Spring 2022

Penalized Regression Splines-Based Tests for Comparing Two Time Series with Unequal Lengths

Mirajul Islam
University of New Hampshire, Durham

Penalized Regression Splines-Based Tests for Comparing Two Time Series with Unequal Lengths

BY

Mirajul Islam
Department of Mathematics and Statistics
University of New Hampshire

THESIS

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy
in
Statistics

May, 2022

This thesis has been examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics by:

Dr. Linyuan Li, (Thesis Director) Professor of Statistics
Department of Mathematics and Statistics
University of New Hampshire

Dr. Ernst Linder, Professor of Statistics
Department of Mathematics and Statisticsy
University of New Hampshire

Dr. Philip Ramsey, Professor of Statistics
Department of Mathematics and Statistics
University of New Hampshire

Dr. Qi Zhang, Assistant Professor of Statistics
Department of Mathematics and Statistics
University of New Hampshire

Dr. Rita Hibschweiler, Graduate Coordinator and Professor
Department of Mathematics and Statistics
University of New Hampshire

May, 2022.

Original approval signatures are on file with the University of New Hampshire Graduate School.

Dedication

To my Parents,

to my wife Afroza Yesmin,

for being who they are.

Acknowledgment

My deepest and heartfelt gratitude is reserved for my supervisor, Prof. Linyuan Li. Thanks to you, your excellence in abstract research and diligence in accomplishing project work have encouraged me all the time in my academic research and daily life. During the course of my doctoral study, you taught me how to be patient and motivated in the work of interest. You have been the best mentor that I could have hoped for. I will be forever grateful to you for your immense support and affectionate guidance. Without you, this thesis would never have been possible.

Besides my supervisor, I would like to thank the members of my thesis committee: Prof. Ernst Linder, Prof. Philip Ramsey, Prof. Rita Hibscweiler, and Prof. Qi Zhang for their effort of evaluating this thesis. Their insightful comments have immensely improved my writing and at the same time, their encouragement motivated me to do more research work even from various perspectives.

My sincere thanks also go to the graduate coordinator, Professor Rita Hibschweiler, and the chairs of the department since my enrollment: Prof Edward Hinson and Professor Karen Graham. Their continuous support in funding my study helped me to put my focus only to study and research. A huge thanks to the distinguished faculties of this department who were my course teacher for their rigorous effort in teaching making me ready for the dissertation research.

I am extremely grateful to the associated academic and professional staff of Shahjalal University of Science and Technology in Bangladesh for granting me study leave for five consecutive

years. I would like to thank also to the office staff of this department and in the graduate school at UNH for their continuous technical support over the years of my study.

I thank all my friends who have had a crucial part in keeping me sane. You all have graced my life with your kindness. You all had kept me going when I was down and out. I couldn't have done this without your love and support.

Last but not the least, I would like to express my love and gratitude to my parents, my wife, and my son. Without their tremendous understanding and encouragement in the past few years, it would be a tough job to concentrate and complete my doctoral study.

Mirajul Islam

New Hampshire, April 2022

Abstract

The Spline-based modeling has been an established tool for parametric and nonparametric regression modeling because of its continuous progress on theoretical and computational fronts over the last three decades. This thesis explores the idea of penalized spline modeling and goodness of fit testing in the context of time series testing in the frequency domain approach. The comparison of different time series is an important topic in statistical data analysis and has various applications in scientific research. One approach to identifying similarities or dissimilarities between two stationary processes is to compare the spectral densities of both time series. This thesis examines whether two stationary and independent time series with unequal lengths have the same spectral density. A new test statistic is proposed based on penalized splines regression. It relies on penalized splines estimator of an unspecified smooth function for the log-ratio of two spectral estimates, which are obtained from averaging out of the blocked periodograms for corresponding time series. Under the null hypothesis that two spectral densities are the same, the theoretical asymptotic distribution of the test statistic is derived. Several tests have been proposed in recent years: some of them are computationally intensive, and some lack stable size. Also, some current tests have low powers. So, we examined a relatively computationally fast and consistent test using penalized splines regression which reveals stable empirical type I error and good power properties. Simulation studies show that our proposed test is very comparable to the current test statistics in almost every case. Another advantage of our proposed test statistic is that it is very simple to construct and computationally fast based on a low-rank estimation technique.

Contents

List of Acronyms

AIC	Akaike Information Criterion
ARIMA	Auto Regressive Integrated Moving average
ARMA	Auto Regressive Moving average
AR	Auto Regressive
B-spline	Basis Spline
CME	Coronal Mass Ejection
CLT	Central Limit Theorem
CV	Cross Validation
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
GAM	Generalized Additive Model
GCV	Generalized Cross Validation
LME	Linear Mixed Effect
LRT	Likelihood Ratio Test
MA	Moving average
MSE	Mean Square Error

OLS	Ordinary Least Square
PRESS	Predicted Residual Error Sum of Squares
REML	Restricted Maximum Likelihood Estimation
RKHS	Reproducing Kernel Hilbert Space
RLRT	Restricted Likelihood Ratio Test
SARIMA	Seasonal Auto Regressive Integrated Moving average
SAR	Seasonal Auto Regressive
SMA	Seasonal Moving Average
TPS	Truncated Power Series

CHAPTER 1

Introduction

When analyzing two time series, it is often useful and important to check whether they have the same underlying process. Understanding this relationship between two time series helps analysts to reveal the mechanisms of complex systems. For a stationary process, its second-order dynamics can be described by the autocovariance structure or equivalently its spectral density. The testing mechanism for assessing whether two or more time series have the same dynamics is of great interest in many fields. For example, an application of predicting cooktop fire may be the comparison of the rate of change of two sensor signals-time-series observed at different time points to understand the pattern of normal cooking and preignition time. In mechanical engineering, engineers may wish to detect damage to a mechanical system by comparing the frequency patterns of current vibration signals to those of reference signals collected under healthy conditions. In space science, it is important to compare the number of solar maximam (peak of sunspot activity) and solar minimam (the lull), produced on the outer surface of the sun throughout solar cycles, for a better understanding of the significant space weather events such as solar flares, CMES, radiation storms, and radio bursts. Many additional applications of this statistical problem in different fields can be found in the literature.

There has been an increasing research activity on time series testing since the 1990s, and the comparison of two stationary time series models has been studied in both time and frequency domain methods. Several periodogram-based tests firstly have been considered by Coates & Diggle (1986). Their tests assumed the hypothesis that two independent time series of wall

thicknesses of a gas pipe at two different locations are realizations of the same stationary process. Pötscher & Reschenhofer (1988) extended their procedure to a situation in which replicated observations are available for each process. Piccolo (1990) introduced a Euclidean distance measure between the coefficients of different ARIMA models. Graphical devices to compare periodograms were proposed by Diggle & Fisher (1991), where Kolmogorov-Smirnov or Cramér-Von Mises types of test statistics have been applied for the analysis of hormonal data. Carmona & Wang (1996) analyzed Lagrangian velocities of drifters at the surface of the ocean by a comparison of spectra. Caiado et al. (2006) proposed a new measure of distance between time series based on the normalized periodogram. They dealt with time series of unequal length by truncating the longer series to the length of the shortest one with a view to using spectral estimates to compute distances across countries.

Lund et al. (2009) considered tests in the time-domain approach to assessing whether two stationary and independent time series have the same autocovariance at all lags. They applied their tests to the analysis of temperatures and precipitations from Atlanta and Athens in Georgia in order to identify a good climatological reference series for given stations. A data-driven nonparametric test procedure, using Legendre polynomials expansion to the log spectral ratio, was proposed in Jin (2011) for comparing two or more dependent/independent and stationary time series. The classification and clustering of time series using frequency-domain methods were discussed by Caiado et al. (2012), where they used an interpolation method for periodogram to overcome the issue of unequal time series lengths. They interpolated periodogram for the longer series at the frequencies defined by the shorter series. Preuß & Hildebrandt (2013) presented a theoretical framework for testing the equality of spectral densities in the bivariate stationary processes with unequal sample sizes, and afterward, they applied it to the cluster of financial time series data of different sample lengths. An adaptive test for comparing two time series using Fan (1996)'s adaptive Neyman method was provided in Lu & Li (2013).

Eichler (2008) developed a general nonparametric approach for testing hypotheses about

the spectral density matrix of multivariate stationary time series based on estimating the integrated deviation from the null hypothesis. This approach covers quite a broad range of spectral analyses. Following Eichler (2008), Dette & Paparoditis (2009) employed a general bootstrap procedure to approximate the null distribution of nonparametric frequency-domain tests about the spectral density matrix of a multivariate time series. Based on an L-2 type test statistic, Jentsch & Pauly (2015) proposed a nonparametric approach to test the equality of spectral densities by using randomization techniques to compute the critical values. The L-2 type test statistic is one from the estimation of the minimal L2 distance between the spectral density matrix of a stationary time series and its approximation in the class of all densities satisfying the null hypothesis. Using this approach, Dette et al. (2011) developed a method that only requires an appropriate summation of the periodogram. Later, this idea was generalized by Dette et al. (2011b) to test whether a locally stationary time series is stationary. A Fourier transform-based test for comparing two independent and periodically correlated (PC) time series has been proposed by Mahmoudi et al. (2019). A computational bootstrap-based statistic via blocks of blocks bootstrap has been proposed by Jin et al. (2019) to check if two time series have the same second-order dynamics.

Most of the tests in current literature assume time series of equal lengths. But, in practice, usually one has multiple time series of unequal lengths. In order to overcome this length issue, one typically either truncates the longer time series into the length of the short one (for example, Caiado et al. (2006)) or uses interpolation as in the Caiado et al. (2009, 2012). Caiado et al. (2009) suggested several interpolated periodogram discrepancy statistics for the comparison of time series with unequal sample sizes in a simulation study and Jentsch & Pauly (2012) provided a theoretical result, which however does not yield a consistent test as it was also pointed out by the authors. Decowski & Li (2015) conducted wavelet-based tests in the frequency domain approach for comparing two independent and stationary time series with unequal lengths. Here, similar to Decowski & Li (2015)'s approach to handling unequal length issues we use the local average of consecutive periodogram values in non-

overlapping blocks as suggested in Coates and Diggle (1986). In this thesis, we consider the same approach of unequal sample sizes as in Decowski & Li (2015) with a different nonparametric testing procedure. We directly computed our test statistic using penalized regression splines method applied to observed data of the log-ratio of two spectral estimates corresponding to the two stationary and independent time series.

Recently, Jin & Wang (2016) proposed consistent test statistics in the time-domain approach to compare the autocorrelation structure of two time series. They determined the order of lags via Akaike Information Criterion (AIC) and showed that the test statistic converges asymptotically to a non-conventional distribution. Their process is time-consuming and the rate of convergence is very slow. Most recently, Li & Lu (2018) provided a test statistic based on the difference between two wavelet-based estimates of the two spectral densities through empirical wavelet coefficients. The asymptotic distributions of the empirical wavelet coefficients are derived based on a Bartlett-type approximation. Their test involved too many parameters and has unstable levels in moderately small samples. Also, the test does not consider the power calculation for seasonal time series models. More details of these previous tests, which rely on the same problem, are discussed and compared later in this thesis. The advantages of our new method over the previous tests are that the derivation of the test statistic is very straightforward and does not need too many parameters. Also, the level and power of the test are not sensitive to the size of the sample.

Consider, two stationary and independent time series $\{X_t,\ t = 1, 2, \ldots, n_1\}$ and $\{Z_t,\ t = 1, 2, \ldots, n_2\}$ with autocovariance functions $\gamma_X(h)$ and $\gamma_Z(h)$, respectively, at lag h. The spectral density of $\{X_t\}$ at fixed frequency ω is then defined as

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{-i\omega h}, \quad -\pi < \omega \leq \pi.$$

Brief details of the spectral density with its components are discussed later. Because of the symmetry of $f_X(\omega)$, the tests have the following hypothesis as in Coates & Diggle (1986)

$$
\begin{aligned}
H_0 &: \ f_X(\omega) = f_Z(\omega) \qquad \text{for all } \omega \in (0, \pi), \\
H_1 &: \ f_X(\omega) \neq f_Z(\omega) \qquad \text{for some } \omega \in (0, \pi).
\end{aligned}
\tag{1.1}
$$

This thesis is centered on a two-fold approach: first, estimating spectral densities by solving unequal length issues of respective time series, followed by a derivation of the test statistic using penalized spline regression modeling from the observed log-ratio of estimated spectral densities. For testing the significance of the regression function in the later part, smoothing spline estimators have been the subject of several research efforts. Eubank & Spiegelman (1990) considered testing the departure from linearity based on fitting cubic smoothing splines to the residuals. The same testing procedure, revised in Jayasuriya (1996), was extended to a kth order polynomial regression model. Similar smoothing spline-based tests on local polynomial regression were also introduced by Cox & Koh (1989) and by Chen (1994). Cantoni & Hastie (2002) considered a test statistic based on a mixed-effects model (LME) with a fixed smoothing parameter. Liu & Wang (2004) compared such smoothing spline-based tests as the locally most powerful test in Cox et al. (1988) generalized maximum likelihood ratio test, and the Wahba (1990) Generalized cross-validation (GCV) test. Another line of work on testing the mean function in a nonparametric regression has used local polynomial smoothing under the alternative. For example, Zhang (2004) assessed the equivalence of nonparametric tests based on smoothing splines and local polynomials and reported their equivalent asymptotic distributions under the null and the alternatives. Other related research on local polynomial smoothing includes the work of Cai et al. (2000), Fan et al. (2001), and Li & Nie (2008).

The intrinsic drawbacks of the above mentioned nonparametric approaches are (i) computationally intensive as it requires the number of knots to be equal to the number of data points (for spline smoothing) and (ii) complex (for selecting bandwidth in local polynomial

smoothing). The overarching goal of this thesis is to develop a new test statistic based on penalized regression splines for testing the preceding hypothesis (1.1) in the frequency domain, which is computationally fast and consistent in terms of its stable size and good power properties. O'Sullivan (1986) used penalized fitting with cubic B-splines for inverse problems. O'Sullivan splines are discussed by Wand & Ormerod (2008). Kelly & Rice (1990) and Besse et al. (1997) used B-spline approximations to the smoothing splines, which they called hybrid splines. Schwetlick & Kunert (1993) decoupled the order of the B-spline and the derivative in the penalty function. The same idea has been promoted by Eilers & Marx (1996), who used a combination of B-splines and a differenced penalty on the spline coefficients. The hypothesis was examined in Wand (2003) by a likelihood ratio test (LRT) through the use of penalized splines and an LME model representation. Crainiceanu & Ruppert (2004) and Crainiceanu et al. (2005) reported asymptotic distributions of the LRT or restricted likelihood ratio test (RLRT). These tests are based on the likelihood of assuming normality of the random effects and the residual errors. The smoothing parameter is taken as the ratio of two variance components and estimated by a restricted maximum likelihood (REML). Many applications and examples of penalized splines are available in Ruppert et al. (2003), and also presented by Li & Ruppert (2008), Ruppert et al. (2009), and Chen & Wang (2011).

Significant progress on the theoretical properties of penalized splines regression and its associated tests has been established through the following scholarly research work over the last 15 years. Claeskens et al. (2009) obtained asymptotic properties of penalized spline estimators and related them to known asymptotic results for regression splines and smoothing splines, which can be seen as the two extreme cases, with penalized splines situated in between. Their theoretical works fundamentally rely on some of the important theoretical results on unpenalized regression splines which are proved by Zhou et al. (1998). Wang et al. (2011) performed an asymptotic analysis of penalized spline estimators. They have shown that a P-spline and a smoothing spline are asymptotically equivalent provided that the number of knots of the P-spline is large enough, and the two estimators have the same

equivalent kernels for both interior points and boundary points. Chen et al. (2014) presented a test of nonparametric function and a test of goodness-of-fit of a polynomial model based on penalized splines. Unlike the test in Cantoni & Hastie (2002), they do not assume a fixed smoothing parameter under the alternative hypothesis, since a reasonable smoothing parameter may not be available in practice. Their proposed test is different from the tests in Crainiceanu & Ruppert (2004) and Crainiceanu et al. (2005) in that it does not rely on LME model representation, and thus relaxes the normality assumption. Finally, Xiao (2019) provided a unified study of the large sample properties of Penalized splines including O-splines using B-splines and T-splines. Here, O-splines are meant by B-splines with an integrated squared derivative penalty. P-splines, formally Penalized splines, use B-splines and a discrete penalty, and the T-splines employ truncated polynomials with a ridge penalty.

The aforementioned cited research has played a vital role in this case for formulating the test statistic and its asymptotic distribution. But in contrast, testing two or more time series through penalized regression splines is not yet explored, especially in the case of unequal lengths. The thesis is organized as follows. In Chapter 2, we describe facts and types of spline basis in detail and introduce the penalized regression splines-based estimator for the mean regression function. In Chapter 3, we illustrate how to obtain observed data from the spectral estimates of two time series of unequal/equal lengths. Then, the asymptotic distribution of the test statistic is derived from penalized splines regression coefficients. Chapter 4 includes investigations of the numerical properties of the proposed test through simulation studies, and the performance of the proposed test is also compared with the previous test. A summary and concluding remarks are given in Chapter 5. In the Appendix, we provide some supplementary materials of theoretical details and of the simulation studies.

CHAPTER 2

Penalized Regression and Splines

2.1 Why we need Penalized Regression?

There is often a tension between applying simple models that don't fit the data satisfactorily but are really easy to interpret and using black-box machine learning methods that do fit the data very well but are hard to interpret. Generalized additive models (GAMs) fit into the gap between these two extremes. It uses highly interpretable splines to model non-linear relationships between covariates and the response that are learned from the data. GAMs are simply a class of statistical models in which the usual linear relationship between the response and predictors is replaced by several nonlinear smooth functions to model and capture the nonlinearities in the data. The GAM model relates a univariate response variable, Y, to some predictor variables, X_i via a structure such as

$$g(\mathrm{E}(Y_i)) = A_i\theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \cdots \tag{2.1}$$

where, g is a link function and $Y_i \sim$ some exponential family distribution (i.e, Binomial, Normal or Poisson, etc.). Here, A_i is a row of the model matrix for any strictly parametric components, θ is a corresponding parameter vector. The functions f_j are smooth functions of the covariates to be estimated by non-parametric means. The representation of smooth functions is best introduced by considering a model containing one smooth function of one covariate Wood (2006, ch. 3),

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1,\, 2,\, \cdots,\, n, \tag{2.2}$$

where $x_i = i/n \in [0, 1]$ are known design points, the noise variables ε_i are assumed to be $i.i.d$ random variables with mean 0 and variance σ^2 and $f(.)$ is a class of unknown function to be estimated. One approach is of course to assume that f has certain shape, such as linear or quadratic, that can be estimated parametrically. The nonparametric portion of the model, f, is only assumed to lie in a general class of functions. Suppose, we estimate f by $\hat{f}_n(x)$ to minimize the mean squares error (MSE)

$$\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2. \tag{2.3}$$

Minimizing MSE over all linear functions (i.e., functions of the form $\beta_0 + \beta_1 x$) yields the least-squares estimator and a straight line fit of the response over the predictors. Scientists are, in practice, increasingly faced with the complex nonlinear functional form with high dimensional data which requires flexible statistical models that can accommodate them. Minimizing MSE for all such complex functions that interpolate the data is an overwhelming job. An efficient way to get the job done is to use a nonparametric smoothing approach, based on a class of functions called penalized regression splines. The penalized regression spline estimators enjoy similarities to both regression splines; without penalty and with fewer knots than data points, and smoothing splines; with knots equal to the data points and a penalty controlling the roughness of fit.

