilon}_k 1}}{\boldsymbol{\epsilon}_k 0 + \alpha_{\boldsymbol{\epsilon}_k 1}} \end{aligned}$$

for any measurable set $B_{\boldsymbol{\epsilon}_k}$.

The first attraction of using the PT priors is that it can be centered on any distribution G, this is accomplished by setting the partitioning subsets as the dyadic quantiles of G, and setting $\alpha_{\boldsymbol{\epsilon}_k 0} = \alpha_{\boldsymbol{\epsilon}_k 1}$. So at level 1 we will have two partitions satisfying $G(B_0) = G(B_1) = 1/2$, and for any $\boldsymbol{\epsilon}_k \in \{0,1\}^k$ the partitions satisfy $G(B_{\boldsymbol{\epsilon}_k 0}|B_{\boldsymbol{\epsilon}_k}) = G(B_{\boldsymbol{\epsilon}_k 1}|B_{\boldsymbol{\epsilon}_k}) = 1/2$. If $\Omega \equiv \mathbb{R}$ ([11]):

$$\begin{aligned}
B_0 &= \left(-\infty, G^{-1}(0.5)\right), \\
B_1 &= \left[G^{-1}(0.5), \infty\right), \\
&\vdots \\
B_{\boldsymbol{\epsilon}_k} &= \left[G^{-1}\{(k^*-1)/2^k, G^{-1}\left(k^*/2^k\right)\right), \\
k &= \{1,2,...\},\ k^* = \{1,...,2^k\}.
\end{aligned}$$

The α's in this test are chosen as constants in each level, i.e. $\alpha_{\boldsymbol{\epsilon}_m 0} = \alpha_{\boldsymbol{\epsilon}_m 1} = c_m$, and $c_m = c/2^m,\ c > 0$ to imply that F is absolutely continuous with probability 1.

The second attraction of using the PT prior is that it has the conjugate property, meaning that given the prior: $F \sim PT(\Pi, \mathcal{A})$ and data $\boldsymbol{y} \sim F$, the posterior will also have a Pólya tree distribution: $F|\boldsymbol{y} \sim PT(\Pi, \mathcal{A}^*)$ with the set of updated parameters $\mathcal{A}^* = \{\alpha^*_{00}, \alpha^*_{01}, \alpha^*_{000}, ...\}$ and $\alpha^*_{\boldsymbol{\epsilon}_i}|\boldsymbol{y} = \alpha_{\boldsymbol{\epsilon}_i} + n_{\boldsymbol{\epsilon}_i}$, where $n_{\boldsymbol{\epsilon}_i}$ is the number of observations in $\boldsymbol{y}$ that belong to $B_{\boldsymbol{\epsilon}_i}$. Then:

$$\theta^*_j|\boldsymbol{y} = Beta(\alpha_{j0} + n_{j0}, \alpha_{j1} + n_{j1}),$$

where n_{j0} and n_{j1} are the numbers of observations going left and right respectively at every knot j in the tree.

The calculation of the marginal likelihood of any set of observations depends on the n_{j0} and n_{j1}:

$$Pr(\boldsymbol{y}|\Theta, \boldsymbol{\Pi}, \mathcal{A}) = \textstyle\prod_j \theta_j^{n_{j0}}(1-\theta_j)^{n_{j1}},$$

where $\theta_j|\mathcal{A} \sim Beta(\alpha_{j0}, \alpha_{j1})$, and $j \in \{0, 1, 00, ...\}$. By integrating out Θ we get the

following marginal likelihood:

$$\begin{aligned} Pr\left(\boldsymbol{y}|\boldsymbol{\Pi},\mathcal{A}\right) &= \int_{\theta_j} Pr\left(\boldsymbol{y}|\Theta,\boldsymbol{\Pi},\mathcal{A}\right)\cdot Pr\left(\theta_j|\boldsymbol{\Pi},\mathcal{A}\right)d(\theta_j) \\ &= \int_{\theta_j}\prod_j \theta_j^{n_{j0}}(1-\theta_j)^{n_{j1}}.\frac{(\theta_j)^{(\alpha_{j0}-1)}(1-\theta_j)^{(\alpha_{j1}-1)}}{\mathbf{B}(\alpha_{j0},\alpha_{j1})}d(\theta_j) \\ &= \prod_j \frac{1}{\mathbf{B}(\alpha_{j0},\alpha_{j1})}\int_0^1 \theta_j^{\alpha_{j0}+n_{j0}-1}(1-\theta_j)^{\alpha_{j1}+n_{j1}-1}d(\theta_j) \\ &= \prod_j \frac{\mathbf{B}(\alpha_{j0}+n_{j0},\alpha_{j1}+n_{j1})}{\mathbf{B}(\alpha_{j0},\alpha_{j1})}. \end{aligned}$$

Hence:

$$Pr\left(\boldsymbol{y}|\boldsymbol{\Pi},\mathcal{A}\right) = \prod_j \left(\frac{\Gamma(\alpha_{j0}+\alpha_{j1})}{\Gamma(\alpha_{j0})\Gamma(\alpha_{j1})}\frac{\Gamma(\alpha_{j0}+n_{j0})\Gamma(\alpha_{j1}+n_{j1})}{\Gamma(\alpha_{j0}+n_{j0}+\alpha_{j1}+n_{j1})}\right). \tag{4}$$

Based on the equation (4), we calculate the Bayes factor of our problem. To provide the weight of evidence in favor of H_0 given the data using PT prior we assume:

$$\begin{aligned} F^{(1)}, F^{(2)} &\sim^{iid} PT\left(\Pi,\mathcal{A}\right), \\ \alpha_{j0} &= \alpha_{j1} = cm^2. \end{aligned}$$

From the Bayes theorem:

$$\begin{aligned} Pr\left(H_0|\boldsymbol{y}^{(1,2)}\right) &\propto Pr\left(\boldsymbol{y}^{(1,2)}|H_0\right)Pr\left(H_0\right), \\ \frac{Pr\left(H_0|\boldsymbol{y}^{(1,2)}\right)}{Pr\left(H_1|\boldsymbol{y}^{(1)},\boldsymbol{y}^{(2)}\right)} &= \frac{Pr\left(\boldsymbol{y}^{(1,2)}|H_0\right)}{Pr\left(\boldsymbol{y}^{(1)},\boldsymbol{y}^{(2)}|H_1\right)}\frac{Pr\left(H_0\right)}{Pr\left(H_1\right)}, \end{aligned}$$

where $\boldsymbol{y}^{(1,2)} = \{\boldsymbol{y}^{(1)},\ \boldsymbol{y}^{(2)}\}$.
Using $Pr\left(\boldsymbol{y}|\boldsymbol{\Pi},\mathcal{A}\right)$ in (4):

$$\begin{aligned} Pr\left(\boldsymbol{y}^{(1,2)}|H_0\right) &: \theta_j^*|\boldsymbol{y}^{(1,2)} \sim Beta(\alpha_{j0}+n_{j0}^{(1)}+n_{j0}^{(2)},\alpha_{j1}+n_{j1}^{(1)}+n_{j1}^{(2)}), \\ Pr\left(\boldsymbol{y}^{(1)},\boldsymbol{y}^{(2)}|H_1\right) &: \theta_j^{*(1)}|\boldsymbol{y}^{(1)} \sim Beta(\alpha_{j0}+n_{j0}^{(1)},\alpha_{j1}+n_{j1}^{(1)}) \\ & \quad\ \theta_j^{*(2)}|\boldsymbol{y}^{(2)} \sim Beta(\alpha_{j0}+n_{j0}^{(2)},\alpha_{j1}+n_{j1}^{(2)}). \end{aligned}$$

Hence the Bayes factor is given ([7]):

$$\frac{Pr\left(\boldsymbol{y}^{(1,2)}|H_0\right)}{Pr\left(\boldsymbol{y}^{(1)},\boldsymbol{y}^{(2)}|H_1\right)}=\prod_j b_j,$$

where:

$$b_j=\frac{\Gamma(\alpha_{j0})\Gamma(\alpha_{j1})}{\Gamma(\alpha_{j0}+\alpha_{j1})}\frac{\Gamma(\alpha_{j0}+n_{j0}^{(1)}+n_{j0}^{(2)})\Gamma(\alpha_{j1}+n_{j1}^{(1)}+n_{j1}^{(2)})}{\Gamma(\alpha_{j0}+n_{j0}^{(1)}+n_{j0}^{(2)}+\alpha_{j1}+n_{j1}^{(1)}+n_{j1}^{(2)})}$$
$$\times\frac{\Gamma(\alpha_{j0}+n_{j0}^{(1)}+\alpha_{j1}+n_{j1}^{(1)})}{\Gamma(\alpha_{j0}+n_{j0}^{(1)})\Gamma(\alpha_{j1}+n_{j1}^{(1)})}\frac{\Gamma(\alpha_{j0}+n_{j0}^{(2)}+\alpha_{j1}+n_{j1}^{(2)})}{\Gamma(\alpha_{j0}+n_{j0}^{(2)})\Gamma(\alpha_{j1}+n_{j1}^{(2)})},$$
$$j\in\{\emptyset,0,1,00,...\}.$$

The Bayesian method requires us to specify $\{\Pi,\mathcal{A}\}$ for our PT prior. The default setting for $\mathcal{A}$: $\alpha_{j0}=\alpha_{j1}=cm^2$, and for Π: start by standardizing the joint data $\boldsymbol{y}^{(1,2)}$, then set $\Pi=\Phi(.)^{-1}$, i.e. the partitions are defined on the quantiles of a standard normal density.

Th proofs of the consistency of the truncation of the Bayes factor are given in the study by Holmes et al. [7]. It is more challenging to prove the consistency in the non-truncated case, but the paper provides a number of simulations that suggest consistency for the non-truncated test ([7]). Following these notations a definition of the truncated Bayes factor is given:

- $n=n_{\emptyset}^{(1)}+n_{\emptyset}^{(2)}$: total sample size,
- $2^{l(\boldsymbol{\epsilon})}$: number of partitions at level $l(\boldsymbol{\epsilon})$.

Truncated Bayes factor:

$$BF_{k_0}=\prod_{\{j|l(j)\leq k_0\}} b_j,$$

where the level of truncation $k_0\in\mathbb{N}$. We also define a truncated Hypothesis test:

$$H_{0,k_0}:\forall\boldsymbol{\epsilon}|l(\boldsymbol{\epsilon})\leq k_0,\quad F^{(1)}\left(B_{\boldsymbol{\epsilon}}\right)=F^{(2)}\left(B_{\boldsymbol{\epsilon}}\right),$$
$$vs.$$
$$H_{1,k_0}:\exists\boldsymbol{\epsilon}|l(\boldsymbol{\epsilon})\leq k_0,\quad F^{(1)}\left(B_{\boldsymbol{\epsilon}}\right)\neq F^{(2)}\left(B_{\boldsymbol{\epsilon}}\right).$$

3.2 Other two-sample tests

The tests we consider here perform very well when dealing with high-dimensional observations with sample sizes that are greater than 3. Considering the time series samples as high dimensional vectors, where the number of time points indicates the dimension of the vector, we will deal with the two-sample problem using the BG-test ([2]).

When using nonparametric tests in this thesis, we wish to test the underlying distribution that the data was drawn from, rather than the parameters of the distribution. Since we have no knowledge of the shape or form of the distribution, rather than assuming it is normal and testing its parameters we test the distribution as a whole without giving these assumptions. Hence this approach is suitable for any kind or size of samples.

3.2.1 BF-test and BG-test

- **BF-test (2004):**

  Baringhaus and Franz [1] introduced a multivariate two-sample test based on interpoint distance between variables from both distributions where $\mathbf{X}, \mathbf{X}^* \sim^{iid} F$, $\mathbf{Y}, \mathbf{Y}^* \sim^{iid} G$ and $\|\cdot\|$ is the Euclidean norm. This measures the difference between both distributions F and G. Using the inequality $2E\|\mathbf{X}-\mathbf{Y}\|-E\|\mathbf{X}-\mathbf{X}^*\|-E\|\mathbf{Y}-\mathbf{Y}^*\| \geq 0$, where it equals 0 if and only if $F = G$, and the empirical representation of $E\|\cdot\|$, the test is given as follows:

$$T^{BF}_{m,n} = \frac{mn}{m+n}\left[\frac{1}{mn}\sum_{j=1}^{m}\sum_{k=1}^{n}\|\mathbf{X_j}-\mathbf{Y_k}\| - \frac{1}{2m^2}\sum_{j=1}^{m}\sum_{k=1}^{m}\|\mathbf{X_j}-\mathbf{X_k}\| - \frac{1}{2n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\|\mathbf{Y_j}-\mathbf{Y_k}\|\right].$$

  The null hypothesis will be rejected for large values of the statistic and accepted otherwise. Based on simulations, Baringhaus and Franz concluded that this test reaches high power similar to the parametric Hotelling's T^2-test in normal location problems, and is sensitive to location changes.

- **BG-test (2013):**

  Since the BF-test was studied for the low-dimension setting, Biswas and Ghosh proposed a test that is applicable in high-dimension low sample size (HDLSS) settings.