2.2 Choice of Spline Basis

With the introduction of GAMs by Hastie & Tibshirani (1987), the use of spline modeling has become an established tool in statistical regression analysis. This is illustrated in Figure 2.1, where a simple least square regression model would hardly give an approximation of the observed pattern since it is obvious that the relationship is not linear. A similar example was discussed in Perperoglou et al. (2019) for triceps skinfold thickness data available in Royston

& Sauerbrei (2008). The data in our case for this exemplification were generated from the model (2.2) with $x \in [0, 1]$ for 200 equispaced points, where

$$f(x) = 1.5 + \sqrt{x(1-x)} + \sin((2.5\pi(x + 0.05)) + log(0.75 + x), \text{ and}$$
$$\varepsilon \sim N\left(0, 0.20^2\right). \tag{2.4}$$

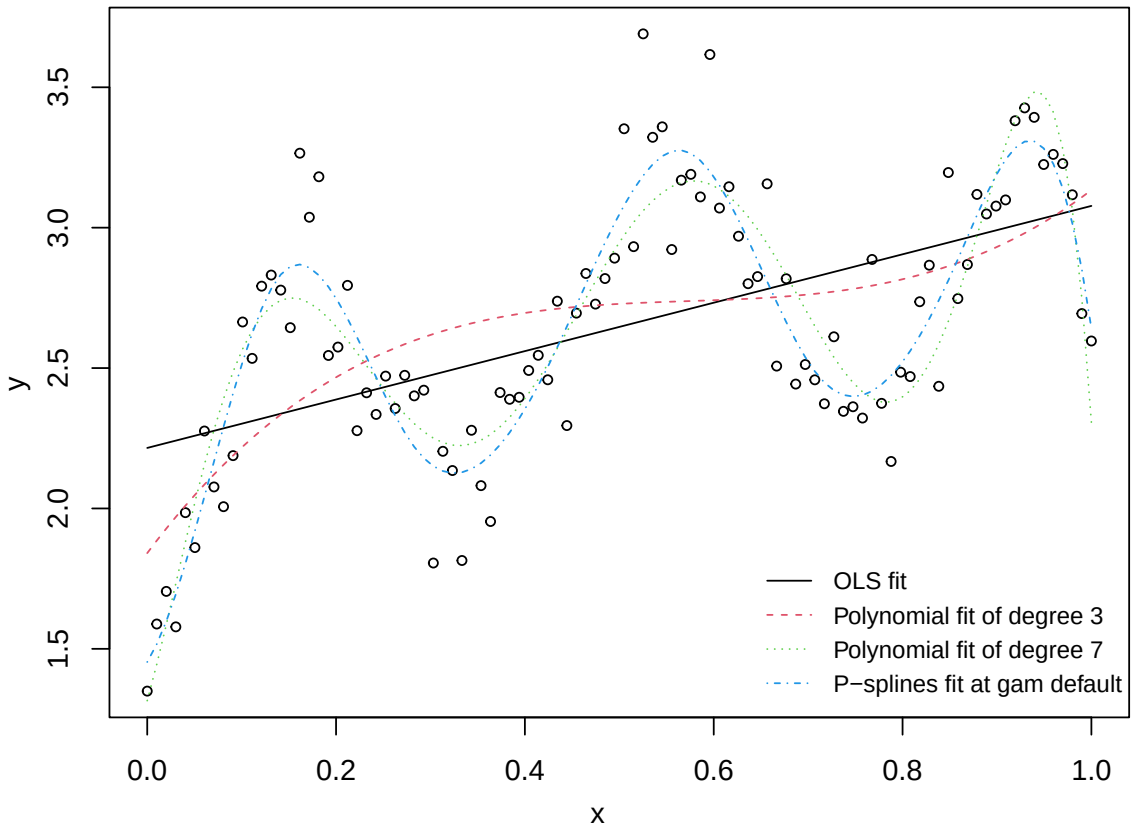

Figure 2.1: An illustration why penalized regression is important over the traditional parametric regression in approximating non-linear response-predictor relationship.

One approach for modeling the nonlinear effects is to extend the liner model by representing $f(x)$ as a collection of known basis functions: if b_j is the j^{th} such basis function, then f is assumed to have a representation

$$f(x) = \sum_{j=1}^{m} \beta_j b_j(x), \tag{2.5}$$

for some unknown parameters, β_j. Substituting (2.5) into (2.2) clearly yields a liner model, despite the fact that it provides a nonlinear function of the predictor varaibles. The model is still linear in coefficients and can be fitted using OLS methods. Suppose that f is believed to be a $(m+1)^{th}$ order polynomial, so that the space of polynomials of order m and below contains f. A basis for this space is $b_0(x) = 1, b_1(x) = x, b_2(x) = x^2, \ \cdots, b_m(x) = x^m$, so that using polynomial basis, (2.2) becomes

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \varepsilon_i.$$

However, Taylor's theorem implies that polynomial bases are useful for situations in which interest focuses on properties of f in the vicinity of a single specified point. But when the questions of interest relate to f over its whole domain, polynomial bases have undesirable side effects: each observation affects the entire curve, even for x values far from the observations. These difficulties are reflected in Figure 2.1, where a 4^{th} order polynomial poorly interpolates the response. A higher-order polynomial could be a better fit (for example, an 8^{th} ordered polynomial), but it is overwhelming and undesirably increases the predicted residual error sum of squares (PRESS) due to the fact that the lower order covariates remain in the model. Higher-order polynomials also oscillate wildly in places, where they interpolate the data at which all derivatives with respect to x are continuous. Therefore, it clearly makes sense to use bases that are good at approximating known functions in order to represent unknown functions. Bases that perform well for interpolating exact observations of a function are also good for the closely related task of smoothing noisy observations of a function. Splines are piecewise polynomials joined together to make a single smooth curve. There are several choices of spline basis which differ from one another with respect to their numerical properties (De Boor (2001), Wood (2017*a*)).

2.2.1 The Piecewise Linear Basis or Tent Basis

Let $\xi_1 < \xi_2 < \cdots < \xi_K$ be a set of K evenly spaced ordered points defined in x—called knots. The basis for piecewise linear functions of a univariate variable is determined entirely by the locations of the function's derivative discontinuities, that is by the knot locations at which the linear pieces join up. For $j = 2, \ldots, k-1$, the tent function representation of piecewise linear basis is

$$b_j(x) = \begin{cases} \frac{x-\xi_{j-1}}{\xi_j - \xi_{j-1}}, & \text{if } \xi_{j-1} < x \leq \xi_j \\ \frac{\xi_{j+1}-x}{\xi_{j+1}-\xi_j}, & \text{if } \xi_j < x \leq \xi_{j+1} \\ 0, & \text{otherwise} \end{cases}$$

For the two basis functions on the edge

$$b_1(x) = \begin{cases} \frac{\xi_2 - x}{\xi_2 - \xi_1}, & \text{if } x \leq \xi_2 \\ 0, & \text{otherwise} \end{cases}$$

$$b_k(x) = \begin{cases} \frac{x-\xi_{k-1}}{\xi_k - \xi_{k-1}}, & x > \xi_{k-1} \\ 0, & \text{otherwise} \end{cases}$$

So, $b_j(x)$ is zero everywhere, except over the interval between the knots immediately to either side of ξ_j. $b_j(x)$ increases linearly from 0 at ξ_{j-1} to 1 at ξ_j, and then decreases linearly to 0 at ξ_{j+1}. Basis functions like this, which are nonzero only over some finite intervals, are said to have compact support. Because of their shape, the b_j are often known as tent functions. Then one can represent $f(x)$ as the linear combination of this basis function and the coefficients β's followed by the linear model $Y = X\beta + \varepsilon$, where $X_{ij} = b_j(x_i)$ for, $i = 1, 2, \cdots, n$; and $j = 1, 2, \cdots, K$. See Figure 2.2, each of the tent function basis peaks to value 1 at their corresponding knot-locations. The interpolation looks plausible, but the degree of smoothness is controlled by the arbitrary choice of basis dimension. The remedy is to use a relatively large number of knots and leave the smoothness to be controlled by the

wiggliness penalty. This penalization technique is described in section 2.3 (for more details refer to Wood (2017*a*)). The piecewise linear bases can also be improved by spline bases having continuity of just a few derivatives.

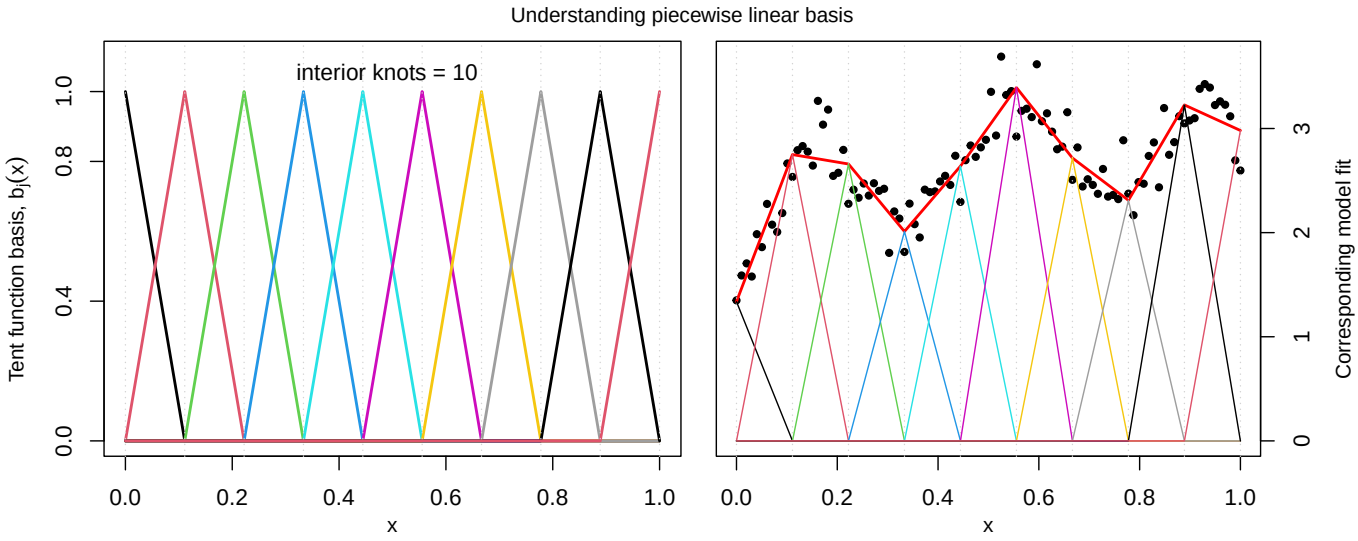

Figure 2.2: The left panel shows an example of piecewise linear basis functions of the covariate x. The right panel is the basis functions each are multiplied by corresponding coefficients, before being summed to provide interpolant, represented as thick redline (data are the black dots). Data were generated from equation (2.4).

2.2.2 Truncated Polynomial Basis

A piecewise linear spline basis makes the function linear within every interval by two successive knots, so it is not smooth as a quadratic spline basis ($m = 2$). Quadratic spline tends to do a better job of fitting peaks and valleys in a scatter plot. However, if one uses enough knots and penalized least squares, then the difference between a linear and quadratic spline fit is usually negligible. We can obtain more flexible smooth curves by increasing the degree of the spline and/or by adding knots. However, there is a trade-off (i) using a fewer number of knots and low degree results in a biased function, and (ii) a high dimension of spline basis (too many knots) increases the risk of overfitting (high variance). The most commonly used splines in practice are piecewise cubic splines ($m = 3$). It is claimed that cubic splines are the

lowest order spline for which the discontinuity at the knots cannot be noticed by the human eye. A cubic spline is a continuous function f such that (i) f is a cubic polynomial over the intervals, being made of two successive knots, and (ii) f has continuous first derivatives and square-integrable second derivatives at the knots. Cubic splines contain $K+4$ degrees of freedom: $(K+1)$ regions $\times$ 4 parameters per region minus K knots $\times 3$ constraints per knot. For example, a cubic spline function with three knots will have 7 degrees of freedom.

An $(m+1)^{th}$ order spline is a piecewise m degree polynomial with $m-1$ continuous derivatives at the knots. Define $T_1(x) = 1,\ T_2(x) = x,\ T_3(x) = x^2, \cdots,\ T_{m+1}(x) = x^m$ and $T_j(x) = (x - \xi_{j-m-1})_+^m\ for\ j = m+2, \cdots, K$. Then the functions $\{T_1, T_2, \cdots, T_{K+4}\}$ form a basis for the set of cubic splines, at K equally spaced knots $\xi_1, \cdots, \xi_K$, called the truncated power series (TPS) basis of order 4. A truncated power base spline is $m-1$ times differentiable at the knots and has $K+m$ degrees of freedom. Thus the cubic spline model with truncated power basis is

$$f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + \sum_{j=5}^{K+4} \beta_j (x - \xi_{j-4})_+^3,$$

where $(.)_+$ denotes the positive portion of its argument such that

$$(x-\xi)_+ = \begin{cases} x - \xi, & if\ x \geq \xi \\ 0, & if\ x < \xi \end{cases}.$$

An unpenalized model fit (on the right) for linear and cubic TPS basis functions along with a cubic TPS basis (on the left) is presented in Figur 2.3. Some examples of this basis functions were documented also in Eilers & Marx (2010), Hastie et al. (2001) and Ruppert et al. (2003). The truncated power bases are useful for understanding the mechanics of spline-based regression, and they can be used in practice with their conceptual simplicity if the knots are selected carefully or a penalized fit is used. Unfortunately, they have a number of numerical flaws. The supports of TPS functions are not local, with some of the bases defined over the whole range of x. This might lead the model matrix to be nearly singular

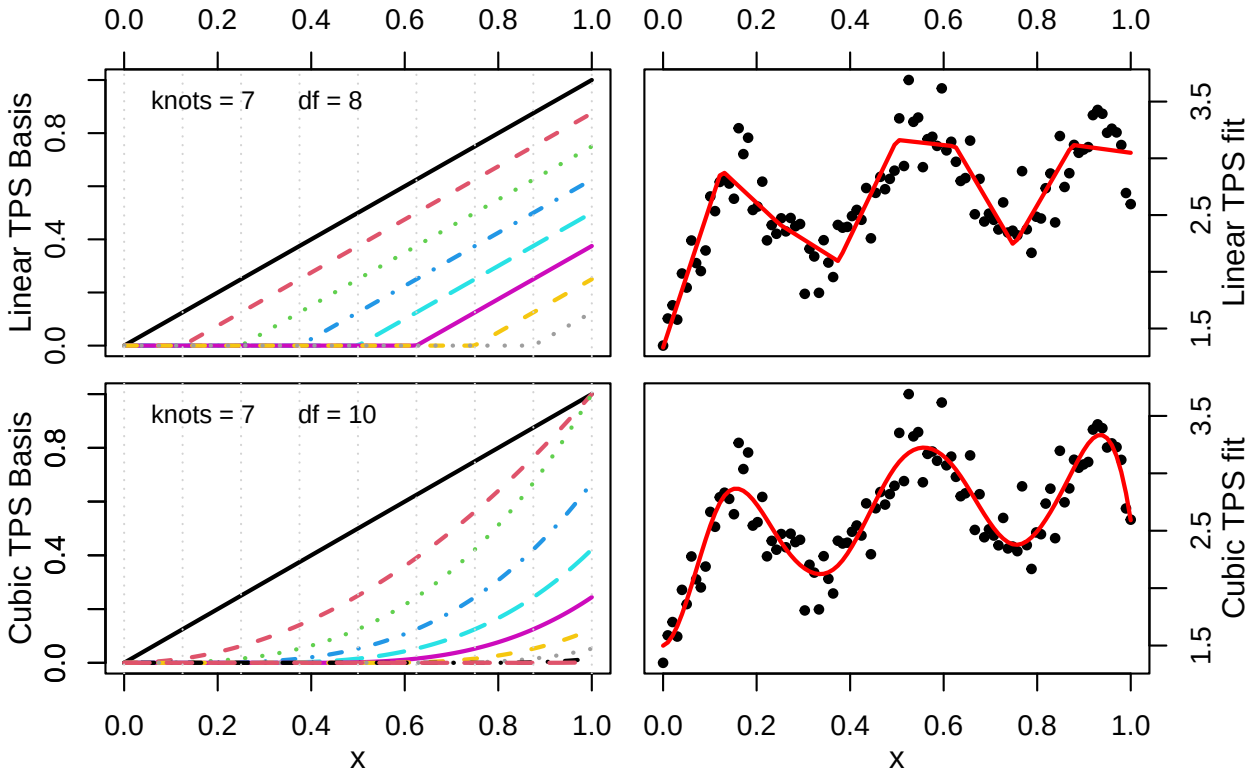

Figure 2.3: Understanding truncated power series basis. The continuous lines (top left) shows an truncated polynomial basis functions of degree 1. The top right plot is the model fit using corresponding basis, displayed by a thick red line, of the data (shown as black dots). The lower panel exhibits cubic truncated polynomial basis (bottom left), and the model fit of the same basis along with observed data (bottom right plot). Data were generated from equation (2.4).

that they are far from orthogonal due to the high correlations between some basis splines. Therefore, in practice, especially for OLS fitting, it is advisable to work with orthogonal and equivalent bases which allow more stable numerical properties.

2.2.3 B-spline Basis

The B-spline basis is a commonly used spline basis that is based on special parameterization of cubic splines (and splines of higher or lower order). The B-spline basis is appealing because the basis functions are strictly local −each basis function is non zero in intervals spanned by $m+2$ adjacent knots. Also, they have compact support which makes it possible to speed up calculations, see Hastie et al. (2001). B-spline curves are composed of many polynomial pieces and are therefore more versatile than Bézier curves (for details of Bézier curves and B-spline formation, please see Racine (2014)). In a B-spline basis, the K interior knots are bounded by the two exterior knots $\xi_0 = x_{(1)}$ and $\xi_{K+1} = x_{(n)}$ such that $\xi_0 \leq \xi_1 < \xi_2 < \cdots < \xi_K \leq \xi_{K+1}$. Now, define m additional knots (where, m being the degree of the B-spline, basis) such that

$$\xi_{-m}, \cdots, \xi_{-1} \leq \xi_0, \text{ and } \xi_{K+1} \leq \xi_{K+2} \leq \cdots \leq \xi_{K+m}.$$

The choice of extra knots is arbitrary; usually one takes $\xi_{-m} = ... = \xi_{-1} = \xi_0 = x_{(1)}$ and $x_{(n)} = \xi_{K+1} = \xi_{K+2} = \cdots = \xi_{K+m}$. Due to the recursive nature of B-spline the extra knots are added m times on either side of the boundary knots. Then for each of the augmented knots $\xi_i, i = 0, \cdots, K + 2m + 1$, ($m = 3$ for cubic spline) we recursively define a set of real-valued functions $B_{i,m}$ as follows:

$$B_{i,1} = \begin{cases} 1 & if\ \xi_i \leq x < \xi_{i+1} \\ 0 & otherwise \end{cases} \ ;\ \text{for } i = 0, 1, \cdots, K + 2m \quad \text{and}$$

$$B_{i,m+1} = \frac{x - \xi_i}{\xi_{i+m} - \xi_i} B_{i,m} + \frac{\xi_{i+m+1} - x}{\xi_{i+m+1} - \xi_{i+1}} B_{i+1,m} \quad for\ i = 1, \cdots, K + m + 1$$

Here, the convention $0/0 = 0$ and the recurrence relation is called the de Boor recurrence

relation, discovered by Carl de Boor (De Boor (2001)). With the use of additional knots, this gives $K+m+1$ basis functions. The functions $\{B_{i,m+1},\ i = 1, \cdots, K+m+1\}$ formed a basis for the set of m degree splines, which we call i^{th} B-spline basis functions of order $(m+1)$, The first term $B_{0,m}$ is often referred to as the 'intercept'. In typical spline implementations the intercept term is being dropped to avoid multiconearity issues.

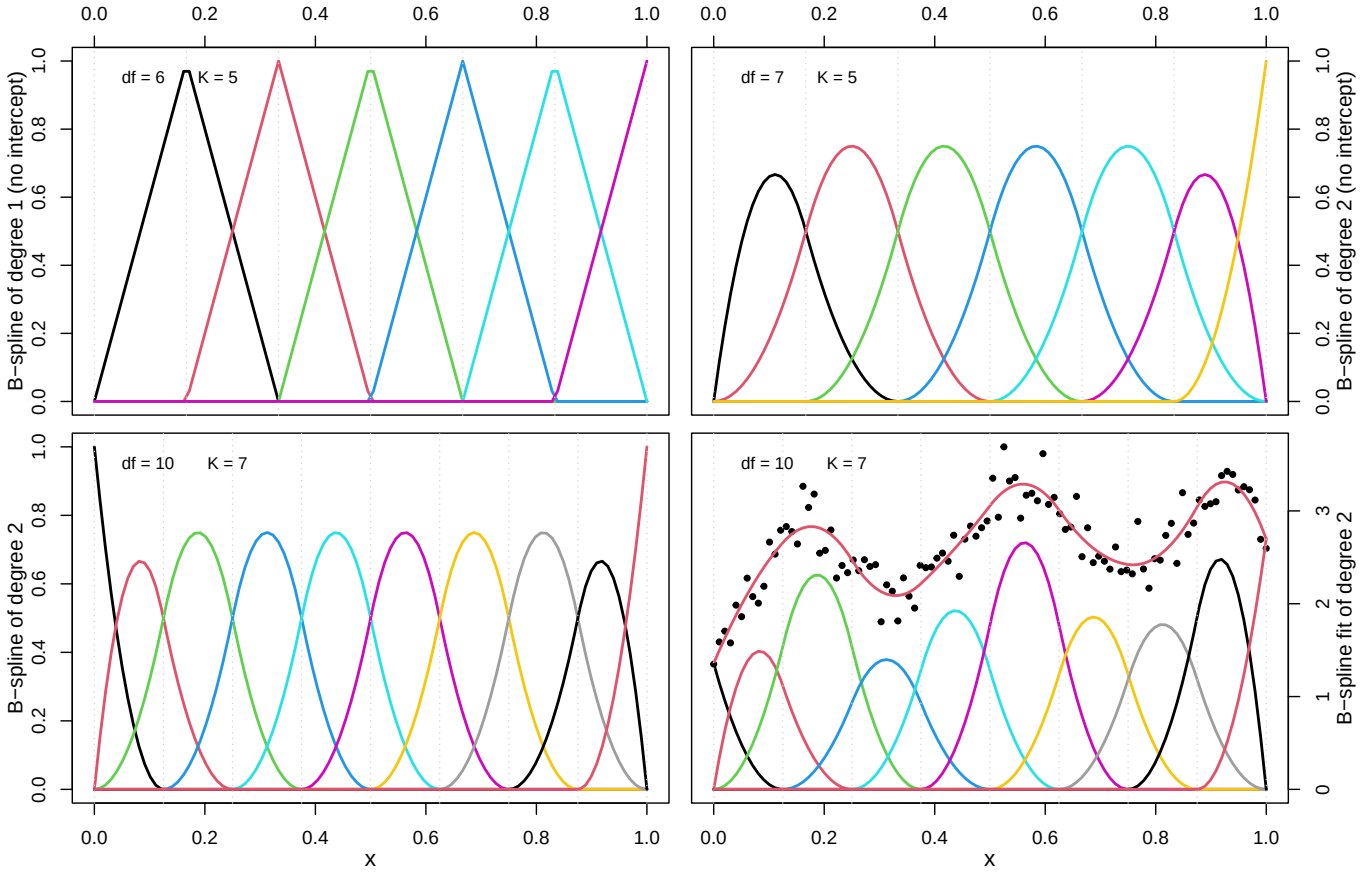

Figure 2.4: Illustration of B-spline basis functions. Top left: An intercept free 1^{st} degree B-spline basis functions. Top rigth: B-spline basis functions of degree 2 (without intercept). The bottom left plot represents a 2^{nd} degree B-spline basis. The bottom right plot illustrates how this basis functions are employed to give the fitted values at given x. Data were generated from equation (2.4).

Figure 2.4 illustrates B-splines of degree1, degree 2 and degree 3 for K interior knots. A first-degree B-spline consists of two linear pieces and is based on three adjacent knots, a quadratic B-spline is spanned by 4 adjacent knots and consists of three quadratic pieces. In

general, an m-degree B-spline is spanned by $m+2$ knots and made of $m+1$ polynomial pieces. B-splines also overlap each other. First-degree B-splines overlap with two neighbors, second-degree B-splines with four neighbors, and so on. The detailed properties of B-splines are discussed in Eilers & Marx (1996) and Racine (2014). B-spline was developed as a stable basis for large-scale spline interpolation which leads to a penalty matrix that is very computationally intensive (see De Boor (1978)).

2.2.4 Natural Cubic Spline

A polynomial spline such as a cubic spline or a B-spline can be erratic at the boundaries of the data. To address this issue, a commonly used modification of the cubic splines is the natural cubic splines. Natural cubic splines have the additional constraints that they are linear in the tails of the boundary knots $(-\infty, x_{(1)}]$ and $[x_{(n)}, +\infty)$. The linearity is enforced through the constraints that the spline function satisfy $f''(x) = f'''(x) = 0$ at the boundary knots. A natural spline cubic spline with K knots has $K+1$ degrees of freedom. A simple but popular functional form of natural cubic spline basis resulting from generalized smoothing splines based on reproducing kernel Hilbert spaces (RKHS) is discussed in the books by Wahba (1990) and Gu (2002). For such a spline basis functions, the model matrix becomes

$$N = \begin{pmatrix} 1 & x_1 & N_1(x_1, \xi_1) & \cdots & N_K(x_1, \xi_K) \\ 1 & x_2 & N_1(x_1, \xi_1) & \cdots & N_K(x_2, \xi_K) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & N_1(x_n, \xi_1) & \cdots & N_K(x_n, \xi_K) \end{pmatrix}.$$

Here, $N_j(x_i, \xi_j) = R(x_i,\ \xi_j),\ for\ j = 1, 2, \cdots, K$, where

$$R(x,z) = \frac{1}{4}\left[(z-0.5)^2 - \frac{1}{12}\right]\left[(x-0.5)^2 - \frac{1}{12}\right] - \frac{1}{24}\left[(|x-z|-0.5)^4 - \frac{1}{2}(|x-z|-0.5)^2 + \frac{7}{240}\right].$$

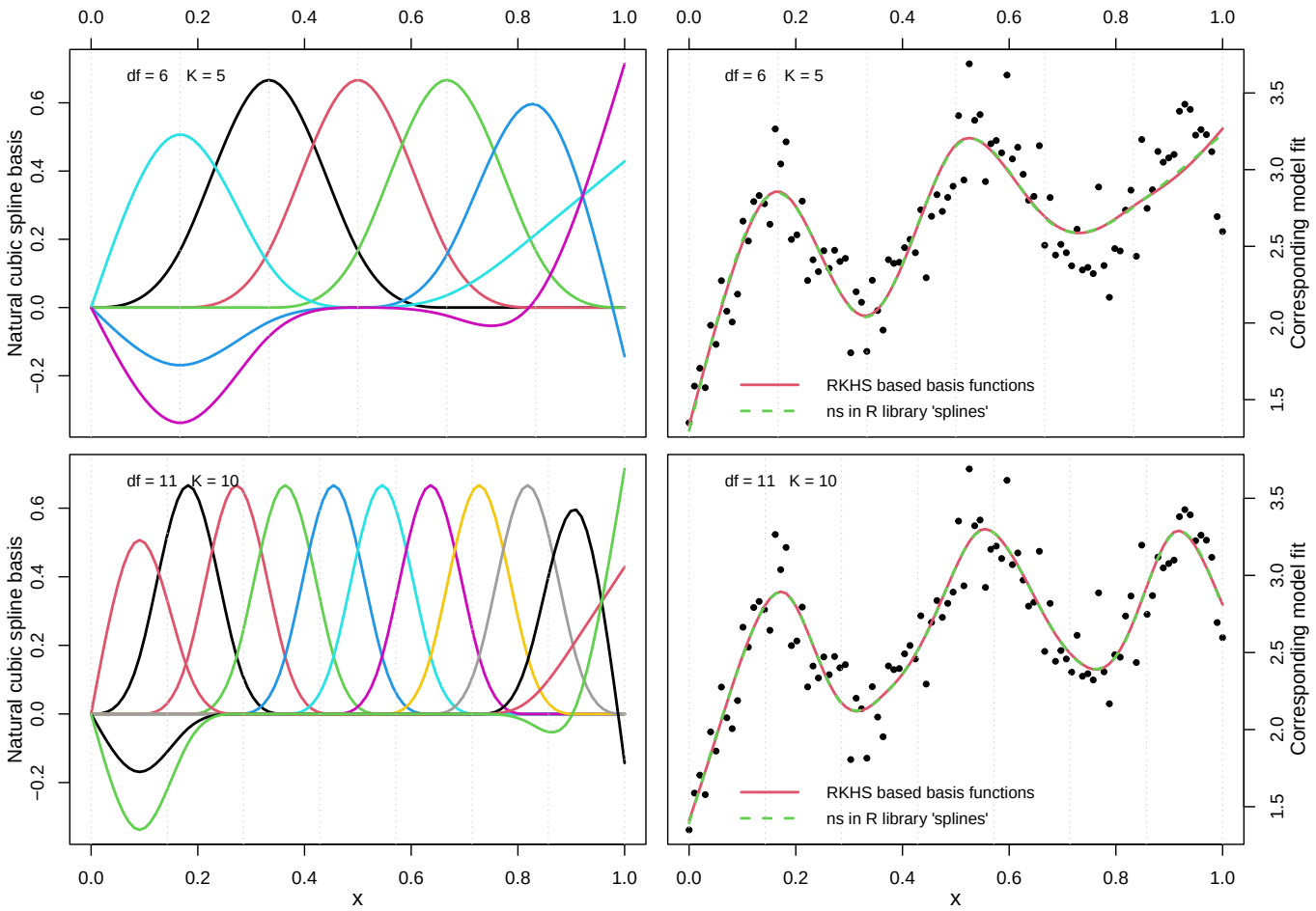

Figure 2.5: Understanding natural cubic splines. The left panel shows the basis function using the command ns in library splines. The plots on the right represent the regression fit from the corresponding spline basis. Data were generated from equation (2.4).

With these basis functions, a reasonable fit of the data can be achieved with a good selection of basis dimensions, see Figure 2.5 on the right panels. The figure also pictured the cubic natural splines (left pannels) from the library 'splines' for different interior knots along with regression fit on the right. The spline basis based on RKHS approach behaves the same as the R command **ns** in the package **splines** in capturing the response predictor relationship.

2.3 Penalized Regression Splines

The splines described so far in the previous section are often referred to as regression splines without the wiggliness penalty. In addition to the choice of the spline basis, deciding an optimal number of knots and their positions is fundamentally crucial and has a large effect on fit. A large number of knots implies more wiggliness of the estimates which may result in overfitting the data at hand. Conversely, a small number of knots may result in biased estimates. A popular approach to facilitate the choice of the knot positions in spline modeling is the use of penalized regression splines. For an i.i.d. sample of data $(x_1, y_1), \ldots (x_n, y_n)$, a penalized spline is the solution to the problem

$$\hat{\beta} = arg \max_{\beta} [l_\beta\{(x_1, y_1), \ldots, (x_n, y_n)\} - \lambda \cdot J_\beta], \tag{2.6}$$

where l_β denotes the log-likelihood β's and J_β is a roughness penalty. Adding a penalty term to the criterion we are optimizing is sometimes called regularization. Generally, penalized splines are based on the idea that the unknown function f is modeled by a spline with a large number of knots, allowing for a high degree of flexibility. But higher flexibility attains a high value of l_β which is subject to a large value of smoothness penalty. So, the solution is a trade-off between smoothness and model fit that is controlled by the smoothing parameter $\lambda \geq 0$ for which the problem in (2.6) becomes minimizing the penalized sums of squares such that

$$\hat{\beta} = arg\min_{\beta} \left[\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \cdot J(f) \right] = arg\min_{\beta} M(\lambda). \tag{2.7}$$

The roughness penalty $J(f)$ here for penalized spline of $f(x)$ will be a second ordered derivative based penalty on the fitted function, i.e., $J(f) = J_\beta = \int_x \{f''(x)\}^2 dx$. Therefore, penalized regression splines minimizes

$$M(\lambda) = \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int_x \{f''(x)\}^2 dx. \tag{2.8}$$

For any given λ, it can be shown that the solution is a natural cubic spline with knot locations that are naturally chosen at the ordered unique data values of x i.e., $x_(1), \cdots, x_(n)$. This type of natural cubic spline is referred to as smoothing spline, Green & Silverman (1993). When $\lambda = 0$, the solution of smoothing spline is interpolating the data. When $\lambda = \infty$, $\hat{f}_n$ converges to the least-squares line. The solution of $\hat{f}_n$ for $0 < \lambda < \infty$ turns out to be surprisingly simple for the class of penalized regression splines. Penalized regression splines are natural cubic splines that minimize (2.8) by using a relatively smaller number of knots placed at quantiles of the covariate x. Since the number of knots in penalized regression splines is less than the data points, the estimation process is computationally faster than the smoothing splines method which requires the number of knots to be equal to the unique data points. Some commonly used penalized regression splines are discussed in the subsequent subsections.

2.3.1 Penalized Regression with TPS splines: T-splines

The penalized regression estimators using truncated polynomial bases with a ridge penalty on the fitted curve are called T-spline estimators. Then T-splines is the estimator

$$\hat{f}_{tp} \equiv \arg\min_{\beta \in \mathbb{R}K} \left[\frac{1}{n} \sum_{i=1}^{n} \left\{ y_i - T\left(x_i\right) \beta \right\}^2 + \lambda_{tp} \beta^\top \tilde{D}_{K,m} \beta \right],$$

where λ_{tp} is a smoothing parametr for T-splines and $\tilde{D}_{K,m} = diag(\mathbf{0}_{m+1\times m+1}, I_K)$; I stands for an identity matrix. For penalized regression with TPS basis functions, the model matrix, say for $K = 4$ and $m = 2$, becomes

$$T(x) = [T_1(x) \quad T_2(x) \cdots T_{K+3}(x)], \text{ where}$$
$$T_j(x) = [T_j(x_1) \quad T_j(x_2) \cdots T_j(x_n)]^\top; \text{ for } j = 1, 2, \cdots, K+3.$$

Again, the corresponding ridge penalty is

$$\tilde{D}_{K,m} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The T-spline smooth function also can be generalized from B-splines bases for O-splines through an invertible transformation matrix, for details please see Xiao (2019).

2.3.2 Penalized Regression with B-splines: O-splines

The penalized regression splines estimator with the B-spline basis functions is the minimizer of the objective function

$$S = \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=1}^{K+m+1} \beta_j B_{j,m+1}(x_i) \right\}^2 + \lambda \int_x \left[\left\{ \sum_{j=1}^{K+m+1} \beta_j B_{j,m+1}(x) \right\}^{(d)} \right]^2 dx. \tag{2.9}$$

The penalty in (2.9) is the integrated squared d^{th} order derivative of the spline function for $d = m-1$, and it is written as $\lambda\beta^t\Delta_d^t R\Delta_d\beta = \lambda\beta^t D_d\beta$, where the matrix R has the elements

$R_{i,j} = \int [B^{(d)}_{j,m+1}(x)][B^{(d)}_{j,m+1}(x)]dx$ and Δ_d is the matrix corresponding to the weighted difference operator (for details refer Claeskens et al. (2009)). Now, define a m-degree B-spline basis functions of $1 \times (K + m + 1)$ dimension as $B(x) = \{B_{1,m+1}(x), \ldots, B_{K+m+1,m+1}(x)\}$, and $n \times (K+m+1)$ spline design matrix $B = \left\{B\left(x_1\right)^t, \ldots, B\left(x_n\right)^t\right\}^t$, and let $D_d = \Delta_d^t R \Delta_d$. With this notation, the penalized spline estimator takes the form of a ridge regression estimator as

$$\hat{f} = B\left(B^t B + \lambda D_d\right)^{-1} B^t Y, \tag{2.10}$$

where $\hat{f} = \left\{\hat{f}\left(x_1\right), \ldots, \hat{f}\left(x_n\right)\right\}^t$ and $Y = (Y_1, \ldots, Y_n)^t$. To make the result less flexible, O'Sullivan (1986) introduced a penalty on the second derivative of the fitted curve. The integrated squared of this derivative penalty has become common as a smoothness penalty, since the seminal work on smoothing splines by Reinsch (1967). The estimator in (2.10) with the O'Sullivan penalty has been considered by Wand & Ormerod (2008), who gave it the name O'Sullivan spline, or just O-spline, estimator. The problem lies in the fact that the squared second order and higher-ordered penalty for B-spline basis functions lead to complex mathematics which requires intense computational algorithm.

2.3.3 P-splines

The real statistical interest in B-spline has resulted from the work of Eilers & Marx (1996, 2010) in using them to develop what they term P-splines. P-splines are, low-rank smoothers, created using B-spline bases, usually of equally spaced knots defined on the range of covariate x, but directly impose a second ordered differenced penalty on the coefficients of adjacent B-splines. Eilers & Marx (1996) approximated the d^{th} order derivative based penalty for B-spline bases in (2.9) for K knots within along the range of x as

$$J_\beta = \int_x \{f''(x)\}^2 dx \equiv \sum_{j=1}^{K+2} \left(\Delta^d \beta_j\right)^2 = \tilde{\beta}^\top D_q \tilde{\beta}; \text{ where } q = K + 2. \tag{2.11}$$

Remember that

$$\begin{aligned}\Delta\beta_j &= \beta_{j+1} - \beta_j,\\ \Delta^2\beta_j &= \Delta\left(\Delta\beta_j\right) = \beta_{j+2} - \beta_{j+1} - \left(\beta_{j+1} - \beta_j\right)\\ &= \beta_{j+2} - 2\beta_{j+1} + \beta_j,\end{aligned}$$

and so on for higher d. Mostly $d = 2$ or $d = 3$ is used. So, if one decide to penalize a squared second ($d = 2$) ordered difference penalty between adjacent β_j, then the penalty is

$$\sum_{j=1}^{K}\left(\beta_{j+2} - 2\beta_{j+1} + \beta_j\right)^2 = \tilde{\beta}^t D_q \tilde{\beta} = \tilde{\beta}^t P_q^t P_q \tilde{\beta}$$

where

$$\begin{bmatrix}\beta_1 - 2\beta_2 + \beta_3\\ \beta_2 - 2\beta_3 + \beta_4\\ \beta_3 - 2\beta_4 + \beta_5\\ \beta_4 - 2\beta_5 + \beta_6\\ \beta_5 - 2\beta_6 + \beta_7\end{bmatrix} = \begin{bmatrix}1 & -2 & 1 & 0 & 0 & 0 & 0\\ 0 & 1 & -2 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & -2 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & -2 & 1 & 0\\ 0 & 0 & 0 & 0 & 1 & -2 & 1\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\\ \vdots\end{bmatrix} = P\tilde{\beta}.$$

By definition, the square root penalty matrix P is a $K \times q$ matrix, where q=K+2 being the degrees of freedom of P-splines. Hence, then the $q \times q$ penalty matrix becomes

$$D_q = \begin{bmatrix}1 & -2 & 1 & 0 & 0 & 0 & 0\\ -2 & 5 & -4 & 1 & 0 & 0 & 0\\ 1 & -4 & 6 & -4 & 1 & 0 & 0\\ 0 & 1 & -4 & 6 & -4 & 1 & 0\\ 0 & 0 & 1 & -4 & 6 & -4 & 1\\ 0 & 0 & 0 & 1 & -4 & 5 & -2\\ 0 & 0 & 0 & 0 & 1 & -2 & 1\end{bmatrix}.$$

The above layout for matrix display is shown for $K = 5$ knots, so that the matrices P and D_q are 5×7 and 7×7 matrices. respectively. Such discrete differenced penalties in P-splines are very easy to generate in R. For example the above matrix P is computed by the R command $diff(diag(7), differences = 2)$. The P-splines are numerically stable in that they have no boundary effects, conserve means and variances of the data, and have polynomial curve fits as limits, Eilers & Marx (1996). The boundary issues is demonstrated in Figure 2.6 between B-splines and P-splines basis functions.