In their illustrative example they showed where the BF-test and other nonparametric test show low performance in HDLSS setting, with examples covering the cases of location, scale and location-scale problems. From the illustration they concluded that the BF-test does not perform well in location-scale problems and performs poorly in scale problems. Considering the cases where the BF-test fails, Biswas and Ghosh [2] identified the problem and proposed the following test:
Let

$$\mathbf{X}, \mathbf{X}^* \sim^{iid} F,\ \mathbf{Y}, \mathbf{Y}^* \sim^{iid} G,$$

where $\mathbf{X}, \mathbf{X}^*, \mathbf{Y},$ and $\mathbf{Y}^*$ are independent random vectors, and:

$$\|\mathbf{X} - \mathbf{X}^*\| \sim D_{FF},\ \|\mathbf{Y} - \mathbf{Y}^*\| \sim D_{GG},$$
$$\text{and } \|\mathbf{X} - \mathbf{Y}\| \sim D_{FG},$$

with respective means: μ_{FF}, μ_{FG} and μ_{GG}, define the bivariate distribution:

$$(\|\mathbf{X} - \mathbf{X}^*\|, \|\mathbf{X} - \mathbf{Y}\|) \sim D_F,$$
$$(\|\mathbf{Y} - \mathbf{X}\|, \|\mathbf{Y} - \mathbf{Y}^*\|) \sim D_G.$$

We can rewrite the hypotheses in the following way:

$$\begin{array}{ll} H_0 : F = G & \\ H_1 : F \neq G & \end{array} \iff \begin{array}{ll} H_0^{"} : \boldsymbol{\mu}_{D_F} = \boldsymbol{\mu}_{D_G} & \\ H_1^{"} : \boldsymbol{\mu}_{D_F} \neq \boldsymbol{\mu}_{D_G} & \end{array},$$

since:

$$F = G \Leftrightarrow \boldsymbol{\mu}_{D_F} = \boldsymbol{\mu}_{D_G} \Leftrightarrow \mu_{FF} = \mu_{FG} = \mu_{GG}.$$

The BG-test statistic is:

$$T_{m,n}^{BG} = \|\hat{\boldsymbol{\mu}}_{D_F} - \hat{\boldsymbol{\mu}}_{D_G}\|^2,$$

where for the data: $\boldsymbol{x}_1, ..., \boldsymbol{x}_m \sim^{iid} F$ and $\boldsymbol{y}_1, ..., \boldsymbol{y}_n \sim^{iid} G$

$$\hat{\boldsymbol{\mu}}_{D_F} = \Bigg[\hat{\mu}_{FF} = \binom{m}{2}^{-1} \sum_{i=1}^{m} \sum_{j=i+1}^{m} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|,$$
$$\hat{\mu}_{FG} = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{y}_j\| \Bigg],$$
$$\hat{\boldsymbol{\mu}}_{D_G} = \Bigg[\hat{\mu}_{FG} = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{y}_j\|,$$
$$\hat{\mu}_{GG} = \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \|\boldsymbol{y}_i - \boldsymbol{y}_j\| \Bigg].$$

Similar to the BF-test, we reject the null hypothesis for large values of the test. By running the same HDLSS example on the BG-test, the test shows excellent performance in the cases where the BF-test examines poor performance.

3.2.2 BG^3-test

From the assumptions of the BG-test, we can see that the test is not applicable to the case that the distributions have the same mean and variance. To overcome this problem, we suggest to use the skewness of the distributions (third moments) to differentiate between them. Hence, in the case of equal means and variances and different third moments we replace x and y in the BG-test with x^3 and y^3, respectively.

- **BG^3-test:**

$$T_{m,n}^{BG^3} = \|\hat{\boldsymbol{\mu}}_{D_F}^{(3)} - \hat{\boldsymbol{\mu}}_{D_G}^{(3)}\|^2,$$

where for the data: $\boldsymbol{x}_1, ..., \boldsymbol{x}_m \sim^{iid} F$ and $\boldsymbol{y}_1, ..., \boldsymbol{y}_n \sim^{iid} G$,

$$\hat{\boldsymbol{\mu}}^{(3)}_{D_F} = \left[\hat{\mu}^{(3)}_{FF} = \binom{m}{2}^{-1} \sum_{i=1}^{m} \sum_{j=i+1}^{m} \|\boldsymbol{x}_i^3 - \boldsymbol{x}_j^3\|, \right.$$
$$\left. \hat{\mu}^{(3)}_{FG} = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \|\boldsymbol{x}_i^3 - \boldsymbol{y}_j^3\| \right],$$
$$\hat{\boldsymbol{\mu}}^{(3)}_{D_G} = \left[\hat{\mu}^{(3)}_{FG} = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \|\boldsymbol{x}_i^3 - \boldsymbol{y}_j^3\|, \right.$$
$$\left. \hat{\mu}^{(3)}_{GG} = \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \|\boldsymbol{y}_i^3 - \boldsymbol{y}_j^3\| \right].$$

Biswas and Ghosh [2] showed the limiting behavior of their proposed test by studying its performance as the dimensions increase to infinity and the sample sizes are fixed. Based on the assumptions they made for their study, which are from the assumptions made by Hall et al. [6], we propose the following:
Assuming that the samples $\boldsymbol{x}_i^3 = (x_1^3, ..., x_d^3) \sim^{iid} \bar{F},\ i = 1, ..., m$ and $\boldsymbol{y}_j^3 = (y_1^3, ..., y_d^3) \sim^{iid} \bar{G},\ j = 1, ..., n$ have fixed sample sizes m and n, $d \to \infty$, let:
$\bar{\boldsymbol{\mu}}_1$ and $\bar{\boldsymbol{\Sigma}}_1$: mean vector and dispersion matrix of $\bar{F}$.
$\bar{\boldsymbol{\mu}}_2$ and $\bar{\boldsymbol{\Sigma}}_2$: mean vector and dispersion matrix of $\bar{G}$.
Hall et al. [6] gave the following assumptions, where they considered a $d-$dimensional observation $\boldsymbol{X} = \left(X_{(1)}, ..., X_{(d)}\right)$ as a truncation of a finite time series $\boldsymbol{X} = \left(X_{(1)}, ...\right)$:

A1 $\exists\, \omega_1^2, \omega_2^2 > 0$ and u, such that:

(a) $trace\left(\bar{\boldsymbol{\Sigma}}_1\right)/d \to \omega_1^2$.

(b) $trace\left(\bar{\boldsymbol{\Sigma}}_2\right)/d \to \omega_2^2$.

(c) $\|\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2\|^2/d \to u^2$ as $d \to \infty$..

A2 Fourth moments of $\boldsymbol{X}^3$ and $\boldsymbol{Y}^3$ are uniformly bounded.

A3 Under some permutation:
for $\left(U_{(q)}^3, V_{(q)}^3\right) = \left(X_{(q)}^3, X_{*(q)}^3\right), \left(X_{(q)}^3, Y_{(q)}^3\right), \left(Y_{(q)}^3, Y_{*(q)}^3\right)$.
The sequence:

$$\left\{ \left(U_{(q)}^3 - V_{(q)}^3\right)^2,\ q \geq 1 \right\} \text{ is } \rho \text{ mixing}$$

meaning:

$$sup_{1 \leq q \leq q' \leq \infty, |q-q'|>r} |corr\left\{ \left(U_{(q)}^3 - V_{(q)}^3\right)^2, \left(U_{(q')}^3 - V_{(q')}^3\right)^2 \right\}| \leq \rho(r)$$

where: $\rho(r) \to 0$ as $r \to \infty$.

Proposition 3.1 *Suppose that we have m independent observations from each of $\bar{F}$ and $\bar{G}$, which satisfy A1-A3. Also assume that either $u^2 > 0$ or $\omega_1 \neq \omega_2$. Then, unless m is very small (i.e., $\binom{2m}{m} \leq 2/\alpha$), the power of the proposed test of level α converges to 1 as d tends to infinity.*

Similar to the BG-test, we reject the null hypothesis for large values of the test, and for a test of level of significance of 0.05 this test also requires a sample size greater than or equal 4. By combining the BG-test and the BG$^3-$test, we can deal with the case that the two samples have the same mean and variance but with different skewness. This point will be further explained in the simulation study in the next chapter.

Chapter 4

Case Study

In this chapter, we illustrate our procedure based on a case study. We will also show the performance of the BG^3-test based on simulation.

4.1 Analysis of real data

We apply our proposed procedure to four real data sets, which are from the national long-term water quality monitoring data. The data sets provide one time series sample from each of the seven sites monitored. We will compare the levels of zinc, lead, PH and alkalinity measured in Microgram/Liter (Ug/L), PH and Milligram/Liter (MG/L) in six of these sites on Yukon river basin, which is the third longest river in North America. Here we ignore the seventh site since it has only 22 data points. The quality monitoring of the river was done by Environment and Climate Change Canada (ECCC) from year 2005 till 2015, and the measures were taken monthly. The data set includes measurements of PH, major ions and metals taken over the mentioned period.

In this thesis, we will work with two metals, zinc and lead. We study the differences between the models of the observations in each site. We will also consider the PH and alkalinity measures. PH is the intensity or level of how acidic or basic the water is. It has a range between 0 and 14, where 0 is acidic, 14 is basic and 7 being neutral. Alkalinity is closely related to PH, where it measures the ability of water to resist changes in PH ([23]). There are a number of species that form alkalinity, however in this data they only measure the levels of bicarbonate HCO_3^-.

First, we consider the zinc data set. We denote the samples from site 1 till site 6 as $s_1, ..., s_6$. In the following, we will use the same site notations in all cases. The locations of

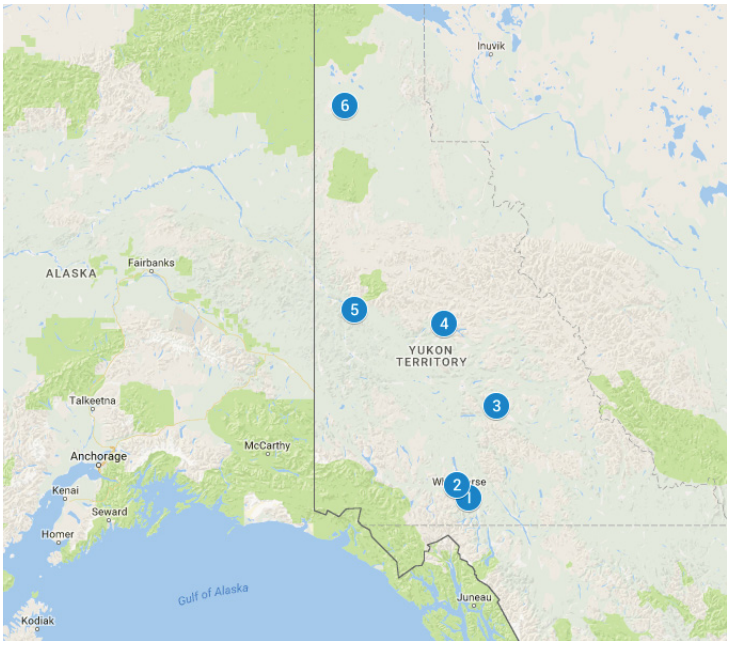
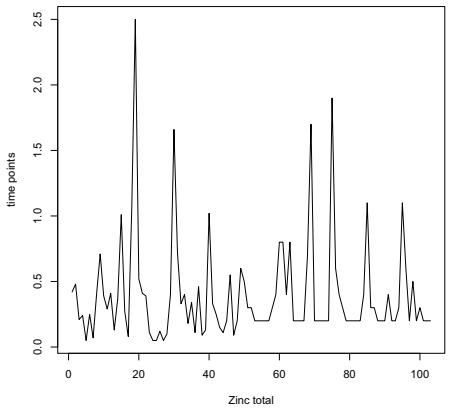


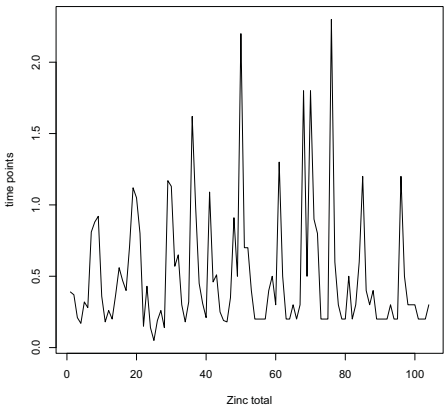
Figure 4.1: Locations of the sites of the samples

the sites are shown in Figure 4.1. The samples have $103, 104, 128, 102, 99, 75$ time points respectively, which are given in Figure 4.2. Note that although the plots show the difference in the range of values and hence have different average values, this does not mean that they do not share the same structure.

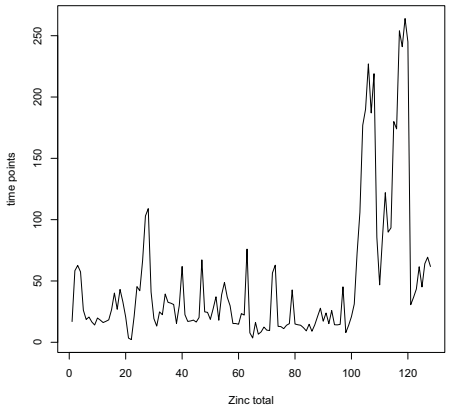
We test the stationarity of the time series of each sample using the package `'tseries'` in R for the KPSS-test [10], and the ADF-test [17]. For the KPSS-test the null hypothesis is stationarity, so for a P-value < 0.05 we reject H_0, however for the ADF-test the null hypothesis is non-stationarity, so when the P-value < 0.05 the time series is trend-stationary. In this set of data we see that the plots do not show any trend in s_1, s_2, s_4 and s_5, see Figure 4.2 and Figure 4.3. Hence, trend stationarity tests are considered as stationarity tests in this case and we will use the KPSS-test for level-stationarity. However for s_3 and s_6 we will test trend-stationarity. The results of the tests are given in Table 4.1.