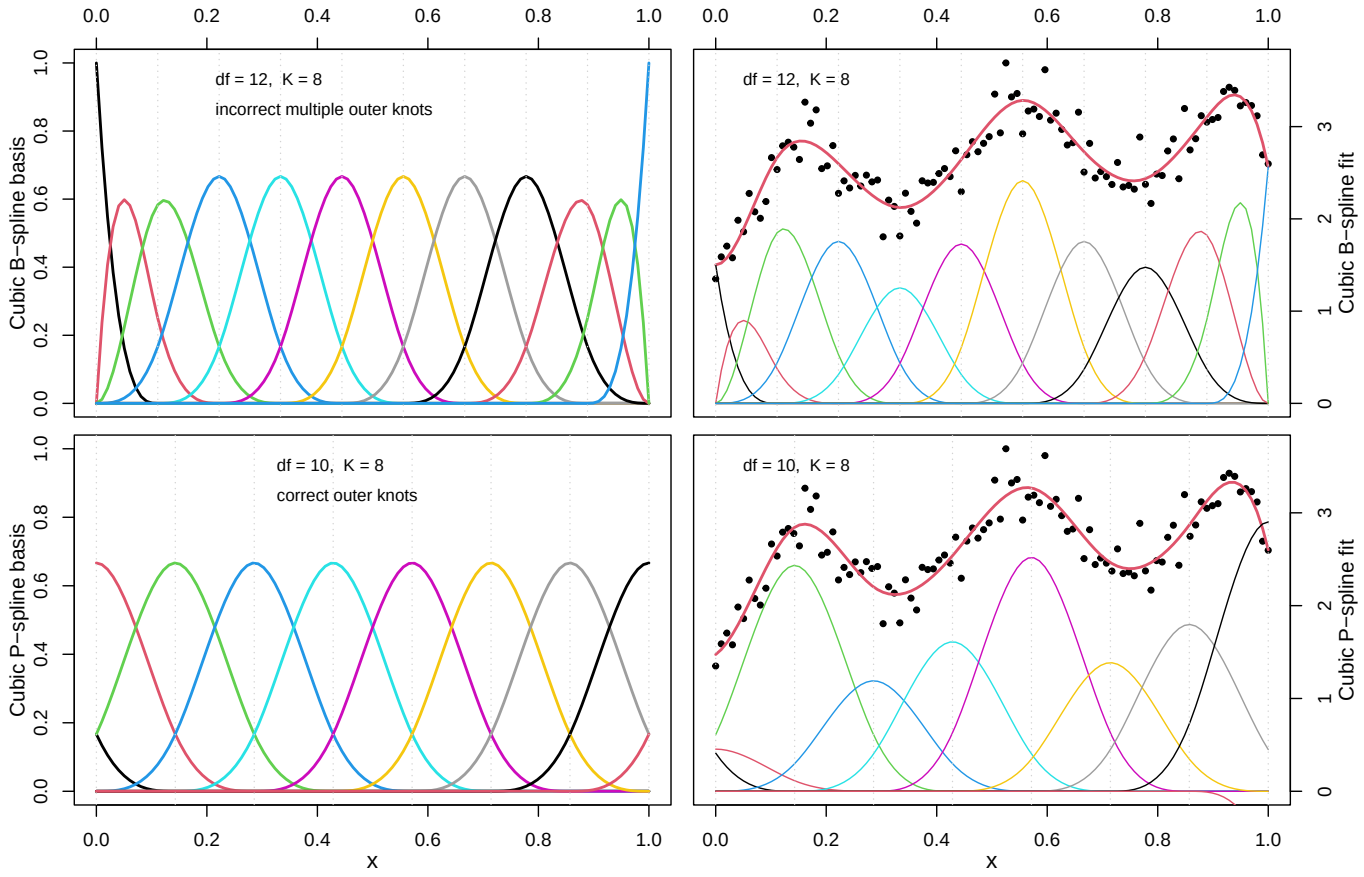

Figure 2.6: P-splines vs. B-splines with different choices of knots. The top panel shows B-spline bases with corresponding curve fitting, where multiple knots at both ends of the domain pose boundary problems, which is the result of the R function bs() and is unsuitable for penalized splines. The Bottom plots, on the other hand, represent proper basis functions for P-splines and corresponding fit, for equally spaced knots, without having any boundary issues. Data were generated from equation (2.4).

P-splines have three properties that make them very popular as reduced rank smoothers

(Wood (2017*b*)): (i) the basis and the penalty are sparse, enabling efficient computation, especially for Bayesian stochastic simulation; (ii) it is possible to flexibly 'mix-and-match' the order of B-spline basis and penalty, rather than the order of penalty controlling the order of the basis as in spline smoothing; (iii) it is very easy to set up the B-spline basis functions and penalties. The discrete penalties are somewhat less interpretable in terms of function's shape than the traditional derivative-based penalties, but tend towards penalties proportional to traditional spline penalties in case of large basis size. However part of the point of P-splines is not to use a large basis size. In addition, for solving functional optimization problems discrete penalties for smoothing may not always be desirable. If one uses the K inner knot B-spline basis functions for the collection of P-spline bases $\{N_j\}_{j=1}^{K}$, then he/she could follow a 6-steps algorithm, described in Wood (2017*a*,*b*), to obtain the derivative-based penalty matrix D_q.

2.4 Parameters Involved in Penalized Regression Splines

There are two parameters involved in the penalized regression spline for regularization of the regression curves: the basis dimension which depends on the number of knots, and the smoothing parameter. In smoothing splines, the location of knots is fixed at the unique values of the independent variable. But, for P-splines the number and the position of knots are not fixed. However, to overcome this problem to a certain degree the philosophy is to use a large number of knots at the quantiles of the covariate x and then let λ control the amount of smoothness. A popular way to determine the optimal value of λ is to use generalized cross-validation (GCV) (described after the next section), or a mixed model representation via the REML approach. GCV is the default method for optimal λ in the **gam** command in R language under the **mgcv** package.

2.4.1 Deciding on the Number of Knots, K

Because smoothing is controlled by the penalty parameter the number of knots is not crucial. There are no standards of how large the K should be, but after a minimum necessary number of knots has been reached, further increases in K often have little effect on the fit. Ruppert (2002) focused their study on selecting the number of knots for penalized splines jointly with penalty parameters by the GCV approach. They discussed two algorithms for the automatic selection of the number of knots. The myopic algorithm stops when no improvement in the GCV statistic is noticed with the last increase in the number of knots. The full search algorithm examines all candidates in a fixed sequence of possible numbers of knots and chooses the candidate that minimizes GCV. For the detail of how these two algorithm works along with different applied examples, please see Ruppert (2002). Since the algorithms used in their study cannot choose 35 knots, they ended up with a solution to the default number of knots, $K = min(n/4, 40)$, instead of the simple default, $K = min(n/4, 35)$, stated by Wand (2000). Ruppert (2002) also suggested that for a larger signal-to-noise ratio, the larger number of knots is preferable, say, 40. So, considering the larger signal-to-noise ratio for some time series models, in our simulation studies we consider $K = min(n/4 + 5, 40)$.

2.4.2 Estimation of the Smoothing Parameter, λ

If λ is too high then the data will be over-smoothed, and if it is too low then the data will be under-smoothed: in both cases, this will mean that the estimate $\hat{f}$ will not be close to the true function. Ideally, it would be good to choose an optimal λ so that $\hat{f}_n$ is as close as possible to f. A suitable criterion to choose λ is to minimize the mean residual sums of squares, called the risk estimate, by using the leave-one-out cross-validation (CV) score.

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{f}_n^{[-i]}(x_i))^2,$$

where, $\hat{f}_n^{[-i]}$ be the model fitted to all data except i^{th} Y value. It is computationally inefficient

to calculate the risk by leaving out one datum at a time, and fitting the model to each of the n resulting data set Wood (2006, p. 131). Fortunately, the ordinary CV score, $\hat{R}(\lambda)$, can be written as

$$CV(\lambda) = \hat{R}(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(\frac{Y_i - \hat{f}_n(x_i)}{1 - L_{ii}})^2, \tag{2.12}$$

where L_{ii} is the i^{th} diagonal element of the influence matrix. The smoothing parameter λ can then be chosen by minimizing $\hat{R}(\lambda)$. Rather than minimizing the ordinary CV score in 2.12, an alternative is to use an approximation called GCV in which each L_{ii} is replaced with its average $n^{-1}\sum_{i=1} L_{ii} = v/n$ where $v = trace(L)$ is the effective degrees of freedom. Thus, we would minimize

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(\frac{Y_i - \hat{f}_n(x_i)}{1 - v/n})^2. \tag{2.13}$$

The GCV has computational advantages over ordinary CV, and it also has advantages in terms of invariance properties (see Wahba (1990, sec. 4.5.2 & 4.5.3). In practice, however, the two are usually similar. The algorithm for estimating the smoothing parameters by this GCV statistic are outlined inWood (2006, 2017a).

CHAPTER 3

Penalized Splines -Based Test Statistic

3.1 Spectral Estimates and our Hypothesis Testing

Spectral analysis is a frequency domain approach in time series analysis. The fundamental objective of spectral analysis is to identify the dominant frequencies in a series and to find an explanation of the system from which the measurements were derived, cycles of unexpected frequencies. The spectral density is a frequency domain representation of a time series that is directly related to the autocovariance time-domain representation. Some technical issues of spectral analysis go well beyond the scope of this thesis. For the hypothesis testing problem in (1.1), the spectral density of $\{X_t\}$ at frequency $\omega \in [0, 2\pi)$ is typically estimated by its periodogram, denoted by $I_X(\omega)$:

$$\hat{f}_X(\omega) = I_X(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t e^{-it\omega} \right|^2 . \tag{3.1}$$

If $\{X_t\}$ is a stationary linear process with *i.i.d.* innovations, then asymptotically $I_X(\omega_l) \sim f_X(\omega_l)\, \chi^2_{(2)}/2$ for $\omega_l = 2\pi l/n_1$ and $1 \leq l < n_1/2$. (For details, see Thm 10.3.2 in Brockwell & Davis (1991).) Also $I_X(\omega_k)$ and $I_X(\omega_l)$ are asymptotically independent for all $k \neq l$. Since we consider two time series $\{X_t, t = 1, 2, \ldots, n_1\}$ and $\{Z_t, t = 1, 2, \ldots, n_2\}$ with $n_1 \neq n_2$, the periodograms $I_X(\omega_l)$ and $I_Z(\omega_k)$ are available at different Fourier frequencies (i.e., $\omega_l = 2\pi l/n_1$ and $\omega_k = 2\pi k/n_2$). We adopt the suggestion from Coates & Diggle (1986) to estimate the spectral densities at the same Fourier frequency using local averages of consecutive periodogram values in non-overlapping blocks.

So, to overcome the unequality issues, $n_1 \neq n_2$, one can factorize $n_i = p_i\, N$ for some fixed integers p_1 and p_2, possibly afterdiscarding few observation, with $1 \leq p_1, p_2 \leq 10$. Now, construct $n := [(N-1)/2]$ ($[x]$ indicates the largest integer less than or equal to x) spectral estimates $I_X^{p_1}(\omega_i^*)$ of $f_X(\omega_i^*)$ and $I_Z^{p_2}(\omega_i^*)$ of $f_Z(\omega_i^*)$ by averaging non-overlapping blocks of p_1 and p_2 periodogram ordinates of series $\{X_t\}$ and $\{Z_t\}$, respectively. For details, see Decowski & Li (2015). Then the log-spectral ratio is defined and denoted as

$$\begin{aligned} Y_i^* &= \ln\Big[\frac{I_X^{p_1}(\omega_i^*)}{I_Z^{p_2}(\omega_i^*)}\Big], \text{ then we have} \\ Y_i^* &= \ln\Big[\frac{f_X(\omega_i^*)}{f_Z(\omega_i^*)}\Big] + \ln\Big[\frac{I_X^{p_1}(\omega_i^*)/f_X(\omega_i^*)}{I_Z^{p_2}(\omega_i^*)/f_Z(\omega_i^*)}\Big] \\ &= h(\omega_i^*) + \ln V_i, \end{aligned} \tag{3.2}$$

where, V_i are independent and

$$V_i \sim \frac{2p_2\chi^2_{(2p_1)}}{2p_1\chi^2_{(2p_2)}}, \text{ because asymptotically } \frac{I_X(\omega_l)}{f_X(\omega_l)} \sim \chi^2_{(2)}/2.$$

.

Therefore, V_i are asymptotically distributed independently as $F_{2p_1,2p_2}$. Let the random variable $V_0 \sim F_{2p_1,2p_2}$ and $v_0 = E\big[\ln V_0\big]$, we write the above equation further as

$$\begin{aligned} Y_i^* &= h(\omega_i^*) + \ln V_i = h(\omega_i^*) + E\big[\ln V_0\big] + \big(\ln V_i - E\big[\ln V_0\big]\big) \\ &= h(\omega_i^*) + v_0 + \varepsilon_i, \end{aligned} \tag{3.3}$$

then the observed log spectral ratio is,

$$Y_i = Y_i^* - v_0 = h(\omega_i^*) + \varepsilon_i = f(x_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n. \tag{3.4}$$

where $f(x_i) = f(i/n) = h(\omega_i^*) = \ln\big[f_X(\omega_i^*)/f_Z(\omega_i^*)\big]$ and $\varepsilon_i = \ln V_i - v_0$ are asymptotically independent with mean zero. Now the equation (3.4) is considered to be the same as the simple regression model (2.2).

Under the null hypothesis, the log-ratio of the spectral density, $h(\omega_i^*)$, is zero and from (3.4), $Y_i = \varepsilon_i$. Then the hypothesis in (1.1) becomes

$$\begin{aligned} H_0 &: \ f(x) = 0, \text{ or } \beta_j = 0 \qquad \text{for all } j = 1, 2, \cdots, q \\ H_1 &: \ f(x) \neq 0, \text{ or } \beta_j \neq 0 \qquad \text{for some } j = 1, 2, \cdots, q. \end{aligned} \tag{3.5}$$

3.2 Estimation of Periodogram and Discrete Fourier Transform

To perform spectral analysis, we first must transform data from the time domain to the frequency domain. The discrete Fourier transform (DFT) is the most important discrete transform, used to perform Fourier analysis in many practical applications. The DFT transforms a finite sequence of evenly spaced signals from its original domain (often time or space) into a same-length sequence of equally-spaced function of frequency. The DFT is therefore said to be a frequency domain representation of the original input sequence. Given data $X_1, \ldots, X_n$, the DFT is then

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} X_t \mathrm{e}^{-2\pi i \omega_j t} \tag{3.6}$$

for $j = 0, 1, \ldots, n-1$, where the frequencies $\omega_j = j/n$, or j cycles in n time points, are called the Fourier or fundamental frequencies. The periodogram $I(\omega_j)$ is defined to be the squared modulus of the DFT. That is

$$I(\omega_j) = |d(\omega_j)|^2$$

for $j = 0, 1, 2, \ldots, n-1$. Note that $I(0) = n\bar{X}^2$, where $\bar{X}$ is the sample mean. Also,

$\sum_{t=1}^{n} \exp\left(-2\pi i t \frac{j}{n}\right) = 0$ for $j \neq 0$, so we can write the DFT as

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} \left(X_t - \bar{X}\right) \mathrm{e}^{-2\pi i \omega_j t} \text{ for } j \neq 0.$$

Thus, for $j \neq 0$,

$$\begin{aligned} I(\omega_j) = |d(\omega_j)|^2 &= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \left(X_t - \bar{X}\right)\left(X_s - \bar{X}\right) \mathrm{e}^{-2\pi i \omega_j (t-s)} \\ &= n^{-1} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} \left(X_{t+|h|} - \bar{X}\right)\left(X_t - \bar{X}\right) \mathrm{e}^{-2\pi i \omega_j h} \\ &= \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-2\pi i \omega_j h} \end{aligned}$$

where h in the autocovariance function $\hat{\gamma}(h)$ is the lag length, i.e., $h = t - s$. In view of the above equation, we may think of the periodogram, $I(\omega_j)$ as the sample spectral density of X_t.

We write the expression of DFT and periodogram again as follows

$$\begin{aligned} d\left(\frac{j}{n}\right) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \mathrm{e}^{-2\pi i t \frac{j}{n}} \\ &= \frac{1}{\sqrt{n}} \left[\sum_{t=1}^{n} X_t \cos\left(2\pi t \frac{j}{n}\right) - i \sum_{t=1}^{n} X_t \sin\left(2\pi t \frac{j}{n}\right)\right], \ \because e^{-i\alpha} = \cos\alpha - i \sin\alpha \\ \left|d\left(\frac{j}{n}\right)\right|^2 &= \frac{1}{n}\left(\sum_{t=1}^{n} X_t \cos\left(2\pi t \frac{j}{n}\right)\right)^2 + \frac{1}{n}\left(\sum_{t=1}^{n} X_t \sin\left(2\pi t \frac{j}{n}\right)\right)^2 \\ &= \frac{n}{4}\left(a_j^2 + b_j^2\right) = \frac{n}{4} P\left(\frac{j}{n}\right), \end{aligned}$$

where, $P\left(\frac{j}{n}\right) = \left(a_j^2 + b_j^2\right)$ is called the scaled periodogram with coefficients a_j and b_j. The coefficients a_j and b_j could be obtained by regressing $X_t;\ t = 1, 2, \ldots, n$ on periodic sines

and cosines (for details. refer to Shumway et al. (2000)). For example, for an odd n,

$$a_0 = \bar{x} = \frac{1}{n}\sum_{t=1}^{n} X_t \text{ for } j = 0, \text{ and}$$
$$a_j = \frac{2}{n}\sum_{t=1}^{n} X_t \cos(2\pi\omega_j t)$$
$$b_j = \frac{2}{n}\sum_{t=1}^{n} X_t \sin(2\pi\omega_j t).$$

for $j = 1, 2, \ldots, n-1$. The scaled periodogram is the estimate of σ_j^2 corresponding to the sinusoid oscillating at a Fourier frequency of $\omega_j = j/n$. Large values of $P(j/n)$ indicate which frequencies (j/n) are dominant in the series, whereas small values of $P(j/n)$ may be associated with noise. Note that

$$P(j/n) = P(1 - j/n), \quad j = 0, 1, \ldots, n-1,$$

so there is a mirroring effect at the folding frequency of 1/2; consequently, the periodogram is typically not plotted for frequencies higher than the folding frequency.