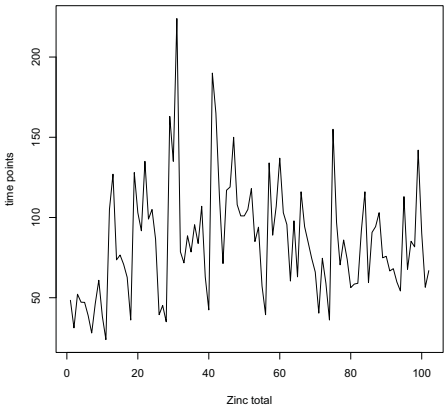
(a) Site 1



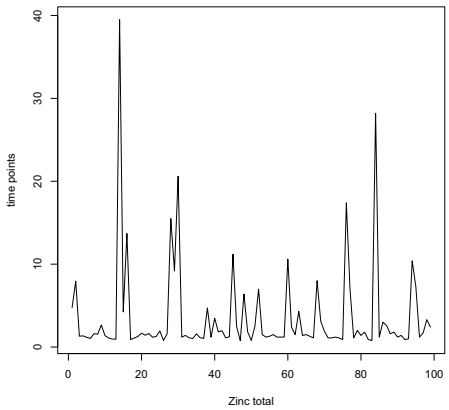
(b) Site 2



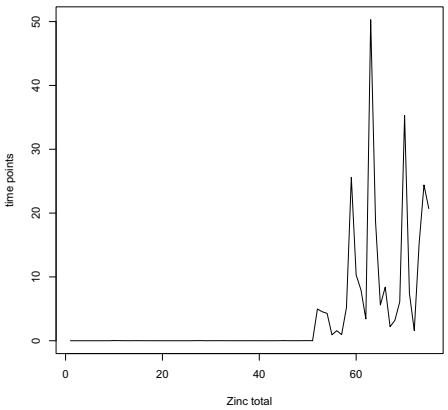
(c) Site 3



(d) Site 4

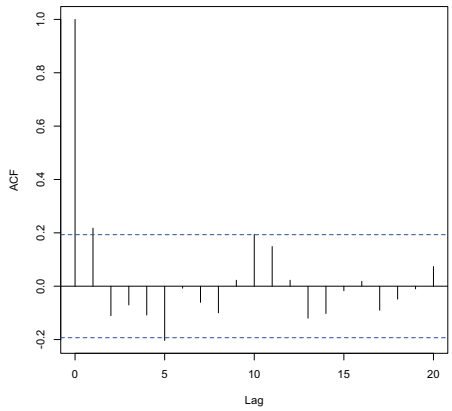


(e) Site 5

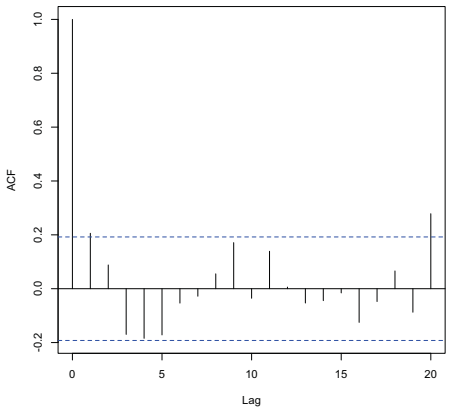


(f) Site 6

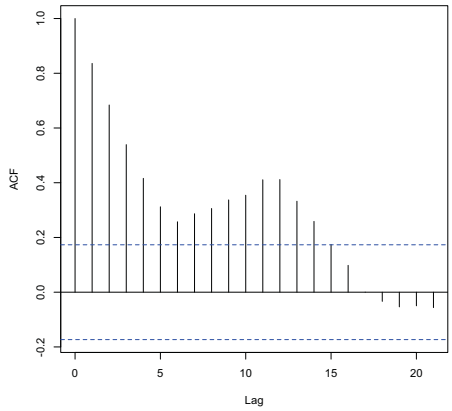
Figure 4.2: Time series plots for zinc levels



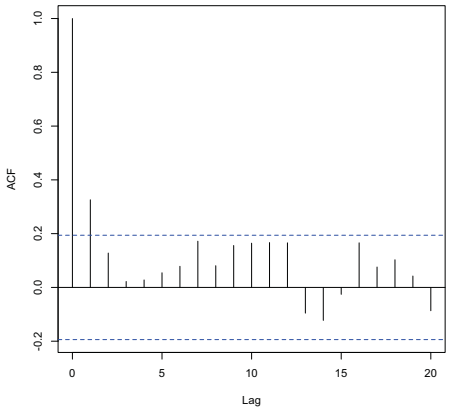
(a) Site 1



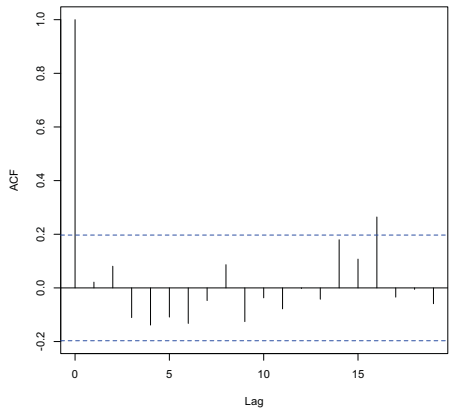
(b) Site 2



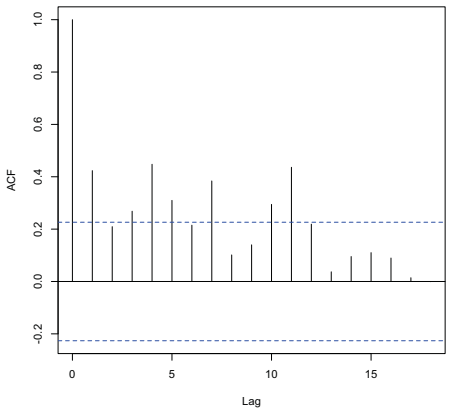
(c) Site 3



(d) Site 4



(e) Site 5



(f) Site 6

Figure 4.3: ACF plots for zinc levels

sample	Test		stationary:
	KPSS	**ADF**	**non-stationary**
s_1	0.10000000	0.01000000	2:0
s_2	0.10000000	0.01000000	2:0
s_3	0.01000000	0.06636761	0:2
s_4	0.1000000000	0.0159199336	2:0
s_5	0.1000000	0.0100000	2:0
s_6	0.0100000000	0.3738797188	0:2

Table 4.1: P-values and stationarity of zinc samples

Based on these results, we will continue this study in two groups. The first group contains the samples that are stationary depending on the stationary over non-stationary results given in Table 4.1, which are s_1, s_2, s_4 and s_5. The rest will form the second group. From this step we know that the models of the samples in the first group are different from the models of the samples in the second group. Considering that the time series of the first group are stationary, we continue the study by following the second step of our procedure, which is comparing their correlation structures. We pair the samples as follows:

$$\begin{array}{lll} \text{(1) } s_1 \text{ vs. } s_2, & \text{(2) } s_1 \text{ vs. } s_4, & \text{(3) } s_1 \text{ vs. } s_5, \\ \text{(4) } s_2 \text{ vs. } s_4, & \text{(5) } s_2 \text{ vs. } s_5, & \text{(6) } s_4 \text{ vs. } s_5. \end{array}$$

For each pair we gather both samples and fit them to an $AR(p)$ model by Yule-walker using R. We obtain the following models:

(1) s_1 vs. s_2:

$$w_t = y_t - 0.2245\, y_{t-1}.$$

(2) s_1 vs. s_4:

$$\begin{aligned} w_t = y_t &- 0.3858\, y_{t-1} - 0.1475\, y_{t-2} - 0.0373\, y_{t-3} - 0.0059\, y_{t-4} - 0.0348\, y_{t-5} \\ &- 0.0734\, y_{t-6} - 0.2048\, y_{t-7} + 0.0437\, y_{t-8} - 0.0683\, y_{t-9} - 0.0887\, y_{t-10} \\ &- 0.1027\, y_{t-11} - 0.1430\, y_{t-12} + 0.2454\, y_{t-13} + 0.1255\, y_{t-14} - 0.0262\, y_{t-15} \\ &- 0.1741\, y_{t-16} + 0.1202\, y_{t-17}. \end{aligned}$$

(3) s_1 vs. s_5:

$$\begin{aligned} w_t = y_t &- 0.0637\, y_{t-1} - 0.1148\, y_{t-2} + 0.0475\, y_{t-3} + 0.0620\, y_{t-4} - 0.0131\, y_{t-5} \\ &+ 0.0765\, y_{t-6} - 0.0415\, y_{t-7} - 0.1573\, y_{t-8} + 0.0704\, y_{t-9} - 0.0285\, y_{t-10} \\ &- 0.0477\, y_{t-11} - 0.0693\, y_{t-12} - 0.0101\, y_{t-13} - 0.1807\, y_{t-14} - 0.0845\, y_{t-15} \\ &- 0.1998\, y_{t-16}. \end{aligned}$$

(4) s_2 vs. s_4:

$$\begin{aligned} w_t = y_t &- 0.3739\, y_{t-1} - 0.1373\, y_{t-2} - 0.0449\, y_{t-3} - 0.0314\, y_{t-4} - 0.0258\, y_{t-5} \\ &- 0.0277\, y_{t-6} - 0.1924\, y_{t-7} + 0.0124\, y_{t-8} - 0.0899\, y_{t-9} - 0.1089\, y_{t-10} \\ &- 0.1047\, y_{t-11} - 0.1629\, y_{t-12} + 0.2104\, y_{t-13} + 0.1159\, y_{t-14} - 0.0290\, y_{t-15} \\ &- 0.1781\, y_{t-16} + 0.1001\, y_{t-17} - 0.0729\, y_{t-18} + 0.0731\, y_{t-19} + 0.1203\, y_{t-20}. \end{aligned}$$

(5) s_2 vs. s_5:

$$\begin{aligned} w_t = y_t &- 0.0624\, y_{t-1} - 0.1152\, y_{t-2} + 0.0496\, y_{t-3} + 0.0637\, y_{t-4} - 0.0124\, y_{t-5} \\ &+ 0.0781\, y_{t-6} - 0.0397\, y_{t-7} - 0.1579\, y_{t-8} + 0.0702\, y_{t-9} - 0.0258\, y_{t-10} \\ &- 0.0476\, y_{t-11} - 0.0689\, y_{t-12} - 0.0095\, y_{t-13} - 0.1808\, y_{t-14} - 0.0840\, y_{t-15} \\ &- 0.1978\, y_{t-16}. \end{aligned}$$

(6) s_4 vs. s_5:

$$\begin{aligned} w_t = y_t &- 0.4115\, y_{t-1} - 0.1596\, y_{t-2} - 0.0615\, y_{t-3} - 0.0111\, y_{t-4} - 0.0128\, y_{t-5} \\ &- 0.0481\, y_{t-6} - 0.1727\, y_{t-7} + 0.0745\, y_{t-8} - 0.0537\, y_{t-9} - 0.0517\, y_{t-10} \\ &- 0.0706\, y_{t-11} - 0.1301\, y_{t-12} + 0.2411\, y_{t-13} + 0.0958\, y_{t-14} - 0.0117\, y_{t-15} \\ &- 0.1726\, y_{t-16}. \end{aligned}$$

Calculating the residuals of the data in each case, and calculating their PACFs, we get the values in Table 4.2.

	$\chi^2_{(p)}$	$\chi^2_{0.05,\ p}$	**Result**
$\mathbf{s_1}$ **vs.** $\mathbf{s_2}$**:**	$\chi^2_{(1)} = 0.1895918$	$\chi^2_{0.05,\ 1} = 3.84$	accept H_0
$\mathbf{s_1}$ **vs.** $\mathbf{s_4}$**:**	$\chi^2_{(17)} = 7.659143$	$\chi^2_{0.05,\ 17} = 27.59$	accept H_0
$\mathbf{s_1}$ **vs.** $\mathbf{s_5}$**:**	$\chi^2_{(16)} = 15.40808$	$\chi^2_{0.05,\ 16} = 26.30$	accept H_0
$\mathbf{s_2}$ **vs.** $\mathbf{s_4}$**:**	$\chi^2_{(20)} = 18.21268$	$\chi^2_{0.05,\ 20} = 31.41$	accept H_0
$\mathbf{s_2}$ **vs.** $\mathbf{s_5}$**:**	$\chi^2_{(16)} = 14.96461$	$\chi^2_{0.05,\ 16} = 26.30$	accept H_0
$\mathbf{s_4}$ **vs.** $\mathbf{s_5}$**:**	$\chi^2_{(16)} = 16.21758$	$\chi^2_{0.05,\ 16} = 26.30$	accept H_0

Table 4.2: Correlation comparison for group 1 of zinc data

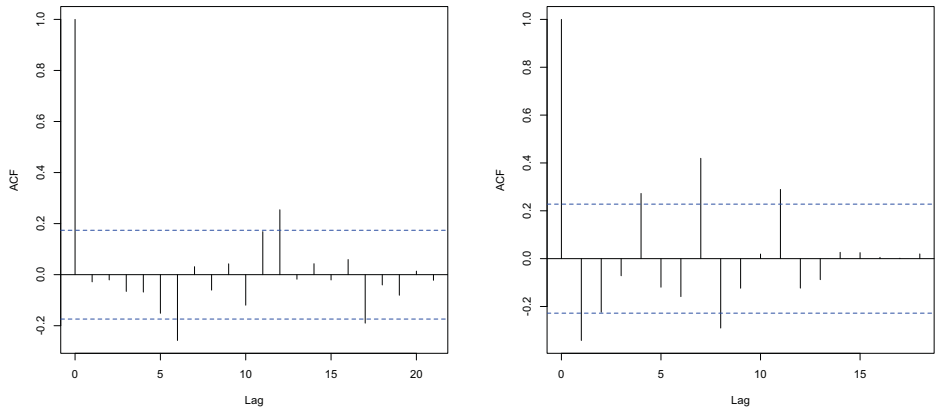
From Table 4.2, we see that s_1, s_2, s_4 and s_5 have the same correlation structure. This allows us to proceed to the third step and test the difference in the distributions of the residuals of the samples, which are calculated based on the fitted model of each pair. Since we have one sample from each site, the most suitable two-sample test to be used is the Bayesian two-sample test [7]. The script of the test is available online for Matlab/Octave by François Caron, who is one of the authors of [7].

We see from the results of this test in Table 4.3 that the samples s_1 and s_2 have the same model, but they only differ from the s_4 and s_5 models in the distributions of residuals.

residuals of :	**Result**	$P(H_0\|y_1, y_2)$
$\mathbf{s_1}$ **vs.** $\mathbf{s_2}$**:**	accept H_0	1.0000
$\mathbf{s_1}$ **vs.** $\mathbf{s_4}$**:**	reject H_0	2.3137×10^{-50}
$\mathbf{s_1}$ **vs.** $\mathbf{s_5}$**:**	reject H_0	6.6413×10^{-24}
$\mathbf{s_2}$ **vs.** $\mathbf{s_4}$**:**	reject H_0	7.2522×10^{-50}
$\mathbf{s_2}$ **vs.** $\mathbf{s_5}$**:**	reject H_0	4.3468×10^{-18}
$\mathbf{s_4}$ **vs.** $\mathbf{s_5}$**:**	reject H_0	4.8405×10^{-18}

Table 4.3: Bayesian two-sample test results for the residual distributions in group 1 of zinc data

We now consider the second group of samples, our first step is to transform our samples to stationary time series. Starting with s_3 we apply differencing, since the sample shows the existence of a unit root. Applying the same transformation to s_6, we see that it is also transformed to a stationary time series. By denoting the transformed data of s_3 and s_6 by t_3, t_6 we see from their ACF plots in Figure 4.4 that they have no trend. We obtain their stationarity (level-stationarity) results in Table 4.4.