If n is a highly composite integer (i.e., it has many factors), the DFT can be computed by an algorithm called the fast Fourier transform (FFT) introduced in Cooley & Tukey (1965), and there is no need to run repeated regressions of X_t on periodic sines and cosines. Different R packages scale the FFT differently. Using FFT, R computes the DFT defined in (3.6) as follows

$$fft(x) = \sum_{t=1}^{n} X_t \mathrm{e}^{-2\pi i t \frac{j}{n}} \cdot \mathrm{e}^{2\pi i \frac{j}{n}}$$
$$\Rightarrow \left| d\left(\frac{j}{n}\right) \right|^2 = \frac{1}{n}\left(\sum_{t=1}^{n} X_t e^{-2\pi i t \frac{j}{n}}\right)^2 = \frac{1}{n}|fft(x)|^2.$$

Thus, $P(j/n) = \frac{4}{n}\left|d\left(\frac{j}{n}\right)\right|^2 = \frac{4}{n^2}|fft(x)|^2$ in R. But, in this thesis for a match to the equation

(3.1), we compute the periodogram as

$$\left| d\left(\frac{j}{n}\right)\right|^2 = \frac{1}{2\pi n}\left|fft(x)\right|^2.$$

The normalizing constant does not have any impact on the final result because as we see in (3.2), the observed data arises as a ratio of two periodogram values.

3.3 Our New Test Statistic

In practice, a change of penalized spline basis does not change the fit though some bases are more numerically stable and allow computation of fit with greater accuracy. Besides numerical stability, reasons for selecting one basis over another are ease of implementation, especially of penalties, and interpretability. Now, we are ready to define penalized regression splines mathematically. For K inner knots and two boundary knots, usually at 0 and 1, let $\{N_j\}_{j=1}^{K}$ denote the collection of penalized spline basis functions (for example, natural cubic splines or B-spline bases for P-splines) and $N = (N(x_1),\ N(x_2),\ \cdots,\ N(x_n))^t$ be the $n \times K$ design matrix of the basis functions. Therefore, the penalized regression spline estimator with P-spline basis and second ordered derivative (or difference) penalty of the fitted function is the minimizer of

$$\sum_{i=1}^{n}\left\{Y_i - \sum_{j=1}^{K}\beta_j N_j(x_i)\right\}^2 + \lambda\int_x \{f''(x)\}^2 dx. \tag{3.7}$$

Because the mean regression function f is linear in the parameters, β_j for $j = 1, 2, \cdots, q$, the roughness penalty in (3.7) can be written as

$$\int_x \{f''(x)\}^2 dx = \int\left[\left\{\sum_{j=1}^{K}\beta_j N_j(x)\right\}''\right]^2 dx = \beta^t D_q \beta,$$

where $D_q = \int[N''(x)]^t[N''(x)dx]$, a second order derivative-based $q \times q$ penalty matrix. For

simplicity here we would apply the natural cubic spline basis function $R(x, \xi)$ described in subsection (2.2.4) with a second order derivative based penalty matrix. So, the element of the penalty matrix D are $D_{j+2,k+2} = R(\xi_j, \xi_k),\ for\ j, k = 1, \cdots, q-2$ while the first two rows and columns of D_q are 0 (Gu (2002, sec. 2.3.3)). The formation of least square estimation of the penalized splines regression and it's computational procedure using the R language is described in Wood (2006, 2017*a*). Therefore, in matrix notation, the penalized regression spline fitting problem is to minimize

$$S(\beta) = (Y - N\beta)^\top (Y - N\beta) + \lambda \beta^\top D_q \beta$$

w.r.t β. With this notation the the least squares version of the penalized spline estimator of β and $f_n(x)$ are then

$$\hat{\beta} = (N^\top N + \lambda D_q)^{-1} N^\top Y, \tag{3.8}$$

and

$$\hat{f}_n(x) = N(N^\top N + \lambda D_q)^{-1} N^\top Y = LY = \sum_{i=1}^{n} Y_i l_i(x) \tag{3.9}$$

respectively. Here, $L = N(N^\top N + \lambda D_q)^{-1} N^\top$ is called the influence (hat) matrix, and l_i is the ith diagonal element of this hat matrix. To test the null hypothesis in (3.5), similar to Chen et al. (2014), we proposed a simple test statistic based on the sum of squared distances of the fitted values as,

$$T_n = \hat{f}_n^\top \hat{f}_n = Y^\top N(N^\top N + \lambda D_q)^{-1} N^\top N(N^\top N + \lambda D_q)^{-1} N^\top Y \tag{3.10}$$

A similar test based on the smoothing spline estimator was proposed in Eubank & Spiegelman (1990), Chen (1994), and Jayasuriya (1996). The average bias and variance can be expressed in terms of the eigen values obtained from the singular value decomposition used in Claeskens

et al. (2009).

$$(N^\top N)^{-1/2} D_q (N^\top N)^{-1/2} = FSF^\top, \tag{3.11}$$

Where, F is the matrix of eigen vectors ad $S = diag(s_1, \cdots, s_q)$ is the diagonal matrix of eigen values. A simpler form of the test statistic from (3.10) similar to Huaihou Chen et al. (2014) can be derived by the following simple matrix algebra. Since the left hand side of equation (3.11) is symmetric positive definite, because $N^\top N$ is symmetric positive definite and D_q is symmetric positive semidefinite, F is an orthogonal matrix such that, $F^\top F = I_q = FF^\top$ or $F^{-1} = F^\top$.

Define a semi-orthogonal matrix $A = N(N^\top N)^{-1/2} F$, such that

$$A^\top A = F^\top (N^\top N)^{-1/2} N^\top N (N^\top N)^{-1/2} F = F^\top F = I_q, \qquad \text{and}$$
$$AA^\top = N(N^\top N)^{-1/2} FF^\top (N^\top N)^{-1/2} N^\top = N(N^\top N)^{-1} N^\top.$$

A simpler from of the test statistic can be formed using eigen values or singular values obtained in (3.11) as follows:

$$
\begin{aligned}
\hat{f}_n(x) &= N\big(N^\top N + \lambda D_q\big)^{-1} N^\top Y \\
&= N(N^\top N)^{-1/2}\big[(N^\top N)(N^\top N)^{-1} + \lambda D_q (N^\top N)^{-1}\big]^{-1}(N^\top N)^{-1/2} N^\top Y \\
&= N(N^\top N)^{-1/2}\big[I_q + \lambda (N^\top N)^{-1/2} D_q (N^\top N)^{-1/2}\big]^{-1}(N^\top N)^{-1/2} N^\top Y \\
&= N(N^\top N)^{-1/2} F\big[I_q + \lambda F diag(S) F^\top\big]^{-1} F^\top (N^\top N)^{-1/2} N^\top Y \\
&= A\big[I_q + \lambda diag(S) F F^\top\big]^{-1} A^\top Y \\
&= A\big(I_q + \lambda S\big)^{-1} A^\top Y \ ; \text{here}, S = S = diag(s_1, \cdots, s_q). \qquad (3.12)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
T_n &= \Big[A\big(I_q + \lambda S\big)^{-1} A^\top Y\Big]^\top \Big[A\big(I_q + \lambda S\big)^{-1} A^\top Y\Big] \\
&= Y^\top A (I_q + \lambda S)^{-2} A^\top Y = Y^\top H_n^2 Y, \qquad (3.13)
\end{aligned}
$$

where,

$$
H_n = A(I_q + \lambda S)^{-1} A^\top.
$$

3.4 Asymptotic Distribution of the Test Statistic

Under the null hypothesis the two spectral densities are the same. So, under the assumption $\varepsilon_i \overset{i.i.d.}{\sim} (0, \sigma^2)$, we have $Y_i = \varepsilon_i$, for $i = 1, \cdots, n$ and $Y_1, \cdots, Y_n \overset{i.i.d.}{\sim} (0, \sigma^2)$ with $E(Y_i) = 0$, and $Var(Y_i) = \sigma^2$. Let, the asymptotic form of the test statistic T_n in (3.13) be denoted as T_0. Then under the null hypothesis, one can adopt the following theorem.

Theorem: Suppose the ε_i are *i.i.d.* with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$, and $K^2 = o(n)$, where K is the number of knots. Then the null distribution of T_n is

$$
T_n \approx T_0 = \sigma^2 \sum_{j=1}^{q} \frac{w_i^2}{(1 + \lambda s_i)^2} \quad \text{as } n \to \infty, \qquad (3.14)
$$

where $w_i \overset{i.i.d.}{\sim} N(0, 1)$ and the error variance σ^2 is estimated as σ_0^2 as in (3.17).

Before presenting the formal proof of the theorem, we need some preparations. We define $G_{K,n} = (N^t N)/n$, and state the following results and lemmas.

Result R1 (*Lemma 6.2 in Zhou et al. (1998)*). There exist constants $0 < c_1 \leq c_2 < \infty$, independent of K and n, such that the lower and upper bounds for the eigenvalues of $G_{K,n}$ become

$$(c_1 + o(1))\delta \leq \tau_{min} \leq \tau_{max} \leq (c_2 + o(1))\delta.$$

Here, τ_{min} and τ_{max} are the minimum and maximum eigen values of $G_{K,n}$, respectively. Since, for large n, $o(1) \to 0$, then for some constants $0 < c_3 \leq c_4 < \infty$,

$$c_3\delta \leq \tau_{min} \leq \tau_{max} \leq c_4\delta.$$

Result R2 (*Lemma 6.3 in Zhou et al. (1998)*). There exist constants $c_5 \in (0, \infty)$ *and* $\gamma \in (0, 1)$ such that for large n,

$$|\{G_{K,n}^{-1}\}_{j,k}| \leq c_5\delta^{-1}\gamma^{|j-k|}.$$

Then,

$$\|G_{K,n}^{-1}\|_\infty \equiv \max_{1\leq j\leq q} \sum_{k=1}^{q} |\{G_{K,n}^{-1}\}_{j,k}| = O(\delta^{-1}).$$

For the proof of results R1 and R2, please refer to Zhou et al. (1998), which are also studied in Claeskens et al. (2009).

Lemma L1 (*Cramér-Wold Device*). The Cramér-Wold device is a useful statistical tool to extend the one dimensional central limit theorem (CLT) to a multi-dimensional setting. Let X_n be a sequence of P -dimensional random variables. It states that X_n converges in distribution to a random variable X if for all $\tilde{a} \in \mathbb{R}^P$, $\tilde{a}^\top X_n$ converges in distribution to $\tilde{a}^\top X$.

Proof: It is given that for any vector $\tilde{a} \in \mathbb{R}^P$,

$$\begin{aligned} &\tilde{a}^\top X_n \xrightarrow{d} \tilde{a}^\top X \\ \Rightarrow &\Phi_{\tilde{a}^\top X_n}(t) \xrightarrow{d} \Phi_{\tilde{a}^\top X}(t) \text{ as } n \to \infty \\ \Rightarrow &\Phi_{\tilde{a}^\top X_n}(1) \xrightarrow{d} \Phi_{\tilde{a}^\top X}(1) \text{ as } n \to \infty \end{aligned} \tag{3.15}$$

Now,

$$\Phi_{\tilde{a}^\top X_n}(1) = E\left[e^{i\tilde{a}^\top X_n}\right] = \Phi_{X_n}(\tilde{a}), \text{ and}$$

$$\Phi_{\tilde{a}^\top X}(1) = E\left[e^{i\tilde{a}^\top X}\right] = \Phi_X(\tilde{a}).$$

Using equation (3.15),

$$\Phi_{X_n}(\tilde{a}) \xrightarrow{d} \Phi_X(\tilde{a}) \; as \; n \to \infty$$

$$\Rightarrow X_n \xrightarrow{d} X \; as \; n \to \infty.$$

Lemma L2 (*A Special case of Lindeberg-Feller Theorem*). Suppose $X_1,\ X_2, \cdots, X_n$ are $i.i.d$ random variables with mean μ and variance σ^2. Let $T_n = \sum_{i=1}^n z_{ni} X_i$ with $\mu_n = E(T_n)$ and $\sigma_n^2 = var(T_n)$, where the z_{ni} are given numbers. Then, $\frac{T_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1)$ provided $\max_{i \le n} \frac{z_{ni}^2}{\sum_{i=1}^n z_{ni}^2} \to 0 \; as \; n \to \infty$. For details about the Lindeberg-Feller Theorem and proof of this result, please see Ferguson (1996).

Lemma L3. The minimum eigenvalue of $G_{K,n}^{1/2}$ has the lower bound $c_0 \delta^{1/2}$, and hence

$$\|G_{K,n}^{-1/2}\|_\infty = O(\delta^{-1/2}).$$

Proof: We considered $G_{K,n} = (N^t N)/n$. Suppose P is any orthogonal matrix such that

$$PP^\top = I \text{ and } \left(P^\top\right)^{-1} = P.$$

Then

$$P(G)P^{\top} = \begin{pmatrix} \tau_1 & 0 & \cdots & 0 \\ 0 & \tau_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tau_q \end{pmatrix},$$

where $\tau_i \geqslant \tau_{(1)} \geqslant c_3\delta$ (from Result 1) and τ_i 's are eigen values of $G_{K,n}$. Now,

$$\begin{aligned} G_{K,n} &= P \begin{pmatrix} \tau_1 & 0 & \cdots & 0 \\ 0 & \tau_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tau_q \end{pmatrix} P^{\top} \\ &= P \begin{pmatrix} \sqrt{\tau_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\tau_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\tau_q} \end{pmatrix} P^{\top} P \begin{pmatrix} \sqrt{\tau_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\tau_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\tau_q} \end{pmatrix} P^{\top} \\ &= \left(G_{K,n}^{\frac{1}{2}}\right)^{\top} \left(G_{K,n}^{\frac{1}{2}}\right) \end{aligned}$$

This implies

$$G_{K,n}^{\frac{1}{2}} = P \begin{pmatrix} \sqrt{\tau_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\tau_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\tau_q} \end{pmatrix} P^{\top}$$

Therefore, using result R1 (Lemma 6.2 of Zhou et al. (1998)) it can be shown that the minimum eigenvalue of $G_{K,n}^{\frac{1}{2}} = \sqrt{\tau_{min}} \geq \sqrt{(c_3\delta)} = c_0\delta^{1/2}$

Again, from the property of eigenvalues and matrix inversion one can write

$$G_{K,n}^{-\frac{1}{2}} = P \begin{pmatrix} 1/\sqrt{\tau_1} & 0 & \cdots & 0 \\ 0 & 1/\sqrt{\tau_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sqrt{\tau_q} \end{pmatrix} P^\top$$

For the similar argument as in result R2, and for large n, one could expect that

$$|\{G_{K,n}^{-\frac{1}{2}}\}_{j,k}| \leq c_0 \delta^{-1/2} \gamma^{|j-k|}, \text{ and hence}$$

$$\|G_{K,n}^{-\frac{1}{2}}\|_\infty \equiv \max_{1 \leq j \leq q} \sum_{k=1}^{q} |\{G_{K,n}^{-\frac{1}{2}}\}_{j,k}| = O(\delta^{-1/2}) = O(K^{1/2}),$$

since for equally spaced knots sequence, $\delta \sim K^{-1}$ and δ is the difference between any two successive knots.

Proof of the main theorem: The test statistic can be written as

$$\begin{aligned} T_n &= \tilde{Y} H_n^2 \tilde{Y} = \tilde{Y} A \left(I_q + \lambda S\right)^{-2} A^\top \tilde{Y} \\ &= \left(A^\top \tilde{Y}\right)^\top \left(I_q + \lambda S\right)^{-2} A^\top Y \\ &= \tilde{V}^\top \left(I_q + \lambda S\right)^{-2} \tilde{V}, \text{ where } A^T \tilde{Y} = \tilde{V}_{q \times 1} \end{aligned}$$

Here, $\widetilde{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1}$ and $Y_{1,}, Y_2, \ldots, Y_n \overset{iid}{\sim} (0, \sigma^2)$ with $E(Y_1) = 0$ and $\text{var}(Y_1) = \sigma^2$.

Now, in order to prove the theorem, we want to show that $\widetilde{V}_{q \times 1} \xrightarrow{d} N\left(\tilde{0}, \sigma^2 \, \mathrm{I}_{q \times q}\right)$, or equivalently $\widetilde{W}_{q \times 1} = \sigma^{-1} \cdot \widetilde{V}_{q \times 1} \xrightarrow{d} N\left(\tilde{0}, \mathrm{I}_{q \times q}\right)$, where I stands for an identity matrix. Because, if $\widetilde{V}_{q \times 1} \xrightarrow{d} N\left(\tilde{0}, \sigma^2 \, \mathrm{I}_{q \times q}\right)$ then one can rewrite the test statistic T_n as

$$T_n = \tilde{V}^\top (I_q + \lambda S)^{-2} \tilde{V} \;\approx\; \sigma^2 \, \widetilde{W}^\top (I_q + \lambda S)^{-2} \widetilde{W}$$

$$= \sigma^2 \begin{pmatrix} W_1 & W_2 & \cdots & W_q \end{pmatrix} \begin{pmatrix} 1/(1+\lambda s_1)^2 & 0 & \cdots & 0 \\ 0 & 1/(1+\lambda s_2)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/(1+\lambda s_q)^2 \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_q \end{pmatrix}$$

$$= \sigma^2 \sum_{j=1}^{q} \frac{w_i^2}{(1+\lambda s_i)^2}.$$

From Cramer-Wold Device or from Lemma 1, to show $\tilde{V}_{q\times 1} \xrightarrow{d} N\left(\tilde{0}, \sigma^2 I_{q\times q}\right)$ is equivalent to show

$$\tilde{a}^\top \tilde{V}_{qx1} \xrightarrow{d} N\left(\tilde{0}, \sigma^2 \tilde{a}^\top \tilde{a}\right) \text{ for all } \tilde{a} \in \mathbb{R}^q$$

$$i.e., \quad \tilde{a}^\top A^\top \tilde{Y} \xrightarrow{d} N\left(\tilde{0}, \sigma^2 \|\tilde{a}\|^2\right).$$

Now, $\tilde{a}^\top A^\top \tilde{Y} = (A\tilde{a})^\top \tilde{Y}$, denote $\tilde{C}_{n\times 1} = A\tilde{a} = A_{n\times q}\tilde{a}_{q\times 1}$

$$\Rightarrow (A\tilde{a})^\top \tilde{Y} = \tilde{C}^T \tilde{Y} = \sum_{i=1}^{n} C_i Y_i$$

From the result of Lindeberg- Feller theorm (Lemma2),

$$\sum_{i=1}^{n} C_i Y_i \xrightarrow{d} N(\tilde{0}, \Sigma) \text{ if } \frac{\max_{1\le i\le n} C_i^2}{\sum_{i=1}^{n} C_i^2} \to 0 \; as \; n \to \infty,$$

$$\text{where,} \quad \Sigma = \text{Var}\left(\sum_{i=1}^{n} C_i Y_i\right) = \tilde{C}^T C \cdot \sigma^2$$

$$= \tilde{a}^\top A^\top A \tilde{a} \cdot \sigma^2 = \tilde{a}^\top I_{q\times q} \tilde{a} \cdot \sigma^2 = \|\tilde{a}\|^2 \cdot \sigma^2$$

Therefore, one needs only to verify

$$\frac{\max_{i\leq i\leq n} C_i^2}{\sum_{i=1}^n C_i^2} \to 0 \text{ as } n \to \infty \tag{3.16}$$

Now, $\tilde{C} = A\tilde{a} = N\left(N^\top N\right)^{-\frac{1}{2}} F\tilde{a} = N\left(\frac{N^\top N}{n}\right)^{-\frac{1}{2}} n^{-\frac{1}{2}}\tilde{b}$, where,

$$\tilde{b}_{q\times 1} = F_{q\times q}\tilde{a}_{q\times 1}, \text{ and } \tilde{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_q \end{pmatrix}$$

Then

$$\|\tilde{b}\|^2 = \tilde{b}^T\tilde{b} = \tilde{a}F^\top F\tilde{a} = \tilde{a}^\top I_{q\times q}\tilde{a} = \|\tilde{a}\|^2, \text{ and}$$
$$\tilde{C} = n^{-\frac{1}{2}} N\left(\frac{N^\top N}{n}\right)^{-\frac{1}{2}}\tilde{b} = n^{-\frac{1}{2}} N\left(G_{K,n}^{-\frac{1}{2}}\right)\tilde{b}.$$

Let, $\tilde{d}_{q\times 1} = \left(G_{K,n}^{-\frac{1}{2}}\tilde{b}\right)_{q\times 1} = (d_1, d_2, \cdots, d_q)^\top$. From Lemma 3, $\left\|G_{K,n}^{-\frac{1}{2}}\right\|_\infty = O\left(K^{\frac{1}{2}}\right)$. So, one has $d_j = \sum_{k=1}^q \left\{G_{K,n}^{-\frac{1}{2}}\right\}_{j,k} \cdot b_k$. This means, $\tilde{C} = n^{-\frac{1}{2}} N_{n\times q} \cdot \tilde{d}_{q\times 1}$, which implies,

$$\begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix} = n^{-\frac{1}{2}} \begin{pmatrix} N_1(x_1,\xi_1) & N_2(x_1,\xi_2) & N_3(x_1,\xi_3) & \cdots & N_q(x_1,\xi_q) \\ N_1(x_2,\xi_1) & N_2(x_2,\xi_2) & N_3(x_1,\xi_3) & \cdots & N_q(x_2,\xi_q) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ N_1(x_n,\xi_1) & N_2(x_n,\xi_2) & N_3(x_n,\xi_3) & \cdots & N_q(x_n,\xi_q) \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_q \end{pmatrix}$$

$$\Rightarrow C_i = n^{-\frac{1}{2}} \sum_{j=1}^q N_j(x_i,\xi_j)\cdot d_j$$

$$\Rightarrow |C_i| \leq n^{-\frac{1}{2}} \sum_{j=1}^q N_j(x_i,\xi_j)\cdot |d_j|, \text{ where}$$

$N_1(x_i,\xi_1) = 1, N_2(x_i,\xi_2) = x_i$ and $N_j(x_i,\xi_j) = R(x_i,\xi_j)$, for $j = 3,4,\cdots,q$.

But,

$$
\begin{aligned}
|d_j| &\leq \left\| G_{K,n}^{-\frac{1}{2}} \right\|_{\infty} \sum_{k=1}^{q} |b_k| = O\left(K^{\frac{1}{2}}\right) \sum_{k=1}^{q} |b_k \cdot 1| \\
&\leq O\left(K^{\frac{1}{2}}\right) \sqrt{\left(\sum_{k=1}^{q} b_k^2\right)\left(\sum_{k=1}^{q} 1\right)} \text{ (using Cauchy Schwarz Inequality)} \\
&= O\left(K^{\frac{1}{2}}\right) \sqrt{\|\tilde{b}\|^2 \cdot q} = O\left(K^{\frac{1}{2}}\right) \sqrt{\|\tilde{a}\|^2 \cdot q} \\
&\approx O\left(K^{\frac{1}{2}}\right) \|\tilde{a}\| K^{\frac{1}{2}} = O(K)\|\tilde{a}\|
\end{aligned}
$$

This indicates,

$$
\begin{aligned}
|C_i| &\leq n^{-\frac{1}{2}} \textstyle\sum_{j=1}^{q} N_j\left(x_i, \xi_j\right) \cdot O\left(K\right) \|\tilde{a}\| \\
&= n^{-\frac{1}{2}} \cdot O\left(K\right) \|\tilde{a}\| = O\left(\tfrac{K}{n^{1/2}}\right) \|\tilde{a}\|
\end{aligned}
\qquad since \begin{cases} N_j\left(x_i, \xi_j\right) \geq 0 \text{ for all } i, j \text{ and} \\ \sum_{j=1}^{q} N_j\left(x_i, \xi_j\right) = 1 \text{ for all } k. \end{cases}
$$

Now, the numerator of (3.16) becomes $\max\limits_{1 \leq i \leq n} C_i^2 \leq O\left(\frac{K^2}{n}\right) \|\tilde{a}\|^2$. Again, the denominator of (3.16) has the following form

$$
\begin{aligned}
\sum_{i=1}^{n} C_i^2 = \|C\|^2 &= \tilde{C}\tilde{C}^{\top} = \tilde{a}^{\top} A^{\top} \cdot A\tilde{a} \\
&= \tilde{a}^{\top} I_{q \times q} \tilde{a} = \tilde{a}^{\top} \tilde{a} = \|\tilde{a}\|^2.
\end{aligned}
$$

Therefore, from equation (3.16) one gets

$$
\frac{\max\limits_{1 \leq i \leq n} C_i^2}{\sum_{i=1}^{n} C_i^2} \leq \frac{O\left(\frac{K^2}{n}\right) \|\tilde{a}\|^2}{\|\tilde{a}\|^2} = O\left(\frac{K^2}{n}\right) \to 0 \text{ as } n \to \infty.
$$

This completes the proof of the main theorem.

The computation of T_0 in (3.14) is very fast because it does not require the observed log-spectral ratios since the value of the smoothing parameter is fixed (determined before the actual simulations, refer to section 4.1).

3.5 Parameters of the Data Model

Since the random variable $V_0 \sim F_{2p_1,2p_2}$, theoretically one could obtain both $v_0 = E[\ln V_0]$ and $\sigma_0^2 = \mathrm{Var}(\ln V_0)$ as a function of p_1 and p_2 as

$$v_0 = \ln(p_2/p_1) + \psi^{(0)}(p_1) - \psi^{(0)}(p_2), \quad \sigma_0^2 = \psi^{(1)}(p_1) + \psi^{(1)}(p_2), \tag{3.17}$$

where in above expressions $\psi^{(l)}(p)$ is a polygamma function, which is defined as $\psi^{(l)}(p) = \frac{d^{(l+1)}}{dp^{(l+1)}} \ln\Gamma(p)$. Thus, one has $\psi^{(0)}(p) = \frac{d}{dp} \ln\Gamma(p) = \Gamma^{(1)}(p)/\Gamma(p)$ and $\psi^{(1)}(p) = \frac{d^2}{dp^2} \ln\Gamma(p) = \Gamma^{(2)}(p)/\Gamma(p) - (\psi^{(0)}(p))^2$. The mathematical derivation of these two data parameters are given in the appendix A. In the statistical software R, there is a built-in function $psigamma(p, l)$ to evaluate the above function $\psi^{(l)}(p)$. The value of σ_0^2 is a decreasing function of p_1 and (or) p_2. Table 3.1 show the estimated values of σ_0^2 for different combination of p_1 and p_2.

Table 3.1: Values of $\sigma_0^2 = \sigma_{2p_1,2p_2}^2$

p_1	p_2									
	1	2	3	4	5	6	7	8	9	10
1	3.2899	2.2899	2.0399	1.9288	1.8663	1.8263	1.7985	1.7781	1.7624	1.7501
2	2.2899	1.2899	1.0399	0.9288	0.8663	0.8263	0.7985	0.7781	0.7624	0.7501
3	2.0399	1.0399	0.7899	0.6788	0.6163	0.5763	0.5485	0.5281	0.5124	0.5001
4	1.9288	0.9288	0.6788	0.5676	0.5051	0.4651	0.4374	0.4170	0.4013	0.3890
5	1.8663	0.8663	0.6163	0.5051	0.4426	0.4026	0.3749	0.3545	0.3388	0.3265
6	1.8263	0.8263	0.5763	0.4651	0.4026	0.3626	0.3349	0.3145	0.2988	0.2865
7	1.7985	0.7985	0.5485	0.4374	0.3749	0.3349	0.3071	0.2867	0.2711	0.2587
8	1.7781	0.7781	0.5281	0.4170	0.3545	0.3145	0.2867	0.2663	0.2506	0.2383
9	1.7624	0.7624	0.5124	0.4013	0.3388	0.2988	0.2711	0.2506	0.2350	0.2227
10	1.7501	0.7501	0.5001	0.3890	0.3265	0.2865	0.2587	0.2383	0.2227	0.2103

CHAPTER 4

Simulation Studies

4.1 Deciding the Value of Smoothing Parameter

The smoothing parameter could be obtained automatically through the model fitting process in R. But, the model fitting is not the main focus of our study. Rather, our goal is to check the consistency of our proposed test through the calculation of the proportion of rejections under null and alternative hypotheses which require the λ to be known to expedite the simulation work. Since both time series have the same spectral density under the null hypothesis, the empirical size is insensitive to the choice of λ. But, in terms of power, there is a strong effect of the λ values, specifically for seasonal vs nonseasonal time series models.

Therefore, to reach out a general agreement on λ values for power calculations, we observe here the GCV score discussed in (2.13) and the observed log-spectral ratios for two specific examples. In either scenario, two independent and stationary time series, $\{X_t\}$ and $\{Z_t\}$ of lengths $n_1 = 500$ and $n_2 = 1000$, respectively, are considered. We started factoring the lengths of the two series by appropriately selecting $N = 200$ or $N = 400$ (the results are only shown for $N = 400$ as the λ values are insensitive to N). From this, the values of p_1 and p_2 were calculated as: $p_1 = [n_1/N] = [500/N]$ and $p_2 = [n_2/N] = [1000/N]$. In some cases, observations were discarded from the end of a series such that $p_1 N = p_2 N$. For example, two time series of lengths 402 and 201 with N selected to be 100 would result in $p_1 = 4,\ p_2 = 2$ and have 3 total observations discarded. Afterwards, the periodograms, $I_X(\omega)$ and $I_Z(\omega)$, of the two series were computed and $n = [(N-1)/2]$ averages were made in non-overlapping blocks. The details of this procedure are discussed in Section 3.1. The averaged periodograms

of the two series, denoted $I_X^{p_1}(\omega)$ and $I_Z^{p_2}(\omega)$, were now of equal length n. The log-spectral ratio of these series, $Y_i = \ln(I_X^{p_1}(\omega_i)/I_Z^{p_2}(\omega_i))$, was then used to estimate our new test statistic.

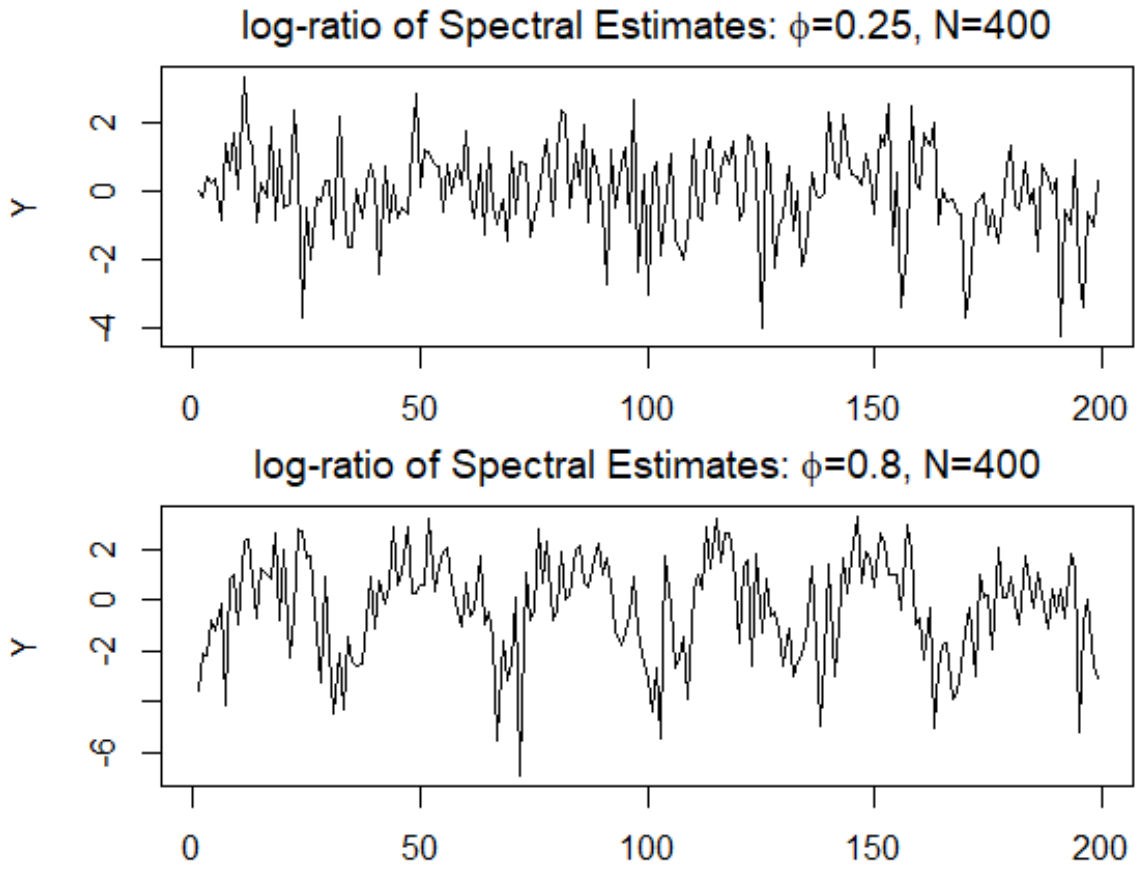

Figure 4.1: Time series plot for observed data: $X_t \sim N(0,1)$ and $Z_t \sim ARMA(1,0)_{12}$ with seasonal autoregressive parameter $\phi = 0.25$ (*top plot*) and $\phi = 0.80$ (*bottom plot*).

4.1.1 Selecting smoothing parameter: seasonality is present in the data.

The first example is illustrated by Figure 4.1 and Figure 4.2 in which the time series $\{X_t\}$ and $\{Z_t\}$ are generated from a standard normal model and a seasonal autoregressive model with frequency 12, respectively. Figure 1 exhibits the pattern of observed log-ratio of spectral estimates whereas Figure 2 shows what could be our optimal smoothing parameter λ by the GCV score. The seasonality becomes more visible when the magnitude of the seasonal

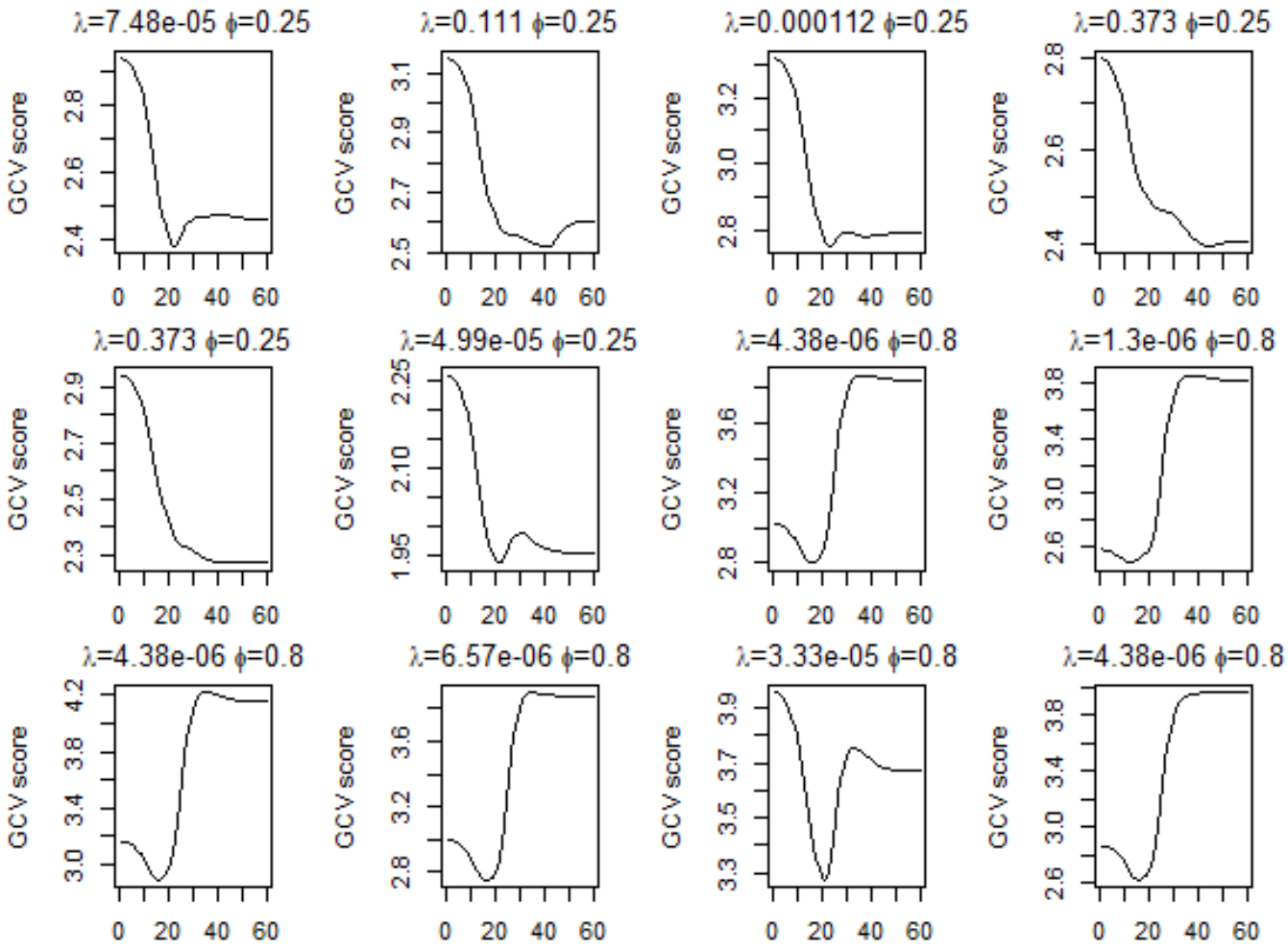

Figure 4.2: Generalized Cross Validation for optimal λ from the same time series used in Figure 4.1

autoregressive parameter ϕ is larger, consequently, the value of λ becomes smaller. We also ran the simulation, as the second example, for a white noise model against a seasonal moving average model of frequency 12. Here, we found the same conclusion that the value of the optimal smoothing parameter varies in the opposite direction of the value of the seasonal parameter, θ, and thus a high value of θ means more presence of seasonality. The plots of observed data and optimal lambda of this kind are presented in Figure B.1, and Figure B.2 in Appendix B. Thus, the penalized spline-based test involving a series with seasonality requires the parameter λ to be relatively small. So, when seasonality is present in the data the optimal value of λ ranges from 10^{-5} to 10^{-7}.

4.1.2 Selecting smoothing parameter: absence of seasonality in the data

We describe the third example by Figure 4.3 and Figure 4.4 in which the time series $\{X_t\}$ and $\{Z_t\}$ are generated from an autoregressive model and a moving average model, respectively. Figure 3 represents the pattern of observed data for example 3 as in Figure 1, and Figure 4 illustrates the optimal value of λ for example 2 as in Figure 2. In this case, seasonality is not present in the data, consequently, there is not a huge variation in the optimal λ values, and they are not very small compared to the case of seasonality. So, in the absence of seasonality in the data model, we fixed our smoothing parameters λ as 0.1, 0.01, and 0.001. We also ran the simulation for observed data and the GCV score when $\{X_t\}$ and $\{Z_t\}$ are $N(0,1)$ and $AR(1)$ process, respectively. The results are shown in Figure B.3 and Figure B.4, where we reached the same conclusion similar to example 2 as there is no seasonality in the data.