(a) Site 3 (b) Site 6

Figure 4.4: ACF plots for transformed zinc samples

sample	**Test**		**stationary:**
	KPSS	**ADF**	**non-stationary**
t_3	0.1000000	0.0100000	2:0
t_6	0.1000000	0.0100000	2:0

Table 4.4: P-values and stationarity of transformed zinc samples

Working with the transformed stationary samples, we compare their correlation structure, by fitting them gathered to an $AR(p)$ model first:

(1) t_3 vs. t_6:

$$\begin{aligned} w_t = y_t &+0.1336\ y_{y-1} + 0.1076\ y_{t-2} + 0.1348\ y_{t-3} + 0.1468\ y_{t-4} + 0.2234\ y_{t-5} \\ &+0.2718\ y_{t-6} + 0.0296\ y_{t-7} + 0.1387\ y_{t-8} + 0.0568\ y_{t-9} + 0.1962\ y_{t-10} \\ &-0.0496\ y_{t-11} - 0.1665\ y_{t-12}. \end{aligned}$$

We then calculate the residuals of each transformed sample, and their PACFs to get the $\chi^2_{(p)}$ values as seen in Table 4.5.

	$\chi^2_{(p)}$	$\chi^2_{0.05,\ p}$	**Result**
$\mathbf{t_3}$ **vs.** $\mathbf{t_6}$**:**	$\chi^2_{(12)} = 29.38165$	$\chi^2_{0.05,\ 12} = 21.03$	reject H_0

Table 4.5: Correlation comparison for group 2 of zinc data

We see that t_3 and t_6 do not share the same correlation, hence, there is no need to compare their residual distributions and they have different models.

Second, we follow the same steps to study the water quality time series for lead levels in the Yukon river basin, where the samples have the lengths $103, 103, 126, 101, 101, 75$, see Figure 4.5.

Using the same notations of the zinc example, and noting that the samples do not have a trend component (Figure 4.6) we run the stationarity tests and get the results in table 4.6.

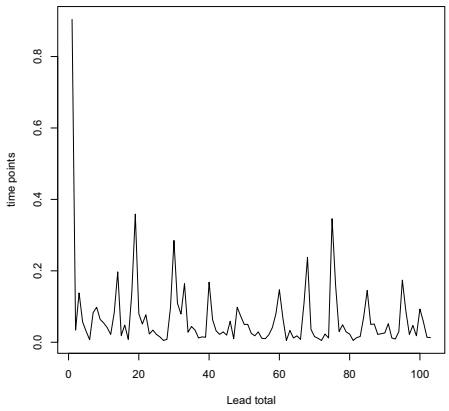
sample	**Test**		**stationary:**
	KPSS	**ADF**	**non-stationary**
s_1	0.06966754	0.01000000	2:0
s_2	0.1000000	0.0100000	2:0
s_3	0.100000000	0.010000000	2:0
s_4	0.1000000	0.0100000	2:0
s_5	0.1000000	0.0100000	2:0
s_6	0.01000000	0.02935544	1:1

Table 4.6: P-values and stationarity of lead samples

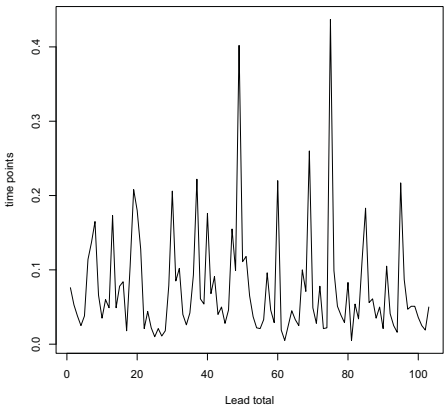
From the stationarity tests, we can see that the only sample that does not pass both tests is s_6, which means that it differs from the rest of the samples. We continue the test with s_1, s_2, s_3, s_4 and s_5 by comparing the two samples correlation structure in each case as follows:

(1) s_1 vs. s_2:

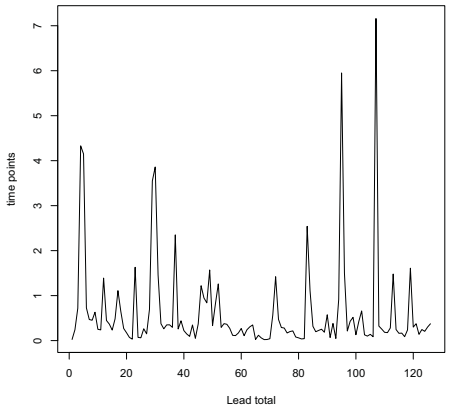
$$w_t = y_t.$$



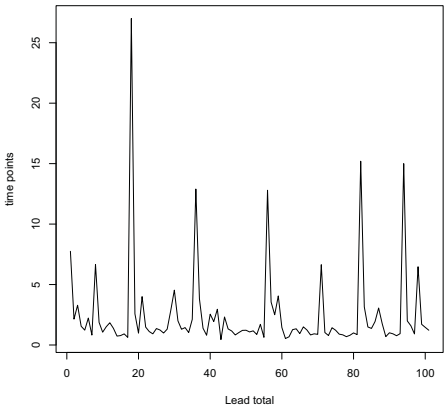
(a) Site 1



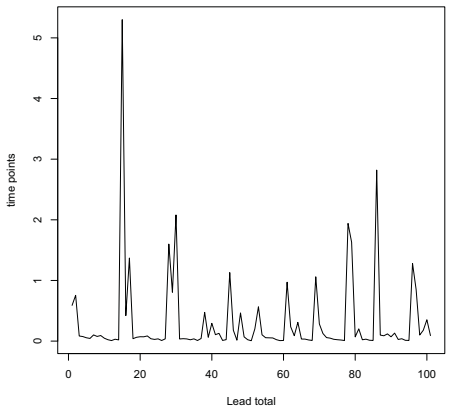
(b) Site 2



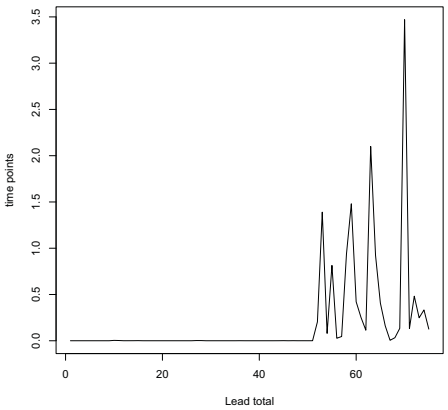
(c) Site 3



(d) Site 4

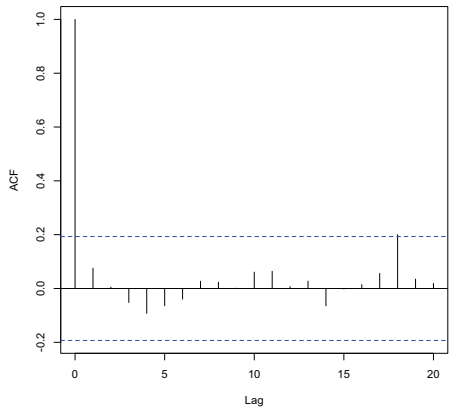


(e) Site 5

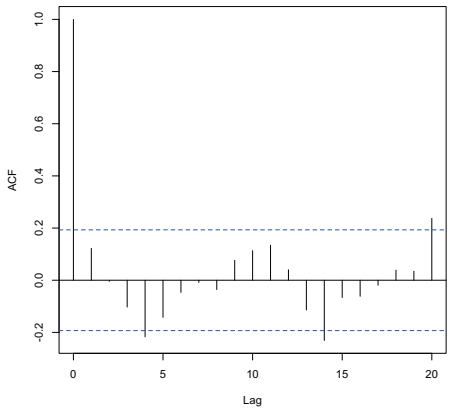


(f) Site 6

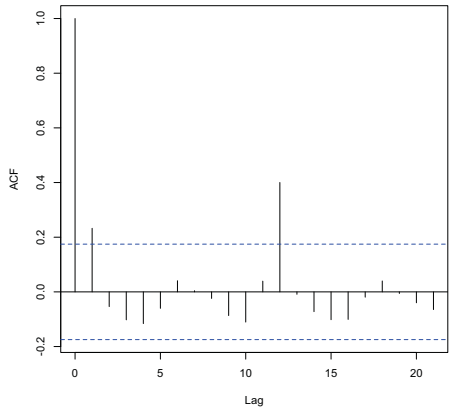
Figure 4.5: Time series plots for lead levels



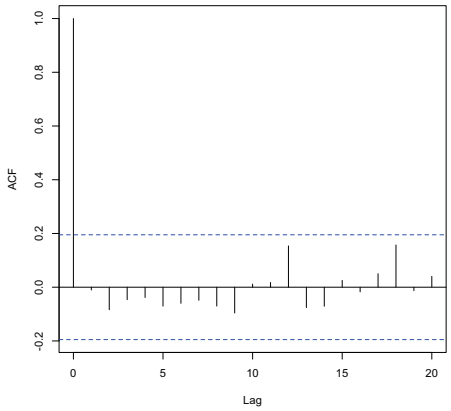
(a) Site 1



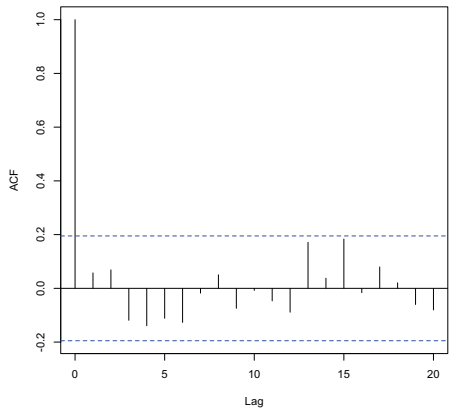
(b) Site 2



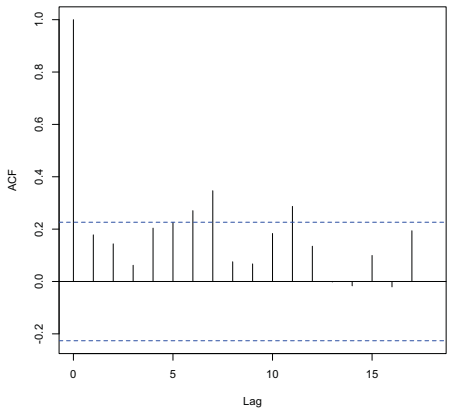
(c) Site 3



(d) Site 4



(e) Site 5



(f) Site 6

Figure 4.6: ACF plots for lead levels

(2) s_1 vs s_3:

$$\begin{aligned} w_t = y_t &- 0.3697\, y_{t-1} + 0.0173\, y_{t-2} - 0.0163\, y_{t-3} + 0.0437\, y_{t-4} + 0.0046\, y_{t-5} \\ &- 0.0661\, y_{t-6} - 0.0314\, y_{t-7} - 0.0381\, y_{t-8} + 0.0359\, y_{t-9} + 0.0309\, y_{t-10} \\ &- 0.0046\, y_{t-11} - 0.4652\, y_{t-12} + 0.2447\, y_{t-13}. \end{aligned}$$

(3) s_1 vs s_4:

$$w_t = y_t - 0.1595\, y_{t-1}.$$

(4) s_1 vs s_5:

$$w_t = y_t - 0.1002\, y_{t-1} - 0.1061\, y_{t-2}.$$

(5) s_2 vs. s_3:

$$\begin{aligned} w_t = y_t &- 0.3701\, y_{t-1} + 0.0187\, y_{t-2} - 0.0142\, y_{t-3} + 0.0436\, y_{t-4} + 0.0042\, y_{t-5} \\ &- 0.0642\, y_{t-6} - 0.0280\, y_{t-7} - 0.0375\, y_{t-8} + 0.0351\, y_{t-9} + 0.0322\, y_{t-10} \\ &- 0.0031\, y_{t-11} - 0.4667\, y_{t-12} + 0.2466\, y_{t-13}. \end{aligned}$$

(6) s_2 vs. s_4:

$$w_t = y_t - 0.1579\, y_{t-1}.$$

(7) s_2 vs. s_5:

$$w_t = y_t - 0.0965\, y_{t-1} - 0.1029\, y_{t-2}.$$

(8) s_3 vs. s_4:

$$w_t = y_t - 0.1142\, y_{t-1}.$$

(9) s_3 vs. s_5:

$$\begin{aligned} w_t = y_t &- 0.1935\, y_{t-1} + 0.0064\, y_{t-2} + 0.0385\, y_{t-3} + 0.0658\, y_{t-4} + 0.0289\, y_{t-5} \\ &- 0.0208\, y_{t-6} - 0.0107\, y_{t-7} - 0.0396\, y_{t-8} + 0.0401\, y_{t-9} + 0.0399\, y_{t-10} \\ &- 0.0084\, y_{t-11} - 0.2825\, y_{t-12}. \end{aligned}$$

(10) s_4 vs. s_5:

$$w_t = y_t - 0.131\, y_{t-1}.$$

	$\chi^2_{(p)}$	$\chi^2_{0.05,\,p}$	**Result**
$\mathbf{s_1}$ **vs.** $\mathbf{s_3}$:	$\chi^2_{(13)} = 15.38609$	$\chi^2_{0.05,\,13} = 22.36$	accept H_0
$\mathbf{s_1}$ **vs.** $\mathbf{s_4}$:	$\chi^2_{(1)} = 0.2432327$	$\chi^2_{0.05,\,1} = 3.84$	accept H_0
$\mathbf{s_1}$ **vs.** $\mathbf{s_5}$:	$\chi^2_{(2)} = 0.2805838$	$\chi^2_{0.05,\,2} = 5.99$	accept H_0
$\mathbf{s_2}$ **vs.** $\mathbf{s_3}$:	$\chi^2_{(13)} = 18.4757$	$\chi^2_{0.05,\,13} = 22.36$	accept H_0
$\mathbf{s_2}$ **vs.** $\mathbf{s_4}$:	$\chi^2_{(1)} = 0.7021482$	$\chi^2_{0.05,\,1} = 3.84$	accept H_0
$\mathbf{s_2}$ **vs.** $\mathbf{s_5}$:	$\chi^2_{(2)} = 0.4781823$	$\chi^2_{0.05,\,2} = 5.99$	accept H_0
$\mathbf{s_3}$ **vs.** $\mathbf{s_4}$:	$\chi^2_{(1)} = 3.360954$	$\chi^2_{0.05,\,1} = 3.84$	accept H_0
$\mathbf{s_3}$ **vs.** $\mathbf{s_5}$:	$\chi^2_{(12)} = 20.66411$	$\chi^2_{0.05,\,12} = 21.03$	accept H_0
$\mathbf{s_4}$ **vs.** $\mathbf{s_5}$:	$\chi^2_{(1)} = 0.1096292$	$\chi^2_{0.05,\,1} = 3.84$	accept H_0

Table 4.7: Correlation comparison for lead data

Since each pair share the same correlation (Table 4.7) we continue to test the distributions of the residuals by the Bayesian two-sample test.

residuals of :	**Result**	$P(H_0\|y_1, y_2)$
$\mathbf{s_1}$ **vs.** $\mathbf{s_2}$**:**	accept H_0	0.9998
$\mathbf{s_1}$ **vs.** $\mathbf{s_3}$**:**	reject H_0	1.4812×10^{-16}
$\mathbf{s_1}$ **vs.** $\mathbf{s_4}$**:**	reject H_0	1.6966×10^{-52}
$\mathbf{s_1}$ **vs.** $\mathbf{s_5}$**:**	accept H_0	0.9998
$\mathbf{s_2}$ **vs.** $\mathbf{s_3}$**:**	reject H_0	2.3196×10^{-17}
$\mathbf{s_2}$ **vs.** $\mathbf{s_4}$**:**	reject H_0	1.4426×10^{-53}
$\mathbf{s_2}$ **vs.** $\mathbf{s_5}$**:**	reject H_0	0.3565
$\mathbf{s_3}$ **vs.** $\mathbf{s_4}$**:**	reject H_0	5.1348×10^{-23}
$\mathbf{s_3}$ **vs.** $\mathbf{s_5}$**:**	reject H_0	0.3089
$\mathbf{s_4}$ **vs.** $\mathbf{s_5}$**:**	reject H_0	2.2098×10^{-29}

Table 4.8: Bayesian two-sample test results for the residual distributions of lead data

From the results of the residual distribution test, we note that the sample s_1 shares the same model with s_2 and s_5.

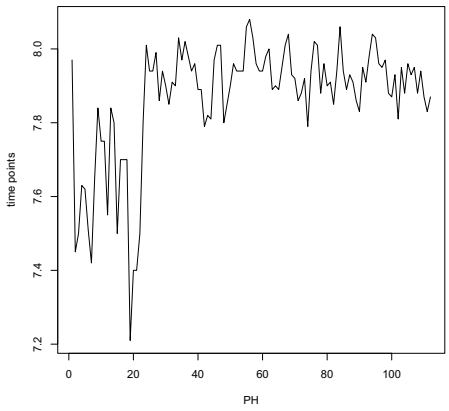
For the PH samples, we follow the same steps to study their models. The samples have the lengths $112, 107, 128, 95, 89, 75$, see Figure 4.7.

Using the same notations, and since there is no trend component in s_3, s_4 and s_5, we apply the stationarity tests. However for s_1, s_2 and s_6 since they exhibit a trend based on Figure 4.8, we apply trend-stationarity tests. The test results are obtained in Table 4.9.

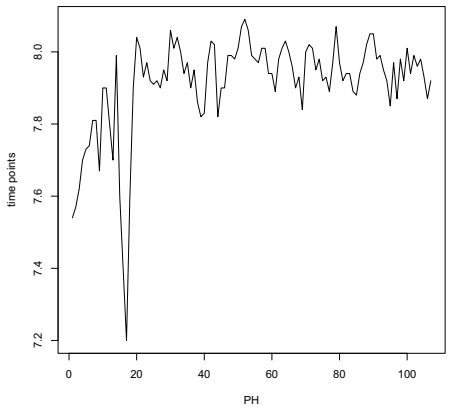
sample	**Test**		**stationary:**
	KPSS	**ADF**	**non-stationary**
s_1	0.01000000	0.07607635	0:2
s_2	0.01000000	0.01102785	1:1
s_3	0.10	0.01	2:0
s_4	0.10	0.01	2:0
s_5	0.10	0.01	2:0
s_6	0.1000000	0.1294852	1:1

Table 4.9: P-values and stationarity of PH samples

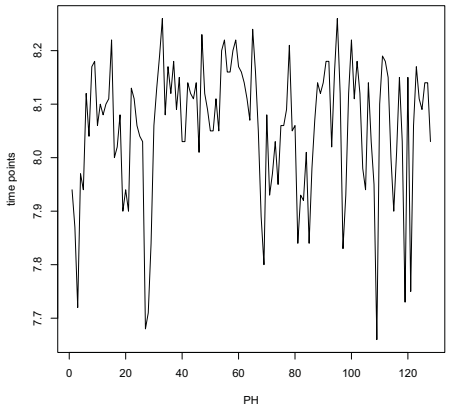
From Table 4.9, based on the results we form two groups. The first consists of s_3, s_4 and s_5 and the second contains s_1, s_2 and s_6. This implies that the samples in the first group have



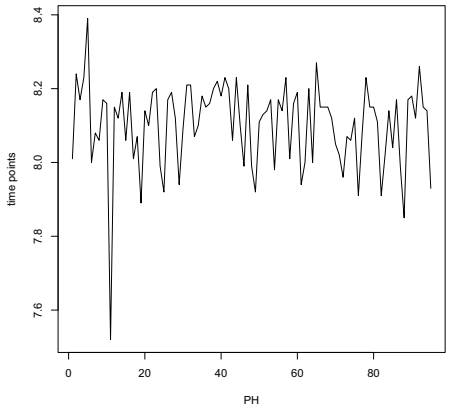
(a) Site 1



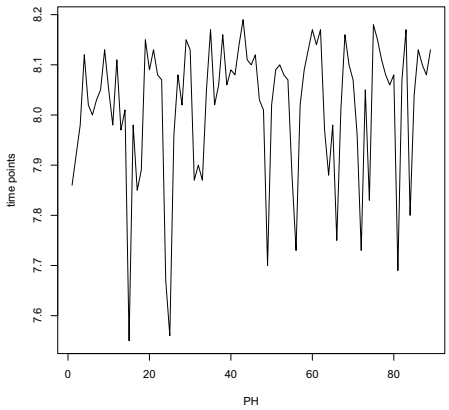
(b) Site 2



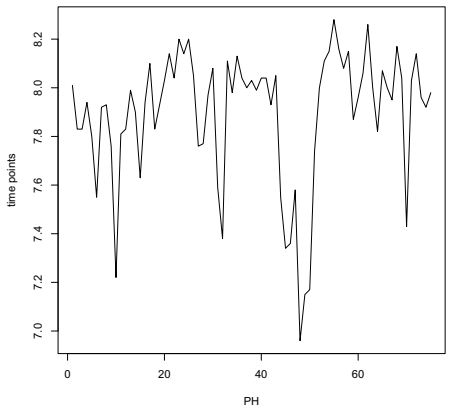
(c) Site 3



(d) Site 4

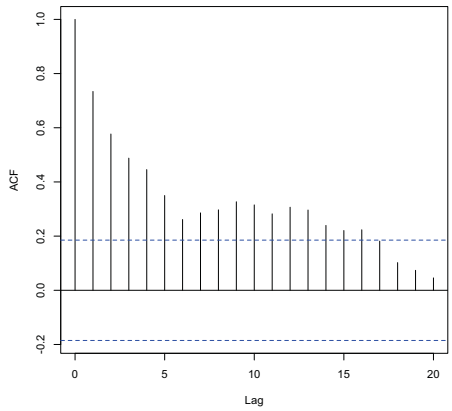


(e) Site 5

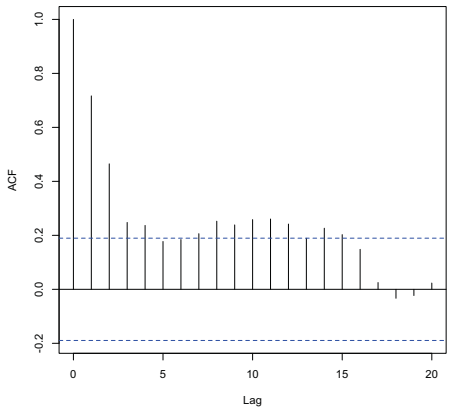


(f) Site 6

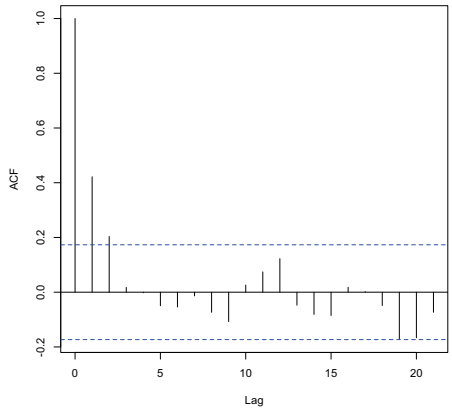
Figure 4.7: Time series plots for PH levels



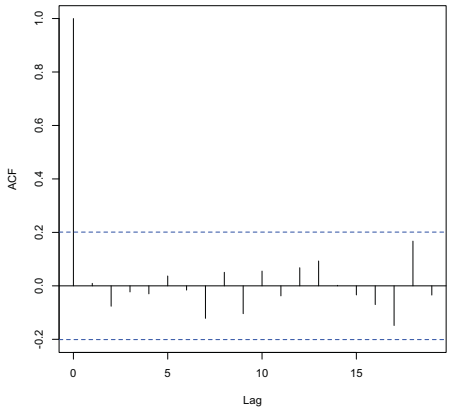
(a) Site 1



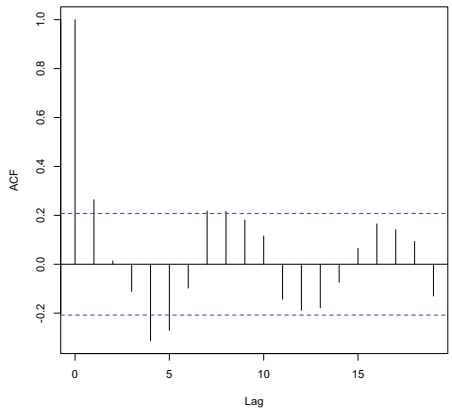
(b) Site 2



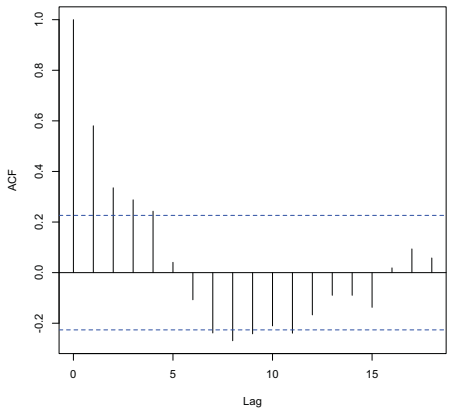
(c) Site 3



(d) Site 4



(e) Site 5



(f) Site 6

Figure 4.8: ACF plots for PH levels

models different from the samples of the second group. Starting with the stationary group, we compared the two-sample correlation structures in each case as follows:

(1) s_3 vs. s_4:

$$w_t = y_t - 0.2868\, y_{t-1}.$$

(2) s_3 vs s_5:

$$w_t = y_t - 0.3651\, y_{t-1}.$$

(3) s_4 vs s_5:

$$w_t = y_t - 0.2522\, y_{t-1}.$$

	$\chi^2_{(p)}$	$\chi^2_{0.05,\,p}$	**Result**
$\mathbf{s_3}$ **vs.** $\mathbf{s_4}$**:**	$\chi^2_{(1)} = 0.04555956$	$\chi^2_{0.05,\,1} = 3.84$	accept H_0
$\mathbf{s_3}$ **vs.** $\mathbf{s_5}$**:**	$\chi^2_{(1)} = 0.005987199$	$\chi^2_{0.05,\,1} = 3.84$	accept H_0
$\mathbf{s_4}$ **vs.** $\mathbf{s_5}$**:**	$\chi^2_{(1)} = 0.02198864$	$\chi^2_{0.05,\,1} = 3.84$	accept H_0

Table 4.10: Correlation comparison for group 1 of PH data

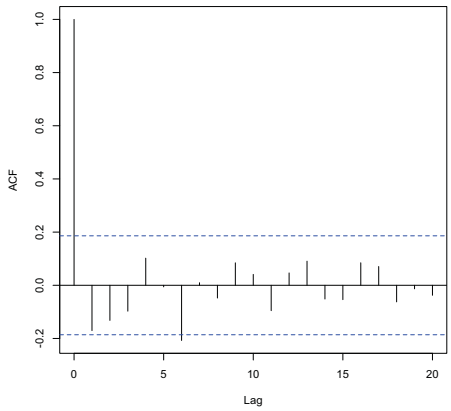
From Table 4.10, we see that all samples of group 1 share the same correlation structure, so we continue to test the distributions of the residuals by the Bayesian two-sample test.

residuals of :	**Result**	$P(H_0\|y_1, y_2)$
$\mathbf{s_3}$ **vs.** $\mathbf{s_4}$**:**	accept H_0	0.9039
$\mathbf{s_3}$ **vs.** $\mathbf{s_5}$**:**	accept H_0	0.9961
$\mathbf{s_4}$ **vs.** $\mathbf{s_5}$**:**	reject H_0	0.0024

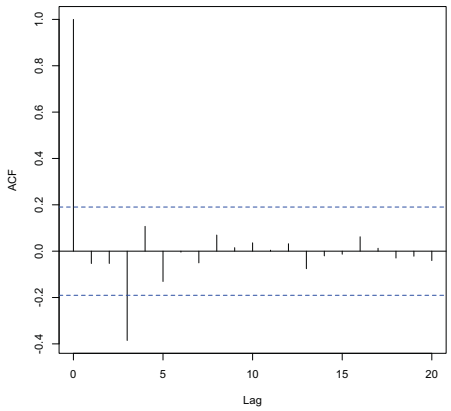
Table 4.11: Bayesian two-sample test results for the residual distributions in group 1 of PH data

Based on residual distribution test results, we see that s_3 shares the same model with s_4 and s_5.