4.2 Empirical Type I Error Rate

This section explores the convergence of our test statistic with empirical size simulations for different combinations of time series lengths, models, N's, and λ's. Here, we again considered two independent and stationary time series of length n_1 and n_2. These simulations result of type I error rate are summarized in the following two tables. In particular, Table 4.1

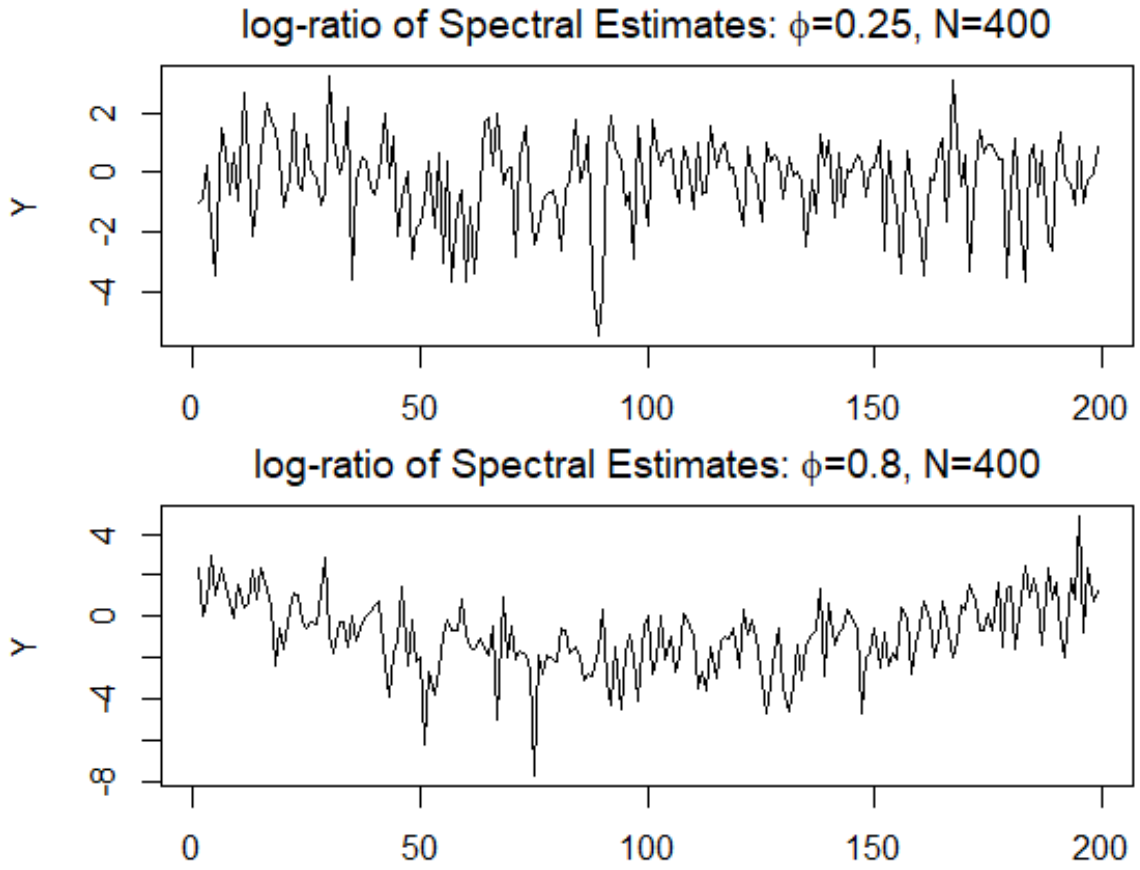

Figure 4.3: Time series plot for observed data: $X_t \sim AR(1)$ with AR parameter $\phi = 0.25$ (*top plot*) and $\phi = 0.80$ (*bottom plot*) and $Z_t \sim MA(1)$ with MA parameter $\theta = \frac{\phi}{|\phi|}\sqrt{\frac{\phi^2}{1-\phi^2}}$.

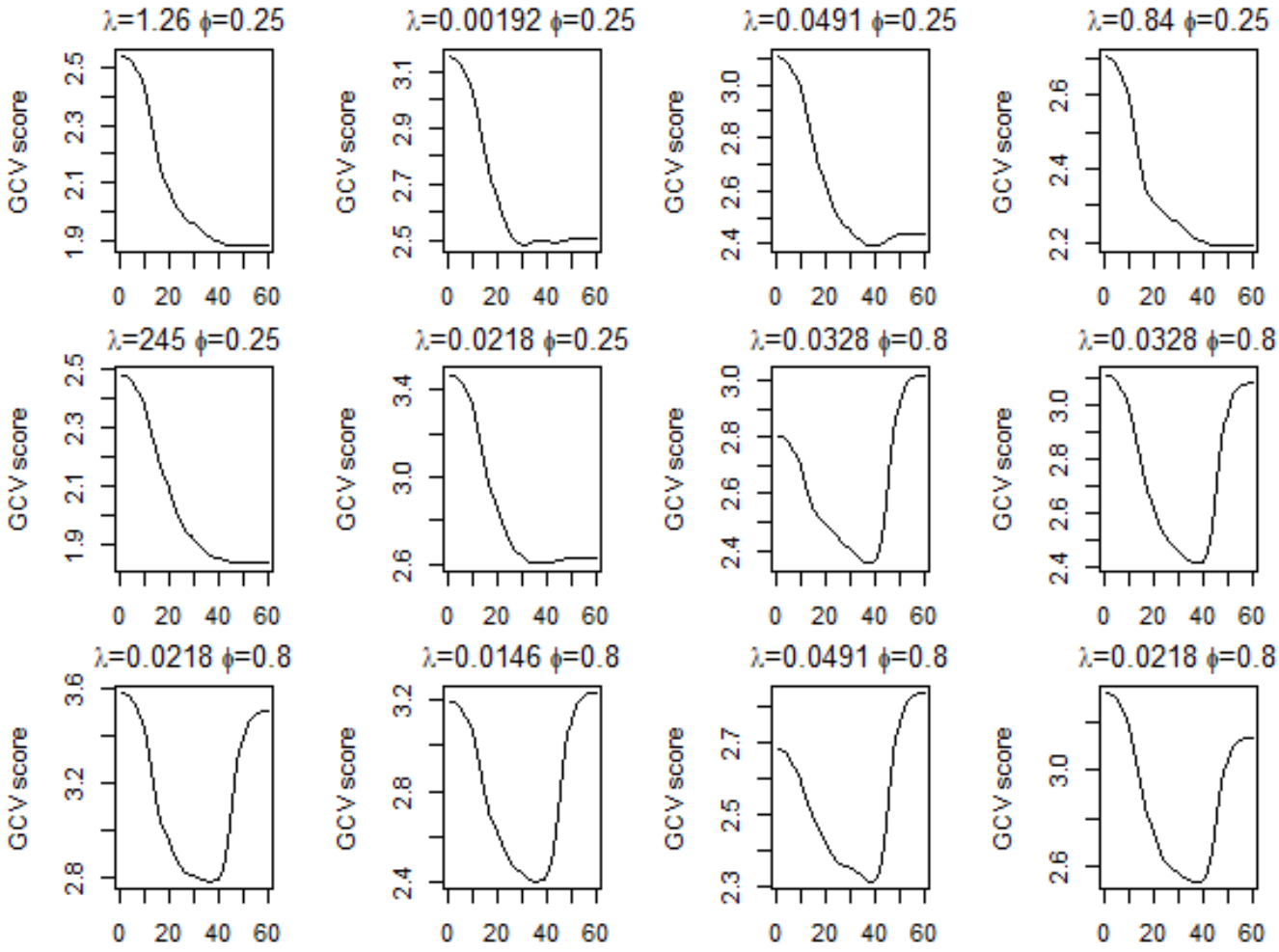

Figure 4.4: Generalized Cross Validation for optimal λ from the same time series used in Figure 4.3.

summarizes the simulation results where both the series are generated from *i.i.d.* standard normal random variates and for results of Table 4.2, the data of two series are generated from ARMA(1,1) model with AR and MA parameters $\phi = 0.3$ and $\theta = 0.4$, respectively. Using the procedure outlined in section 3.3 and with the selected values for λ in previous section, we computed the value of the test statistics T_n in (3.13) and T_0 in (3.14) from 10000 simulations. The critical value is the 95^{th} percentile value of T_0 and is the same for different time series under the same parameter combinations. The proportion of the test statistic, T_n, exceeding the critical value under H_0 is the empirical type I error.

Table 4.1: Empirical type I error rates in percentage for *i.i.d.* normal errors

$p_1, p_2; N$	Penalized Regression Splines-based Test					$p_1, p_2; N$	Wavelet -based Test				
	$\lambda = 1$	$\lambda = .1$	$\lambda = .01$	$\lambda = .001$	$\lambda = .0001$		J=1	J=2	J=3	J=4	J=5
			$\mathbf{n_1 = n_2 = 300}$						$\mathbf{n_1 = n_2 = 300}$		
1,1; 300	4.70	5.18	4.98	4.96	6.04	1,1; 300	4.94	4.66	5.55	6.04	7.25
2,2; 150	5.08	4.94	4.86	4.92	4.94	2,2; 150	5.25	5.35	5.94	6.65	7.94
3,3; 100	4.78	5.48	4.72	5.16	4.52	3,3; 100	5.08	4.83	5.35	6.48	
4,4; 75	4.94	4.86	4.98	4.62	4.44	4,4; 75	4.88	5.28	5.82	5.97	
5,5; 60	4.66	5.32	4.06	4.72	4.60	5,5; 60	5.12	4.85	5.12		
			$\mathbf{n_1 = n_2 = 1000}$						$\mathbf{n_1 = n_2 = 1000}$		
1,1; 1000	5.62	4.58	5.38	4.70	5.98	1,1; 1000	5.06	5.13	5.45	5.43	5.68
2,2; 500	5.10	4.84	4.70	5.16	5.92	2,2; 500	4.96	5.16	5.01	5.32	5.20
3,3; 333	5.24	5.16	5.20	4.80	5.10	3,3; 333	4.90	4.46	4.74	5.41	6.36
4,4; 250	4.96	5.12	4.82	4.34	4.66	4,4; 250	5.05	5.47	5.44	6.23	7.19
5,5; 200	5.14	5.06	4.70	4.98	4.28	5,5; 200	4.97	4.63	4.75	5.28	5.89
6,6; 166	4.00	4.36	5.32	5.00	4.82	6,6; 166	5.09	5.34	5.32	6.46	6.98
8,8; 125	5.56	5.02	4.68	4.68	4.96	8,8; 125	4.85	5.26	5.47	6.07	
10,10; 100	4.98	4.22	5.20	4.64	4.76	10,10;100	5.08	5.14	4.93	5.53	
			$\mathbf{n_1 = 200, n_2 = 400}$						$\mathbf{n_1 = 200, n_2 = 400}$		
1,2; 200	5.08	5.18	4.68	5.04	4.88	1,2; 200	4.98	5.23	5.71	6.64	9.34
2,4; 100	5.26	4.74	5.06	4.98	4.14	2,4; 100	5.14	4.99	5.34	7.09	
3,6; 66	5.00	4.14	4.48	4.18	4.52	3,6; 66	5.15	6.40	12.10	31.27	
4,8; 50	5.18	4.82	5.04	4.40	4.54	4,8; 50	4.90	7.68	15.23		
5,10; 40	4.02	5.04	3.98	4.20	4.44	5,10; 40	4.74	4.71	5.59		
			$\mathbf{n_1 = 600, n_2 = 1000}$						$\mathbf{n_1 = 600, n_2 = 1000}$		
2,3; 300	4.46	5.16	5.66	5.10	5.12	2,3; 300	4.94	4.86	5.19	5.26	6.21
3,5; 200	5.00	5.80	5.30	4.90	4.66	3,5; 200	5.03	5.15	5.09	5.57	6.51
6,10; 100	5.12	4.74	4.08	4.52	4.30	6,10;100	4.38	4.89	4.95	6.18	

Our test is not sensitive to λ values in terms of its type 1 error rate. So, we considered λ as fixed as 1, 0.1, 0.01, 0.001 and 0.0001 for different time series models. The flexibility of our test statistic is reflected in the size shown in Table 4.1 and Table 4.2. Irrespective of the values of λ, for almost every combination of N, n_1 and n_2 we found that our test results achieved the stable size and they are very close to the theoretical type I error rate leveled at 5% even for small sample sizes. We also performed simulations for very small sample sizes, for instances $n_1 = n_2 = 100$ and $n_1 = 100,\ n_2 = 200$ under both $N(0,1)$ and $ARMA(1,1)$ models (for brevity results are not shown here). A very few of these empirical sizes are a little below the theoretical value. Conversely, some of the previous tests showed some instability in empirical size as it is sensitive to N and J and consequently to the fixed integers p_1 and p_2. In the wavelet-based test in Decowski & Li (2015), many of the combinations failed to obtain empirical size since for those combinations the test can not estimate the Haar wavelet. Also, the test revealed a very large type I error rate in some cases, particularly, in the small-scale sample size.

In addition to the error rate for normal white noise models and ARMA(1,1) models presented in Table 4.1 and Table 4.2, we also did simulation studies for both X_t and Z_t following AR(1) models, and in the case when they both follow MA(1) models. Both of these situations yielded similar type I errors as those found in the white noise and ARMA(1, 1) processes. For the conciseness of this thesis report, the results of the size of the test for an AR(1) with $\phi = 0.5$ and a MA(1) with $\theta = 0.5$ are reported in Table C.1 and in Table C.2. All the above models used standard normal innovations. Unlike wavelet-based tests, we found that our test statistic is not sensitive to the above-generating processes. When N is extremely small, the test statistic does not always converge to the theoretical level but its limit is not too far away. So, when sample size n_1, n_2 are small, we recommend choosing moderately larger N, say above 50, because the actual length of observed log-ratio of periodogram values is equal to $n = [(N-1)/2]$.

Furthermore, we also wanted to check the convergence of our proposed test statistic to its asymptotic null distribution by the empirical size simulation of the test on some other time series models. Thus, in the same way, we simulated the type I error rate for a seasonal AR(1) model, for both the time series, with frequency 12, and again when both of the series are generated from a higher-order autoregressive model, an AR(3) model. The corresponding empirical type I error rates of the proposed test are documented in Appendix C in Table C.3, and Table C.4, respectively. Both of these tables illustrate a very good convergence of the test in terms of the empirical size of the test, even for very small value of N's, the seasonal model only (i.., $n1 = 200,\ n2 = 400$ with $N = 40,\ 50$), the error rates are not far away from the theoretical value at 5%.

Table 4.2: Empirical type I error rates in percentage for for ARMA(1, 1) model

$p_1, p_2; N$	Penalized Regression Splines-based Test					$p_1, p_2; N$	Wavelet -based Test				
	$\lambda = 1$	$\lambda = .1$	$\lambda = .01$	$\lambda = .001$	$\lambda = .0001$		J=1	J=2	J=3	J=4	J=5
			$\mathbf{n_1 = n_2 = 300}$						$\mathbf{n_1 = n_2 = 300}$		
1,1; 300	5.06	5.16	5.18	5.08	4.90	1,1; 300	5.15	5.22	5.46	5.90	7.31
2,2; 150	5.40	5.08	5.12	4.46	5.26	2,2; 150	5.46	5.44	5.28	6.61	8.36
3,3; 100	5.12	4.80	4.78	4.66	4.42	3,3; 100	4.98	4.95	5.12	6.61	
4,4; 75	4.62	4.56	4.82	4.24	5.04	4,4; 75	5.02	4.81	5.56	6.48	
5,5; 60	5.02	4.66	4.72	4.38	4.22	5,5; 60	4.89	5.13	5.10		
			$\mathbf{n_1 = n_2 = 1000}$						$\mathbf{n_1 = n_2 = 1000}$		
1,1; 1000	5.72	5.88	5.60	4.82	5.82	1,1; 1000	4.92	4.92	5.13	5.39	6.02
2,2; 500	5.16	5.30	4.98	5.34	4.68	2,2; 500	5.03	5.19	5.03	4.84	5.60
3,3; 333	5.04	4.24	5.34	5.20	4.68	3,3; 333	4.97	4.81	5.43	5.37	6.34
4,4; 250	4.68	4.30	4.92	5.38	4.32	4,4; 250	5.16	5.04	5.95	6.51	7.75
5,5; 200	5.04	4.02	5.00	4.52	4.50	5,5; 200	4.66	5.18	5.15	5.01	5.99
6,6; 166	4.92	4.56	4.70	4.50	4.70	6,6; 166	4.68	4.84	4.85	5.42	7.32
8,8; 125	5.94	5.08	5.24	4.98	4.92	8,8; 125	4.79	5.11	5.27	5.70	
10,10; 100	4.86	4.92	4.92	4.02	4.06	10,10;100	4.50	5.69	4.92	5.54	
			$\mathbf{n_1 = 200, n_2 = 400}$						$\mathbf{n_1 = 200, n_2 = 400}$		
1,2; 200	5.28	5.26	5.18	5.52	5.70	1,2; 200	5.41	5.68	5.91	6.88	9.47
2,4; 100	5.24	4.34	4.80	5.68	5.02	2,4; 100	4.53	4.95	5.44	7.31	
3,6; 66	4.04	4.66	4.74	4.78	4.92	3,6; 66	5.09	7.01	12.71	31.12	
4,8; 50	5.08	4.24	4.06	4.30	4.88	4,8; 50	5.50	7.71	14.34		
5,10; 40	4.10	4.80	4.62	4.56	4.12	5,10; 40	5.26	4.77	5.30		
			$\mathbf{n_1 = 600, n_2 = 1000}$						$\mathbf{n_1 = 600, n_2 = 1000}$		
2,3; 300	4.98	4.86	4.68	4.82	4.96	2,3; 300	4.89	5.07	5.17	5.41	6.05
3,5; 200	5.10	5.34	5.24	4.62	4.62	3,5; 200	4.88	4.95	5.18	5.43	6.53
6,10; 100	4.90	4.92	4.78	4.24	4.62	6,10;100	4.98	4.65	5.07	6.18	

4.3 Current Test Statistics and Power Comparisons

We computed the power of our penalized splines regression-based test statistic T_n under the alternative hypothesis that the two independent time series are different. In this section we also compare the power of our proposed test to the powers of the recent time series tests from Jin (2011), Lu & Li (2013), Decowski & Li (2015), and Li & Lu (2018). The comparisons will follow brief introductions of these published papers. For more details, refer to their published papers.

- Jin (2011) introduced a data driven time-series test in the frequency domain to compare multiple time series. He considered the function $\psi(\omega) = \log(f_X(\omega)/f_Y(\omega))$, which was estimated using the discrete samples $D_k = \ln(I_X(\omega_k)/I_Y(\omega_k))$, where $I_X(\omega_k)$ is the tapered periodogram of time series $\{X_t\}$. In the case of unequal series lengths, assuming $n_X < n_Y$, the periodogram values of $I_Y(\omega_k)$ were adjusted

$$I_Y^A(\omega_k) = \frac{\sum_{i=1}^{\lfloor N_Y/2 \rfloor} I_Y(2\pi i/N_Y) I_{[\omega_k - \pi/N_X,\ \omega_k + \pi/N_X)}(2\pi i/N_Y)}{\sum_{i=1}^{\lfloor N_Y/2 \rfloor} I_{[\omega_k - \pi/N_X,\ \omega_k + \pi/N_X)}(2\pi i/N_Y)},$$

  so $I_Y^A(\omega_k)$ have the same Fourier frequencies as the shorter periodogram $\{X_t\}$.