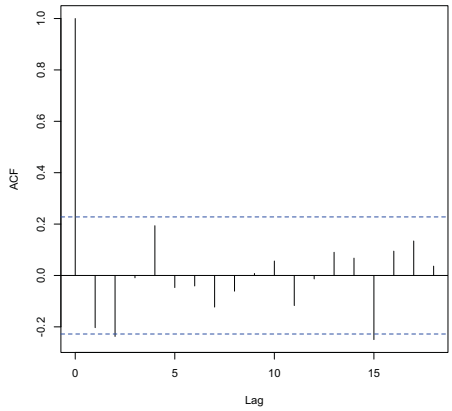
We now consider the second group of samples. Applying the differencing transformation on s_1 and running the stationarity tests, we get a stationary sample. Applying the same



(a) Site 1



(b) Site 2



(c) Site 6

Figure 4.9: ACF plots for transformed PH samples

transformation to s_2 and s_6, and testing their stationarity as well we get the results in Table 4.12. Since there is no trend in the ACF plots of the transformed samples (Figure 4.9) we will use stationarity tests.

sample	**Test**		**stationary:**
	KPSS	**ADF**	**non-stationary**
t_1	0.10	0.01	2:0
t_2	0.10	0.01	2:0
t_6	0.1000000	0.0126347	2:0

Table 4.12: P-values and stationarity of transformed PH samples

Fitting the gathered transformed samples to $AR(p)$ models to test their correlation structures, we get:

(1) t_1 vs. t_2:

$$\begin{aligned} w_t = y_t &+0.2034\, y_{t-1} + 0.2087\, y_{t-2} + 0.3339\, y_{t-3} + 0.0581\, y_{t-4} + 0.1883\, y_{t-5} \\ &+0.2403\, y_{t-6} + 0.0969\, y_{t-7} + 0.1237\, y_{t-8}. \end{aligned}$$

(2) t_1 vs. t_6:

$$\begin{aligned} w_t = y_t &+0.3223\, y_{t-1} + 0.3055\, y_{t-2} + 0.1458\, y_{t-3} - 0.0515\, y_{t-4} + 0.0712\, y_{t-5} \\ &+0.1352\, y_{t-6} + 0.2100\, y_{t-7} + 0.2255\, y_{t-8} + 0.1074\, y_{t-9}. \end{aligned}$$

(3) t_2 vs. t_6:

$$\begin{aligned} w_t = y_t &+0.3106\, y_{t-1} + 0.3026\, y_{t-2} + 0.2585\, y_{t-3} + 0.0482\, y_{t-4} + 0.2255\, y_{t-5} \\ &+0.1884\, y_{t-6} + 0.2365\, y_{t-7} + 0.2763\, y_{t-8} + 0.2009\, y_{t-9} + 0.1217\, y_{t-10} \\ &+0.1474\, y_{t-11} + 0.1477\, y_{t-12} + 0.1104\, y_{t-13} + 0.0746\, y_{t-14} + 0.2057\, y_{t-15} \\ &+0.0286\, y_{t-16} - 0.0411\, y_{t-17} + 0.0463\, y_{t-18} - 0.0452\, y_{t-19} + 0.0804\, y_{t-20} \\ &-0.0243\, y_{t-21} - 0.2352\, y_{t-22}. \end{aligned}$$

The $\chi^2_{(p)}$ values based on the PACFs of the residuals are given in Table 4.13.

	$\chi^2_{(p)}$	$\chi^2_{0.05,\ p}$	**Result**
$\mathbf{t_1}$ **vs.** $\mathbf{t_2}$**:**	$\chi^2_{(8)} = 9.410847$	$\chi^2_{0.05,\ 8} = 15.51$	accept H_0
$\mathbf{t_1}$ **vs.** $\mathbf{t_6}$**:**	$\chi^2_{(9)} = 6.217005$	$\chi^2_{0.05,\ 9} = 16.92$	accept H_0
$\mathbf{t_2}$ **vs.** $\mathbf{t_6}$**:**	$\chi^2_{(22)} = 14.78596$	$\chi^2_{0.05,\ 22} = 33.92$	accept H_0

Table 4.13: Correlation comparison for group 2 of PH data

Based on the Bayesian two-sample test results in Table 4.13, s_1 and s_2 share the same model.

residuals of :	**Result**	$P(H_0\|y_1, y_2)$
$\mathbf{t_1}$ **vs.** $\mathbf{t_2}$**:**	accept H_0	0.8026
$\mathbf{t_1}$ **vs.** $\mathbf{t_6}$**:**	reject H_0	9.7893×10^{-06}
$\mathbf{t_2}$ **vs.** $\mathbf{t_6}$**:**	reject H_0	2.4720×10^{-05}

Table 4.14: Bayesian two-sample test results for the residual distributions in group 2 of PH data

Finally, we consider the alkalinity(HCO_3^-) levels in the studied six sites. The sample lengths are as follow $111, 107, 128, 103, 102, 75$, see Figure 4.10.

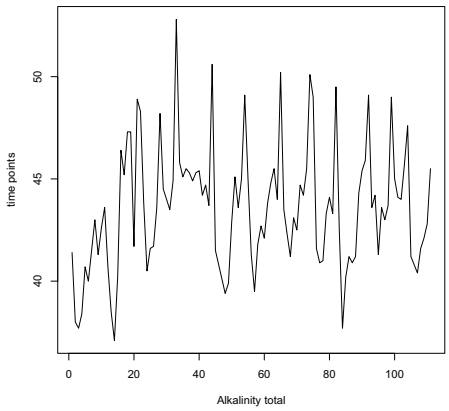
sample	**Test**		**stationary:**
	KPSS	**ADF**	**non-stationary**
s_1	0.10	0.01	2:0
s_2	0.10	0.01	2:0
s_3	0.10	0.01	2:0
s_4	0.10	0.01	2:0
s_5	0.10	0.01	2:0
s_6	0.07777487	0.07544538	1:1

Table 4.15: P-values and stationarity of alkalinity (HCO_3^-) samples

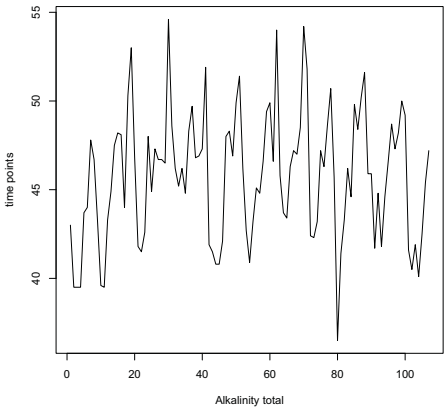
Running the stationarity tests on the samples, since there is no trend (Figure 4.11) we see that all samples are stationary except s_6. We continue the study with the stationary samples by comparing their correlation structure in each case as follows:

(1) s_1 vs. s_2:

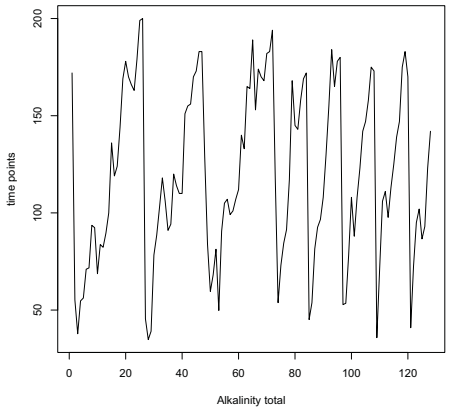
$$\begin{aligned} w_t = y_t &- 0.6027\, y_{t-1} + 0.0710\, y_{t-2} + 0.0383\, y_{t-3} - 0.0008\, y_{t-4} + 0.0524\, y_{t-5} \\ &+ 0.0124\, y_{t-6} - 0.1622\, y_{t-7} - 0.0084\, y_{t-8} + 0.0208\, y_{t-9} - 0.0196\, y_{t-10} \\ &- 0.2808\, y_{t-11} + 0.0359\, y_{t-12} + 0.1728\, y_{t-13} - 0.0400\, y_{t-14} - 0.0234\, y_{t-15} \\ &+ 0.0988\, y_{t-16} - 0.2073\, y_{t-17} + 0.2194\, y_{t-18} - 0.0522\, y_{t-19} + 0.0564\, y_{t-20} \\ &- 0.1989\, y_{t-21} + 0.1317\, y_{t-22}. \end{aligned}$$



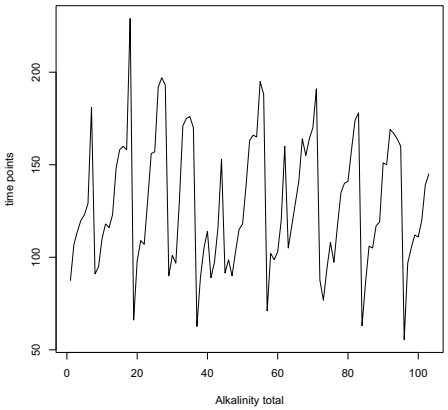
(a) Site 1



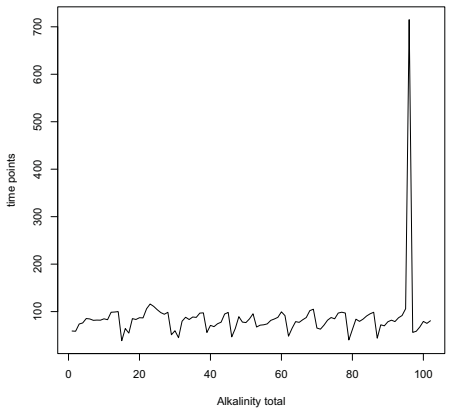
(b) Site 2



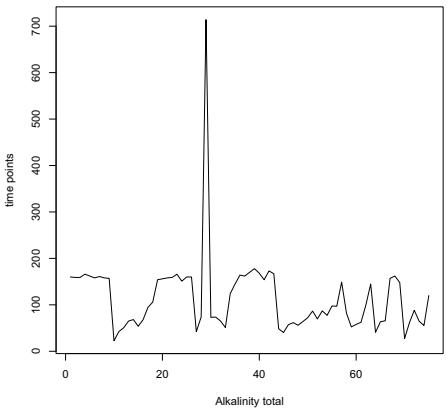
(c) Site 3



(d) Site 4

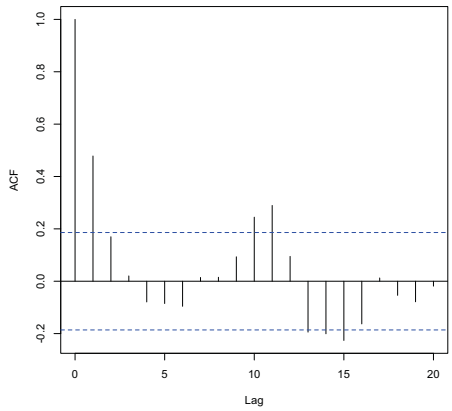


(e) Site 5

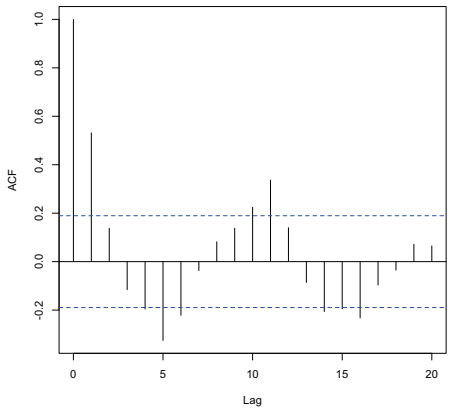


(f) Site 6

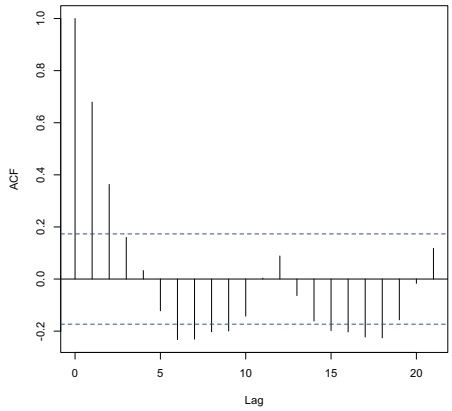
Figure 4.10: Time series plots for alkalinity (HCO_3^-) levels



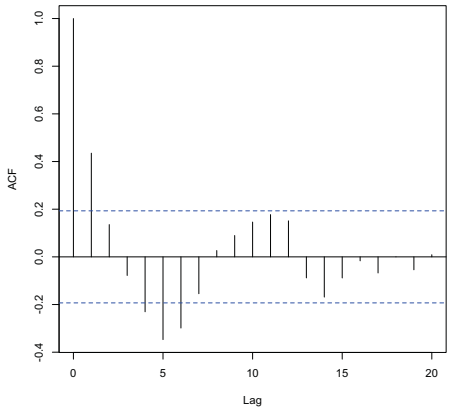
(a) Site 1



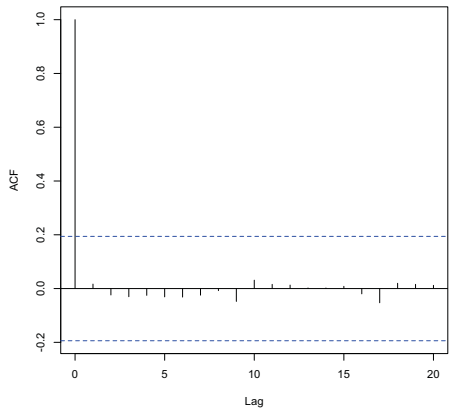
(b) Site 2



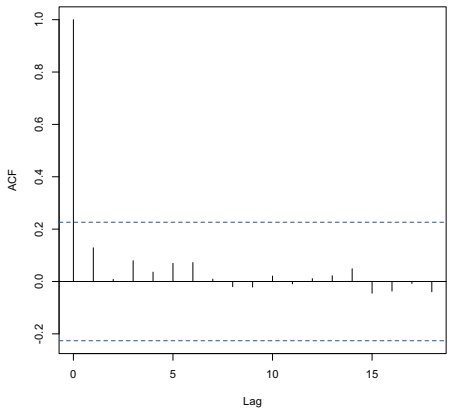
(c) Site 3



(d) Site 4



(e) Site 5



(f) Site 6

Figure 4.11: ACF plots for alkalinity (HCO_3^-) levels

(2) s_1 vs s_3:

$$\begin{aligned} w_t = y_t &- 0.8113\, y_{t-1} + 0.0454\, y_{t-2} - 0.0003\, y_{t-3} - 0.0728\, y_{t-4} + 0.0310\, y_{t-5} \\ &+ 0.0596\, y_{t-6} - 0.0509\, y_{t-7} - 0.0619\, y_{t-8} + 0.0694\, y_{t-9} + 0.0034\, y_{t-10} \\ &- 0.1353\, y_{t-11} - 0.2313\, y_{t-12} + 0.2312\, y_{t-13}. \end{aligned}$$

(3) s_1 vs s_4:

$$\begin{aligned} w_t = y_t &- 0.6254\, y_{t-1} - 0.0806\, y_{t-2} - 0.0429\, y_{t-3} - 0.0087\, y_{t-4} + 0.0908\, y_{t-5} \\ &- 0.0352\, y_{t-6} - 0.0071\, y_{t-7} - 0.2369\, y_{t-8}. \end{aligned}$$

(4) s_1 vs s_5:

$$w_t = y_t - 0.1485\, y_{t-1} - 0.1016\, y_{t-2} - 0.0947\, y_{t-3} - 0.1036\, y_{t-4}.$$

(5) s_2 vs. s_3:

$$\begin{aligned} w_t = y_t &- 0.8114\, y_{t-1} + 0.0446\, y_{t-2} + 0.0029\, y_{t-3} - 0.0707\, y_{t-4} + 0.0281\, y_{t-5} \\ &+ 0.0621\, y_{t-6} - 0.0522\, y_{t-7} - 0.0617\, y_{t-8} + 0.0673\, y_{t-9} + 0.0063\, y_{t-10} \\ &- 0.1337\, y_{t-11} - 0.2364\, y_{t-12} + 0.2362\, y_{t-13}. \end{aligned}$$

(6) s_2 vs. s_4:

$$\begin{aligned} w_t = y_t &- 0.6232\, y_{t-1} - 0.0805\, y_{t-2} - 0.0426\, y_{t-3} - 0.0083\, y_{t-4} + 0.0930\, y_{t-5} \\ &- 0.0351\, y_{t-6} - 0.0091\, y_{t-7} - 0.2370\, y_{t-8}. \end{aligned}$$

(7) s_2 vs. s_5:

$$w_t = y_t - 0.1503\, y_{t-1} - 0.1035\, y_{t-2} - 0.1008\, y_{t-3}.$$

(8) s_3 vs. s_4:

$$\begin{aligned} w_t = y_t &- 0.6335\, y_{t-1} + 0.0253\, y_{t-2} + 0.0518\, y_{t-3} + 0.0083\, y_{t-4} + 0.0891\, y_{t-5} \\ &+ 0.0630\, y_{t-6} + 0.0001\, y_{t-7} - 0.0406\, y_{t-8} + 0.0950\, y_{t-9} + 0.0423\, y_{t-10} \\ &- 0.0753\, y_{t-11} - 0.1818\, y_{t-12} + 0.2871\, y_{t-13}. \end{aligned}$$

(9) s_3 vs. s_5:

$$w_t = y_t - 0.2893\, y_{t-1} - 0.0943\, y_{t-2}.$$

(10) s_4 vs. s_5:

$$w_t = y_t - 0.2359\, y_{t-1}.$$

Based on the fitted model in each case, and by calculating the PACFs of the residuals of each sample, we compare the correlation structures depending on the $\chi^2_{(p)}$ values in Table 4.16.

	$\chi^2_{(p)}$	$\chi^2_{0.05,\ p}$	**Result**
$\mathbf{s_1}$ **vs.** $\mathbf{s_2}$:	$\chi^2_{(22)} = 5.690048$	$\chi^2_{0.05,\ 22} = 33.92$	accept H_0
$\mathbf{s_1}$ **vs.** $\mathbf{s_3}$:	$\chi^2_{(13)} = 7.48145$	$\chi^2_{0.05,\ 13} = 22.36$	accept H_0
$\mathbf{s_1}$ **vs.** $\mathbf{s_4}$:	$\chi^2_{(8)} = 9.182556$	$\chi^2_{0.05,\ 8} = 15.51$	accept H_0
$\mathbf{s_1}$ **vs.** $\mathbf{s_5}$:	$\chi^2_{(4)} = 9.911388$	$\chi^2_{0.05,\ 4} = 9.49$	reject H_0
$\mathbf{s_2}$ **vs.** $\mathbf{s_3}$:	$\chi^2_{(13)} = 8.155671$	$\chi^2_{0.05,\ 13} = 22.36$	accept H_0
$\mathbf{s_2}$ **vs.** $\mathbf{s_4}$:	$\chi^2_{(8)} = 9.590492$	$\chi^2_{0.05,\ 8} = 15.51$	accept H_0
$\mathbf{s_2}$ **vs.** $\mathbf{s_5}$:	$\chi^2_{(3)} = 11.68419$	$\chi^2_{0.05,\ 3} = 7.81$	reject H_0
$\mathbf{s_3}$ **vs.** $\mathbf{s_4}$:	$\chi^2_{(13)} = 7.66186$	$\chi^2_{0.05,\ 13} = 22.36$	accept H_0
$\mathbf{s_3}$ **vs.** $\mathbf{s_5}$:	$\chi^2_{(2)} = 23.57606$	$\chi^2_{0.05,\ 2} = 5.99$	reject H_0
$\mathbf{s_4}$ **vs.** $\mathbf{s_5}$:	$\chi^2_{(1)} = 9.372886$	$\chi^2_{0.05,\ 1} = 3.84$	reject H_0

Table 4.16: Correlation comparison for alkalinity (HCO_3^-) data

From the correlation structure results, we see that s_1, s_2, s_3 and s_4 share the same correlation structure, hence, we continue to compare their noise distributions. We conclude that the sample from s_5 has a different model.

residuals of :	**Result**	$P(H_0\|y_1, y_2)$
$\mathbf{s_1}$ **vs.** $\mathbf{s_2}$**:**	accept H_0	0.5554
$\mathbf{s_1}$ **vs.** $\mathbf{s_3}$**:**	reject H_0	3.6059×10^{-27}
$\mathbf{s_1}$ **vs.** $\mathbf{s_4}$**:**	reject H_0	4.8698×10^{-26}
$\mathbf{s_2}$ **vs.** $\mathbf{s_3}$**:**	reject H_0	6.6108×10^{-27}
$\mathbf{s_2}$ **vs.** $\mathbf{s_4}$**:**	reject H_0	1.0041×10^{-22}
$\mathbf{s_3}$ **vs.** $\mathbf{s_4}$**:**	accept H_0	0.9890

Table 4.17: Bayesian two-sample test results for the residual distributions of alkalinity (HCO_3^-) data

Table 4.17 shows that only the samples s_1 and s_2, s_3 and s_4 share the same model.

Based on the above analysis, we conclude that in the time series models of zinc, lead, PH and alkalinity levels the time series from the first and second site share the same model. This can be a result of the relatively close distance between both sites (around 50 km). Also, considering the PH and alkalinity results, we see that in both cases the samples from the third and the fourth site share the same model, this is compatible with the definition of PH and alkalinity. Since both sites have the same resistance model for change of PH (alkalinity), they also have the same PH models. We also note that in the PH case the sample from site 3 shares the same model with the sample from site 5, this does not happen in the alkalinity case. This could be relied on the fact that alkalinity constitutes of more than one anion, which contributes in resisting the change in PH levels. However, the data we studied measures alkalinity based on the HCO_3^- levels only (cf. [22] and [23]). In the lead samples, we concluded that the sample from s_1 has the same model as the sample from s_2 and the samples from s_1 and s_5 have the same model, but s_2 and s_5 do not share the same model. This does not contradict with our procedure, since this procedure is for comparing two samples only. Also, the difference between the models of s_2 and s_5 is in the residual terms, this means that if the residuals of s_1 and s_2 are similar and the residuals of s_1 and s_5 are also similar, then the residuals of s_2 and s_5 could be less similar than the other pairs, hence the models differ. A similar phenomenon happens in the PH study with the samples from s_3, s_4 and s_5.

4.2 Illustrative examples

To see the advantage of using the BG$^3-$test, we consider the illustrative examples in Biswas and Ghosh [2] by adding the equal means and variances case. Given two time series models, we will test the performance of both tests in location, scale, location-scale, and equal means and variances problems:

$$M_1 : X_t = 0.1X_{t-1} + e_t,$$
$$M_2 : Y_t = 0.1Y_{t-1} + w_t,$$
$$t = 3, ..., 400.$$

By taking 20 samples from each model and repeating the tests 200 times we calculate the powers of both tests as the time points (dimensions) increase. A power of a test is represented by the proportion of rejection of the null hypothesis. The problems considered are:

(a) Location problem: $e_t \sim N(0, 1)$ and $w_t \sim N(0.3, 1)$.

(b) Scale problem: $e_t \sim N(0, 1)$ and $w_t \sim N(0, 1.3)$.

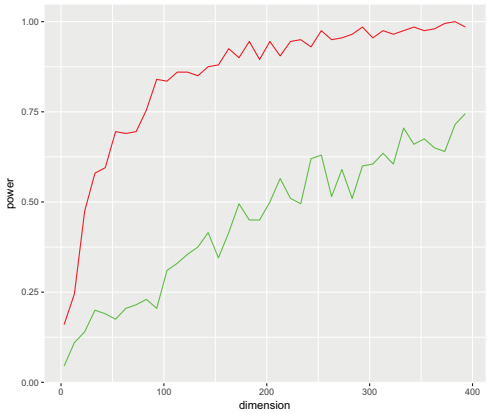
(c) Location-scale problem: $e_t \sim N(0, 1)$ and $w_t \sim N(0.2, 1.2)$.

(d) Different distributions, with equal means and variance:

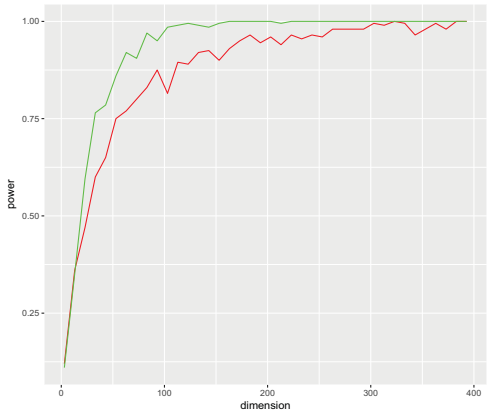
$$e_t \sim N(1, 1) \text{ and } w_t \sim Exp(1).$$
$$e_t \sim N(\frac{1}{2}, \frac{1}{12}) \text{ and } w_t \sim Unif(0, 1).$$

As we can see in Figure 4.12, the BG^3-test outperforms the BG-test in cases (b),(c),(d),(e) and (f), but does not perform very well in the location problem (a). The reason for this is similar to why the BF-test performs better in location problems, based on the simulated data. The added term $\hat{\mu}_{FF} - \hat{\mu}_{GG}$ in the BG-test which serves as noise in the location problem increases in the BG^3-test. Running the same simulation but using a permutation Monte-Carlo-bootstrap instead of a normal Monte-Carlo-bootstrap to obtain a critical value, we get the power plots in Figure 4.13. Similar to the first figure, the BG^3-test outperforms the BG-test in (d) and (e) where the samples have the same mean and variance but differ in the third moment.

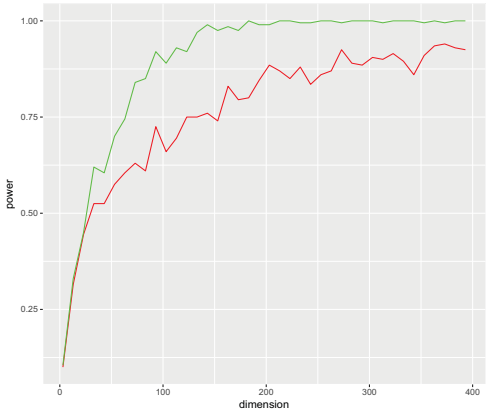
We can conclude that the combination of the results of both tests will give a better result to the general two-sample problem. If the BG-test accepts the null hypothesis, we note from the simulations that this result could be false in the same mean and variance case. Hence, we need to apply the BG^3-test to confirm the result.



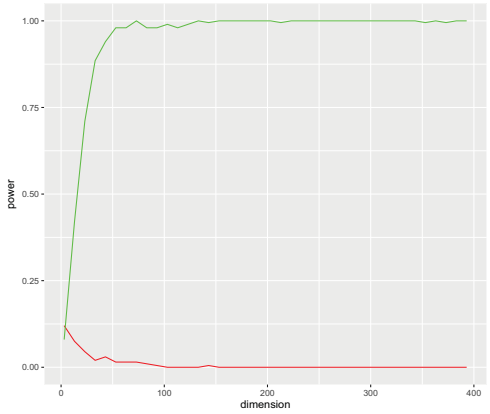
(a) Location problem: $\mu = 0.3, \sigma^2 = 1$



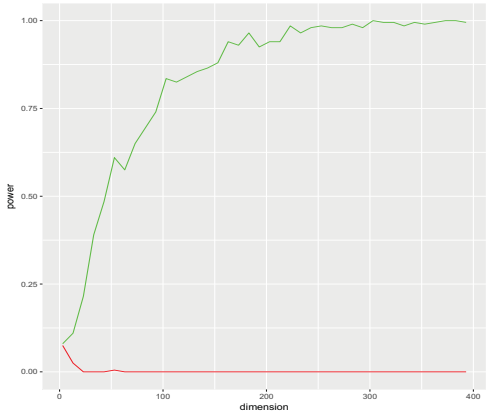
(b) Scale problem: $\mu = 0, \sigma^2 = 1.3$



(c) Location-scale problem: $\mu = 0.2, \sigma^2 = 1.2$

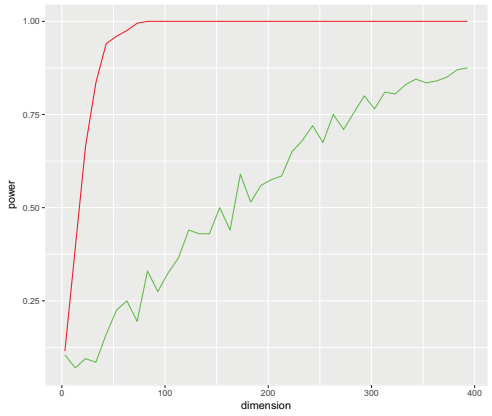


(d) Different distributions with equal means an variance: $N(1,1)$ and $Exp(1)$

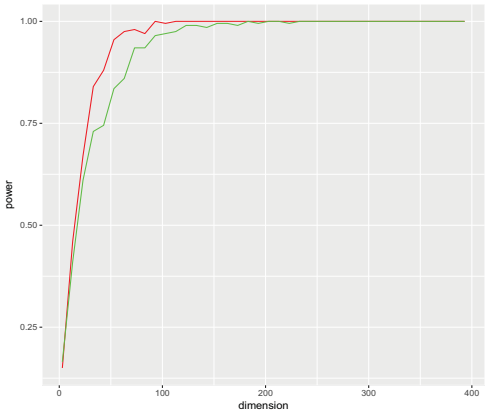


(e) Different distributions with equal means an variance: $N(\frac{1}{2}, \frac{1}{12})$ and $Unif(0,1)$

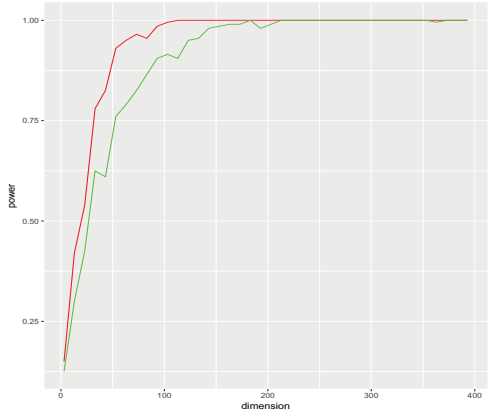
Figure 4.12: Powers of BG-test (red) and BG^3-tests (green) via normal Monte-Carlo-bootstrap



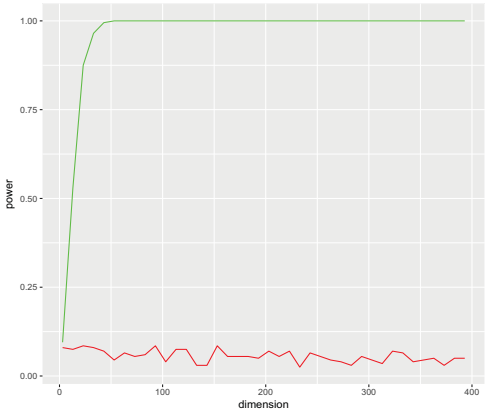
(a) Location problem: $\mu = 0.3, \sigma^2 = 1$



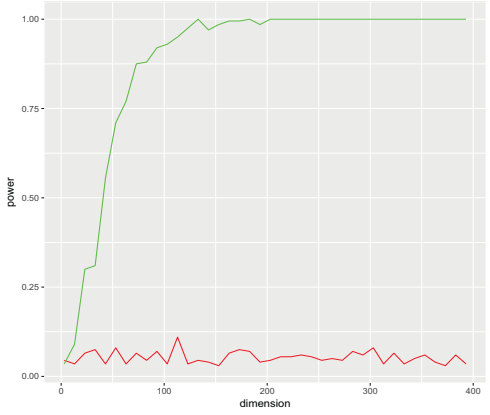
(b) Scale problem: $\mu = 0, \sigma^2 = 1.3$



(c) Location-scale problem: $\mu = 0.2, \sigma^2 = 1.2$



(d) Different distributions with equal means an variance: $N(1, 1)$ and $Exp(1)$



(e) Different distributions with equal means an variance: $N(\frac{1}{2}, \frac{1}{12})$ and $Unif(0, 1)$

Figure 4.13: Powers of BG-test (red) and BG^3-tests (green) via permutation Monte-Carlo-bootstrap

Future work

To the best of our knowledge, there is no direct Bayesian test for two-sample time series unless the noise distribution is known and the noise terms are iid. We would like to generalize the Bayesian two-sample test given by Holmes et al. [7] to the non-iid setting. Also, we would like to consider the problem in a multivariate setting, where there is more than one time series sample from each model. Another issue for future work would be to consider explicit forms of the Bayes factor using other prior distributions, for example, the Dirichlet process mixtures (DPM) as in Borgwardt et al. [3]. We are interested in comparing the performances of different Bayesian tests.

Appendix A

R Codes

We include the codes from the zinc example for s_1 vs. s_2.

Correlation comparison

Fitting an $AR(p)$ model to the gathered data of $s_1 = x1$ and $s_2 = y1$:

```
ar(x = c(x1,y1), method = "yule-walker")
```

Residual calculation of each data set based on the fitted $AR(p)$model:
For x1:

```
w<-vector("numeric",103)
for(i in 2:104){
y=c(0,x1)
w[i]<-y[i]-0.2245*y[i-1]
}


w1=w[-c(1)]
```

For y1:

```
w<-vector("numeric",104)
for(i in 2:105){
y=c(0,y1)
w[i]<-y[i]-0.2245*y[i-1]
}
```

```
w2=w[-c(1)]
```

Calculating the PACFs for the χ^2-test:

```
w<-vector("numeric",20)
for(i in 1:20){
v=c(pacf(w1,plot=F)$acf)
v1=c(pacf(w2,plot=F)$acf)
w[i]<-(v[i]-v1[i])^2/ ((1/(103-i))+(1/(104-i)))
}
```

Residual Comparison: Other Two-Sample Tests

Based on the `cramer.test` code source from the ``cramer'' package, we modify its code in the following way:

```
#require(boot)

.BG.statistic<-function(daten,indexe,mm,nn,lookup) {
    xind<-indexe[1:mm]
    yind<-indexe[(mm+1):(mm+nn)]
    (sum(lookup[xind,yind])/(mm*nn)-sum(lookup[xind,xind])/(mm^2))^2
    +(sum(lookup[xind,yind])/(mm*nn)-sum(lookup[yind,yind])/(nn^2))^2
}

BG.test<-function(x,y,conf.level=0.95,replicates=1000,sim="ordinary",
just.statistic=FALSE,kernel="phiCramer") {
    RVAL <- list(method = paste("nonparametric BG-Test with kernel",
    kernel,"\n(on equality of two distributions)"),
                 d = 0,
                 m = 0,
                 n = 0,
                 statistic = 0,
                 conf.level = conf.level,
                 crit.value = 0,
```

```
            p.value = 0,
            result = 0,
            sim = sim,
            replicates = replicates
            )
if ((is.vector(x))&&(is.vector(y))) RVAL$d<-1
if ((is.matrix(x))&&(is.matrix(y))) if (ncol(x)==ncol(y))
RVAL$d<-ncol(x)
if (RVAL$d==0) stop("types of x and y incompatible or
 inappropriate.")
if (RVAL$d==1) {
    RVAL$m<-length(x)
    RVAL$n<-length(y)
    daten<-matrix(c(x,y),ncol=1,byrow=TRUE)
} else {
    RVAL$m<-nrow(x)
    RVAL$n<-nrow(y)
    daten<-matrix(c(t(x),t(y)),ncol=ncol(x),byrow=TRUE)
}
lookup<-matrix(rep(0,(RVAL$m+RVAL$n)^2),ncol=(RVAL$m+RVAL$n))
for (i in 2:(RVAL$m+RVAL$n))
     for (j in 1:(i-1)) {
         lookup[i,j]<-sum((daten[i,]-daten[j,])^2)
         lookup[j,i]<-lookup[i,j]
     }
lookup<-eval(call(kernel,lookup))

if (just.statistic) {
    RVAL$statistic<-.BG.statistic(daten,1:(RVAL$m+RVAL$n),
    RVAL$m,RVAL$n,lookup)
} else if (sim!="eigenvalue") {
    b<-boot(data=daten,statistic=.BG.statistic,mm=RVAL$m,
    nn=RVAL$n,lookup=lookup,sim=RVAL$sim,stype="i",R=RVAL
    $replicates)
    RVAL$statistic<-b$t0
```

```
        RVAL$p.value<-1-rank(c(b$t0,b$t))[1]/(replicates+1)
        RVAL$crit.value<-sort(b$t)[round(RVAL$conf.level*RVAL
        $replicates)]
        if (RVAL$statistic>RVAL$crit.value) RVAL$result<-1
    }
        class(RVAL) <- "BGtest"
    return(RVAL)
}

print.BGtest<-function(x,...) {
    cat("\n",x$d,"-dimensional ",x$method,"\n\n")
    cat("\tx-sample: ",x$m," values        ")
    cat("y-sample: ",x$n," values\n\n")
    if (x$crit.value>0) {
        cat("critical value for confidence level ",
        format(100 * x$conf.level),"% : ",x$crit.value,"\n")
        cat("observed statistic ",x$statistic,",
         so that\n\t hypothesis (\"x is distributed as y\") is ")
        cat(ifelse(x$result==0," ACCEPTED"," REJECTED"),".\n")
        cat("estimated p-value = ",x$p.value,"\n\n")
if (x$sim!="eigenvalue") {
          cat("\t[result based on ",x$replicates,"
          ",x$sim," bootstrap-replicates]\n\n");
}
    } else {
        cat("observed statistic ",x$statistic,"\n\n")
    }
    invisible(x)
}

phiCramer<-function(x) return(sqrt(x)/2)
```

The following code is based on the location problem in the illustrative examples:
Application of the BG-test:

```
w1=rep(1000,40)
```

```
for (i in 1:40){
m=seq(3,400,10)
w1[i]=sum(replicate(200,{
 xx=replicate(20,{  x11=arima.sim(model=list(ar=0.1),n=m[i],
    innov=rnorm(m[i],mean=0,sd=1))
    x1=c(x11)})
    x=matrix(xx,byrow=TRUE,ncol=m[i])
 yy=replicate(20,{y11=arima.sim(model=list(ar=0.1),n=m[i],
    innov=rnorm(m[i],mean=0.3,sd=1))
    y1=c(y11)})
    y=matrix(yy,byrow=TRUE,ncol=m[i])
    BG.test(x,y)$result}))}
```

Application of the BG^3-test:

```
w2=rep(1000,40)
for (i in 1:40){
m=seq(3,400,10)
w2[i]=sum(replicate(200,{
 xx=replicate(20,{  x11=arima.sim(model=list(ar=0.1),n=m[i],
    innov=rnorm(m[i],mean=0,sd=1))
    x1=c(x11)})
    x=matrix(xx,byrow=TRUE,ncol=m[i])
 yy=replicate(20,{y11=arima.sim(model=list(ar=0.1),n=m[i],
    innov=rnorm(m[i],mean=0.3,sd=1))
    y1=c(y11)})
    y=matrix(yy,byrow=TRUE,ncol=m[i])
    BG.test(x^3,y^3)$result}))}
```

References

[1] L. Baringhaus, C. Franz, On a new multivariate two-sample test, Journal of Multivariate Analysis 88 (2004) 190-206.

[2] M. Biswas, A. Ghosh, A nonparametric two-sample test applicable to high dimensional data, Journal of Multivariate Analysis 123 (2014) 160-171.

[3] K.M. Borgwardt, Z. Ghahramani, Bayesian two-sample tests, eprint arXiv:0906.4032 (2009).

[4] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, New York: Springer, 2002.

[5] H.E. Daniels, The approximate distribution of serial correlation coefficients, Biometrika 43 (1956) 169-185.

[6] P. Hall, J.S. Marron, A. Neeman, Geometric representation of high dimension, low sample size data, Journal of the Royal Statistical Society Series B Statistical Methodology 67 (2005) 427-444.

[7] C.C. Holmes, F. Caron, J.E. Griffin, D.A. Stephens, Two-sample Bayesian nonparametric hypothesis testing, Bayesian Analysis 10 (2015) 297-320.

[8] G.M. Jenkins, Tests of hypotheses in the linear autoregressive model II, Biometrika 43 (1956) 186-199.

[9] R.E. Kass, A.E. Raftery, Bayes factors, Journal of the American Statistical Association 90 (1995) 773-795.

[10] D. Kwiatkowski, P.C.B. Philips, P. Schmidt, Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root, Journal of Econometrics 54 (1992) 159-178.

[11] M. Lavine, Some aspects of Pólya tree distributions for statistical modelling, Annals of Statistics 20 (1992) 1222-1235.

[12] M. Marcellino, J.H. Stock, M.W. Watson, A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series, Journal of Econometrics 135 (2006) 499-526.

[13] P. Montero-Manso, J.A. Vilar, A time series two-sample test based on comparing distributions of pairwise distances, (2016), https://aaltd16.irisa.fr/files/2016/08/AALTD16_paper_12.pdf.

[14] S. Nabeya, K. Tanaka, Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative, Annals of Statistics 16 (1988) 218-235.

[15] M. Pandya, Bayesian estimation of AR (1) with change point under asymmetric loss functions, Statistics Research Letters 2 (2013) 53-62.

[16] M.H. Quenouille, The comparison of correlations in time-series, Journal of the Royal Statistical Society Series B Methodology 20 (1958) 158-164.

[17] S.E. Said, D.A. Dickey, Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika 71 (1984) 599-607.

[18] J. Scott, Nonparametric Bayesian multiple hypothesis testing of autoregressive time series, (2008), http://www2.stat.duke.edu/~james/research/papers/08-09.pdf.

[19] P.E. Smith, R.W. Eppley, Primary production and the anchovy population in the Southern California Bight: comparison of time series, Limnology and Oceanography 27 (1982) 1-17.

[20] O. Stegle, K.J. Denby, E.J. Cooke, D.L. Wild, Z. Ghahramani, K.M. Borgwardt, A robust Bayesian two-sample test for detecting intervals of differential gene expression in microarray time series, Journal of Computational Biology 17 (2010) 355-367.

[21] X. Zhang, T. Zhang, A.A. Young, X. Li, Applications and comparisons of four time series models in epidemiological surveillance data, PLoS ONE 9 (2014) e88075.

[22] http://www.fondriest.com/environmental-measurements/parameters/water-quality/ph/#p4.

[23] http://weppi.gtk.fi/publ/foregsatlas/text/HCO3_.pdf.