  Let $L_k = \ln(I_X(\omega_k)/I_Y^A(\omega_k))$ for $k = 1, \ldots, \lfloor (n_X - 1)/2 \rfloor$, $v = \{1, v_1, v_2, \ldots\}$ be the set of Legendre polynomials and define the model $E(\mathbf{L}) = \beta_0 1 + \beta_1 v_1 + \beta_2 v_2 + ... + \beta_r v_r$. Through regression, the least square estimate $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}'_{r2})'$ can be obtained along with its variance-covariance matrix $\tilde{P}$. Under the null $H_0 : \psi(\omega) = c$, Jin (2011) showed

$$T_r = \hat{\beta}'_{r2} \tilde{P}_{r,2}^{-1} \hat{\beta}_{r2} \xrightarrow{d} \chi_r^2$$

  as $n \to \infty$, where $\hat{\beta}_{r2}$ and $\tilde{P}_{r,2}^{-1}$ are the lower r vector of $\hat{\beta}$ and lower-right $r \times r$ sub-matrix of $\tilde{P}$ respectively. In practice, the user wouldn't know how many basis elements

$(r+1)$ to include in the model. Thus, Jin (2011) used an AIC adaptive method to bypass this parameter selection and proposed

$$D_r = \max_{1 \le r \le R} T_r - 2r.$$

Note the parameter R is selected by the user and the distribution of D_r is found via simulations. One rejects the null if D_r is greater than the $(1-\alpha)$ percentile of the simulated distribution. Jin demonstrated that his proposed test statistic performed about the same as Lund et al. (2009) auto-covariance test when innovations were symmetric, and slightly outperformed when innovations were generated from a skewed distribution.

- Lu & Li (2013) proposed a time series test in the time domain. Building on the test C_L proposed in Lund et al. (2009), Lu and Li proposed

$$C^*_{AN} = \max_{0 \le L \le L_0} \frac{C_L - (L+1)}{\sqrt{2(L+1)}},$$

where L_0 is the maximum dimension tested and the lag length L is selected adaptively. Since a large value of L_0 is too computationally intensive, they used $L_0 = 20$ and declared it to be large enough to capture differences between two short memory processes. A test statistic C_{AN}, the normalized version of C^*_{AN}, was used in their comparison.

Using Fan (1996)'s Neyman adaptive approach Lu & Li (2013) constructed another test in the frequency domain. Similar to Jin (2011), the log ratio of the periodogram values $D_i = \ln(I_X(\omega_i)/I_Y(\omega_i))$, for $i = 1, 2, ..., N$, represent discrete samples from the function $f(\omega) = \ln(f_X(\omega)/f_Y(\omega)), \omega \in (0, \pi)$. A Fourier transform of $\{D_i\}$ (or $\{\bar{D}_i\}$) yielded coefficients $\{D^*_i\}$ (or $\{\bar{D}^*_i\}$) and made the adaptive test more powerful since $\{D^*_i\}$ (or $\{\bar{D}^*_i\}$) are sorted in ascending frequencies. Under the null $H_0 : f(\omega) = 0$,

their adaptive test statistic becomes

$$\bar{T}^*_{AN} = \max_{1\le k\le N_m} \frac{1}{\sqrt{k\hat{\sigma}_2^2}} \sum_{i=1}^{k} (\bar{D}^*_i)^2 - \hat{\sigma}_1^2.$$

Here, $\sigma_m^2 = \mathrm{Var}(\ln F_{2m,2m})$ and for the details on $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, refer to the published article. Similar to C^*_{AN}, $\bar{T}^*_{AN}$ can be normalized into $\bar{T}_{AN}$ and critical values can be determined via simulation.

- Decowski & Li (2015) proposed a test to compare two time series of unequal lengths based on the wavelet transform from two independent series X_t and Y_t. Their proposed linear wavelet estimator of the mean regression function $g(x)$ (same as $f(x)$ in (2.2)) is

$$\hat{g}(x) = \hat{\alpha}\phi(x) + \sum_{j=0}^{J_n} \sum_{k=0}^{2^j-1} \hat{\beta}_{jk}\psi_{jk}(x),$$

where

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} Y_i\phi(x_i), \qquad \textit{and} \quad \hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} Y_i\psi_{jk}(x_i)$$

are the empirical wavelet coefficients. The smoothing parameter J_n, the resolution size, satisfies $2^{J_n+1} \le n$ (for details please refer to the article). Since $\hat{\alpha}$ and $\hat{\beta}_{jk}$ are consistent estimators of the α and β_{jk}'s, under the null hypothesis of equal spectral densities, they proposed their wavelet-based test as

$$T = (\hat{\alpha} - v_0)^2 + \sum_{j=0}^{J_n} \sum_{k=0}^{2^j-1} \hat{\beta}_{jk}^2,$$

for some large J_n. The normalized version of their test statistic is

$$X^2 = \frac{n}{\sigma_0^2}\left[\left(\hat{\alpha} - v_0\right)^2 + \sum_{j=0}^{J} \sum_{k=0}^{2^j-1} \hat{\beta}_{jk}^2\right] \xrightarrow{d} \chi^2(2^{J+1}), \quad \text{as } n \to \infty.$$

- A new test statistic has been proposed by Jin & Wang (2016) in the time domain for checking equality of the correlation structures of two independent time series. In particular, they proposed the following order selection test statistic:

$$\hat{S}_R = \max_{1 \le r \le R} \left(\frac{T_X T_Y}{T_X + T_Y} \Delta_r^T \hat{\Gamma}_r^{-1} \Delta_r - 2r \right),$$

where Δ_r is the difference between the two sample autocorrelation functions of two time series having lengths T_X and T_Y, respectively. $\hat{\Gamma}_r$ is a consistent estimate of Γ_r, which is defined as $(T_Y \Gamma_X + T_X \Gamma_Y)/(T_X + T_Y)$ where Γ_X and Γ_Y are the asymptotic covariance of sample autocorrelation functions of the two given time series. The parameter R controls the maximum dimension tested and is selected in the simulation study to meet the asymptotic assumption in its limit theory.

Though their proposed test statistic is more powerful than existing test statistics, their test has certain limitations. They derived the limit distribution of the proposed test statistic under the null hypothesis with an unconventional probability distribution. The type I error is seriously distorted using the asymptotic critical value because the convergence rate of their test statistic $\hat{S}_R$ to its limit distribution is very slow. So, one has to rely on Monte Carlo simulations to determine the critical values under the null hypothesis for each given sample size, which may depend on the sample sizes of the time series and the underlying time series models as well.

- Most recently, Li & Lu (2018) provided a test statistic based on the difference between two wavelet-based estimates for assessing whether two stationary and independent linear processes, $\{X_t\}$ and $\{Y_t\}$, with unequal lengths have the same spectral densities. Their wavelet-based estimator for the spectral density f_X can be expressed as

$$\hat{f}_X(w) = \hat{\alpha}_{00,X} \frac{1}{\sqrt{2\pi}} + \sum_{j=0}^{J_{n_1}} \sum_{k=0}^{2^j - 1} \hat{\beta}_{jk,X} \Psi_{jk}(w), \quad w \in [-\pi, \pi],$$

where the estimators of wavelet coefficients are:

$$\hat{\alpha}_{00,X} = (n_1^{-1}/\sqrt{2\pi}) \sum_{h=-\infty}^{\infty} (X_t - \bar{X})^2 \quad \text{and}$$

$$\hat{\beta}_{jk,X} = 1/\sqrt{2\pi}) \sum_{h=-(n_1-1)}^{n_1-1} \hat{r}_X(h)\hat{\Psi}_{jk}(h).$$

Here, J_{n_1} satisfies $2^{J_{n_1}+1} \leq n_1 < 2^{J_{n_1}+2}$ with $\hat{r}_X(h) = n_1^{-1} \sum_{t=|h|+1}^{n_1} (X_t - \bar{X})(X_{t-|h|} - \bar{X})$. Define, $\hat{V}_{J,X}$ and $V_{J,X}$ be the subsets of the empirical and the true wavelet coefficients up to the resolution level J, where J is a fixed integer. Their proposed test statistic is defined as

$$X_J^2 = (2\pi n_1) \left(\hat{V}_{J,X} - \hat{V}_{J,Y}\right)^T \hat{\Sigma}^{-1} \left(\hat{V}_{J,X} - \hat{V}_{J,Y}\right),$$

where $\hat{\Sigma}$ is the variance-covariance matrix for the difference of the above two subsets of empirical wavelet coefficients (i.e.,$\hat{V}_{J,X} - \hat{V}_{J,Y}$). Under the null hypothesis $H_0 : V_{J,X} = V_{J,Y}$, for all J,

$$X_J^2 \longrightarrow_d \chi^2(2^J), \quad \text{as } n_1 = \rho n_2 \to \infty,$$

for any fixed $J(1 \leq J < \min\{J_{n_1}, J_{n_2}\})$. Though this test can be used for two time series with arbitrary lengths without throwing away any small number of observations, it involved too many parameters and has unstable levels in moderately small samples. Also, the power of this test is noticeably varied with some parameters, particularly, the resolution size J.

Now, for power comparison here we consider three specific examples as in Decowski & Li (2015): two from non-seasonal time series and one from seasonal time series. The critical values are calculated in the same process as discussed in section 4.2. Since we consider the λ as fixed, the critical value does not depend on the data. So, for a fixed λ irrespective of

the time series parameters ϕ (for AR models) and θ (for MA model), the critical values will be the same across different examples as long as the lengths of the data series are the same. In the following Examples 4.1, 4.2, and 4.3, we simulated empirical critical values of our test statistic in the same way we did in the case of size estimation. Assuming the error terms in the data model are the realization of an $i.i.d.$ distribution with zero mean and constant variance, under the alternative hypothesis that the two time series are different, the test statistic T_n in (3.13) is calculated repeatedly for 10000 times for every combination of N and λ. The proportion of this test statistic exceeding the empirical critical value is the empirical power.

4.3.1 **Example 4.1**

Two independent time series series $\{X_t\}$ and $\{Z_t\}$ with $n_1 = n_2 = 1024$ were generated from an AR(1) process with an autoregressive parameter ϕ and from an MA(1) process with moving average parameter $\theta = \frac{\phi}{|\phi|}\sqrt{\frac{\phi^2}{1-\phi^2}}$, respectively. Both of these time series had standard normal innovations. This example was first discussed in Lund et al. (2009), and re-visited Lu & Li (2013), Li & Lu (2018), Decowski & Li (2015), and Jin and Wang (2016). Similar to those paper, we let $N = 512$ or $N = 256$. Then we consider the value of the smoothing parameter λ to be fixed at 0.1, 0.01 and 0.001, and computed our penalized regression spline-based statistic $T_{N,\lambda} = T_n$ in (3.13). These results are reported in Table 4.3, and compared with the empirical power of Decowski and Li's (2015)'s wavelet-based test $X^2_{N,J}$, Jin and Wang (2016)'s order selection test $\hat{S}_R$, and Lu and Li (2018)'s Bartlett type wavelet-based test X^2_J. The results from these previous tests are quoted from their published articles. Overall, the penalized spline-based test $T_{N,\lambda}$ performed best, for every combination of N and λ, to the previous tests except for the $\hat{S}_R$ test. A graphical display of this power comparison is shown in Figure 4.5. We noticed that our proposed test is also very comparable to the $\hat{S}_R$ test, specifically, when $N = 256$ and $\lambda = 0.01$ and 0.001. We computed power for other non-seasonal time series of this kind and found similar results. We point out that

our test offers fast implementation and could be extended to compare a series of unequal lengths, which we presented in example 4.3.

Table 4.3: Empirical powers for AR(1) vs. MA(1)

ϕ	Previous Tests							Our Proposed Test		
	$X^2_{512,1}$	$X^2_{256,1}$	$X^2_{256,2}$	$\hat{S}_R$	X^2_1	X^2_2	X^2_3	$T_{512,.01}$	$T_{256,.01}$	$T_{256,.001}$
-0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.50	0.997	0.999	1.000	1.000	0.114	0.998	1.000	1.000	1.000	1.000
-0.25	0.126	0.142	0.118	0.215	0.045	0.141	0.105	0.156	0.166	0.146
0.00	0.050	0.049	0.051	0.049	0.049	0.045	0.037	0.050	0.052	0.049
0.25	0.133	0.142	0.122	0.176	0.047	0.138	0.110	0.148	0.165	0.149
0.375	0.579	0.641	0.583	0.791	0.050	0.681	0.621	0.694	0.763	0.711
0.50	0.997	0.999	0.999	0.999	0.112	0.999	1.000	1.000	1.000	1.000
0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

4.3.2 **Example 4.2**

Here, both series had Gaussian innovations with unit variance and of lengths $n_1 = n_2 \in \{256, 512, 1024\}$. The first series $\{X_t\} \sim N(0, 1)$, and then the second series $\{Z_t\}$ was an AR(1) process with frequency 12 and seasonal AR parameter ϕ. This example was also performed in Lu & Li (2013) and Decowski & Li (2015). The parameter N was selected to be $N = n_1/2 = n_2/2 = n/2$ or $n_1/4 = n_2/4 = n/4$ so that $p_1 = p_2 = 2$ or $p_1 = p_2 = 4$, respectively. After choosing the value of λ as $10^{-5}, 10^{-6}$ and 10^{-7} the power for our proposed test $T_{N,\lambda}$ is computed and presented in Table 4.4 and Figure 4.6. Our test provides far better power for series with subtle to moderate seasonality. The power for Lu and Li (2013)'s time domain test C_{AN} seems slightly better at $\phi = .25, n = 1024$ and $\phi = 0.50, n = 256$ only. Except for these two cases, for any other combination of λ, ϕ and N our proposed test has the better power than any other previous tests. As shown in Figure 5, our proposed test outperform all other test in powers at $\lambda = 10^{-6}, and N = n/4$. Even, it is comparable to the the C_{AN} test at $\phi = .25, n = 1024$ and $\phi = 0.50, n = 256$. We have also checked the power

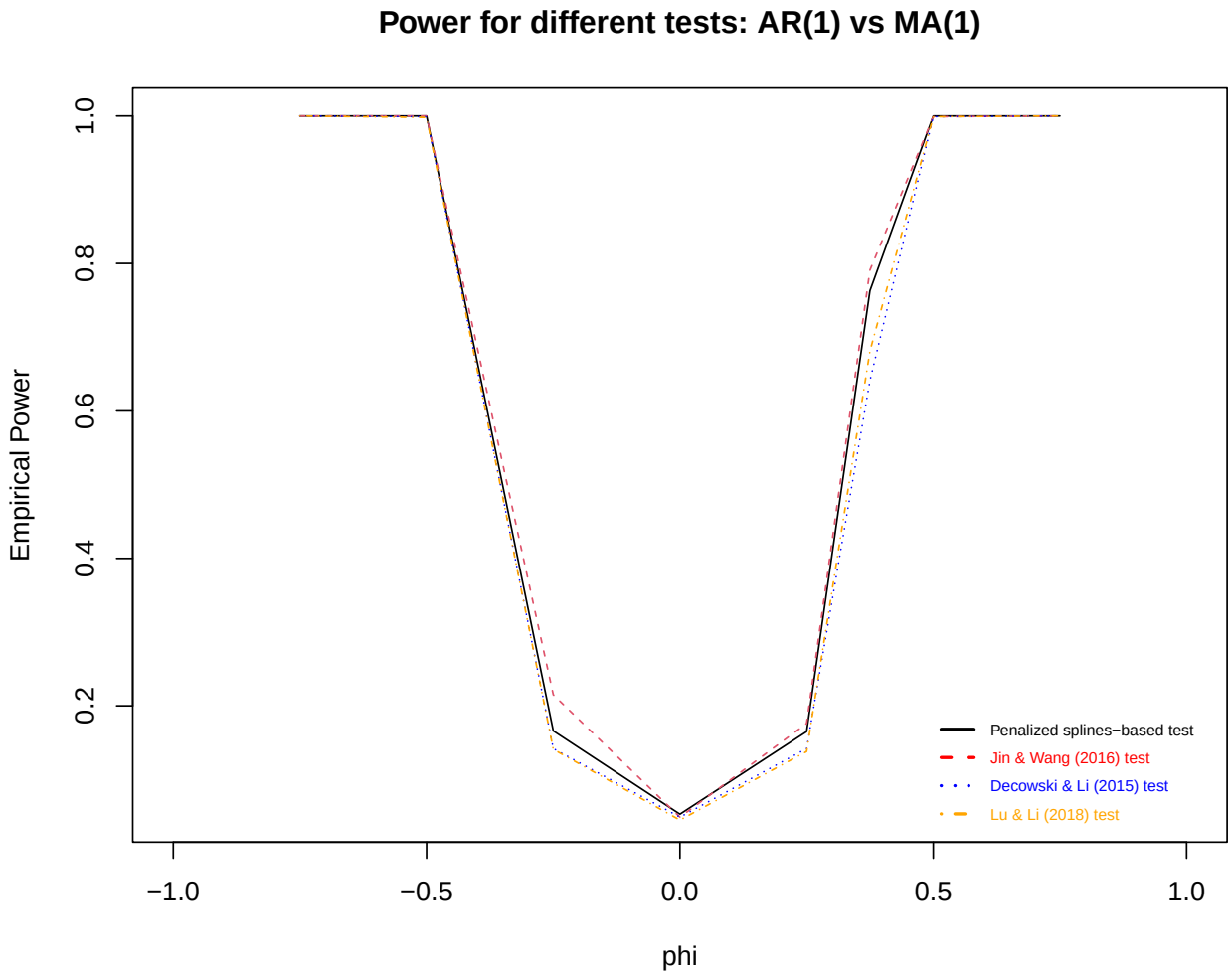

Figure 4.5: Power comparison for our proposed test with current tests: $X_t \sim AR(1)$ model and $Z_t \sim MA(1)$ model.

for another non seasonal series, for instance, a $N(0,1)$ vs. a seasonal MA(1) with frequency 12, in which case we found plausible power of the test for moderately large sample sizes and θ value away from zero. The numerical value of this simulation is documented in Table C.5.

Table 4.4: Empirical powers for N(0, 1) vs. seasonal AR(1) model with frequency 12

ϕ	n	Previous Tests					Our Proposed Tests			
		$\bar{T}_{AN}$	C_{AN}	$X^2_{n/2,3}$	$X^2_{n/4,3}$	$X^2_{n/2,4}$	$T_{n/2,10^{-5}}$	$T_{n/2,10^{-6}}$	$T_{n/4,10^{-6}}$	$T_{n/4,10^{-7}}$
0.10	256	0.054	0.054	0.065	0.615	0.067	0.065	0.071	0.067	0.060
0.10	512	0.064	0.064	0.077	0.076	0.083	0.087	0.093	0.089	0.093
0.10	1024	0.083	0.107	0.114	0.113	0.119	0.156	0.138	0.152	0.124
0.25	256	0.101	0.123	0.139	0.138	0.154	0.161	0.198	0.200	0.205
0.25	512	0.268	0.420	0.302	0.309	0.348	0.438	0.432	0.507	0.482
0.25	1024	0.686	0.913	0.654	0.698	0.725	0.859	0.816	0.886	0.823
0.50	256	0.372	0.863	0.591	0.627	0.643	0.705	0.812	0.823	0.853
0.50	512	0.931	1.000	0.950	0.978	0.973	0.997	0.996	0.999	0.998
0.50	1024	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

4.3.3 **Example 4.3**

This example illustrates the power performances of our proposed test for two series of unequal lengths. The data from two time series of unequal lengths n_1 and n_2 were generated from a $N(0,1)$ process and an AR(1) process with parameter ϕ. Both the series have standard normal innovations. In the absence of seasonality, we considered the λ values fixed as 0.1 and 0.01. The value of N was chosen appropriately based on the series lengths. Here, the simulation was performed in Jin's (2011) test $D_{R,p}$, where p is the tapering rate from the series using a Cosine-Bell taper. Also, it is discussed in Decowski and Li's (2015)'s wavelet-based test $X^2_{N,J}$, Jin and Wang's (2016) test $\hat{S}_R$, and Lu and Li's (2018) test X^2_J. The empirical wavelet coefficients in Lu and Li's (2018) test do not approximate normal limit distribution very well for small sample sizes, so relatively larger-sized time series are considered in this example for an appropriate comparison. Table 4.5 summarizes the empirical power in mod-

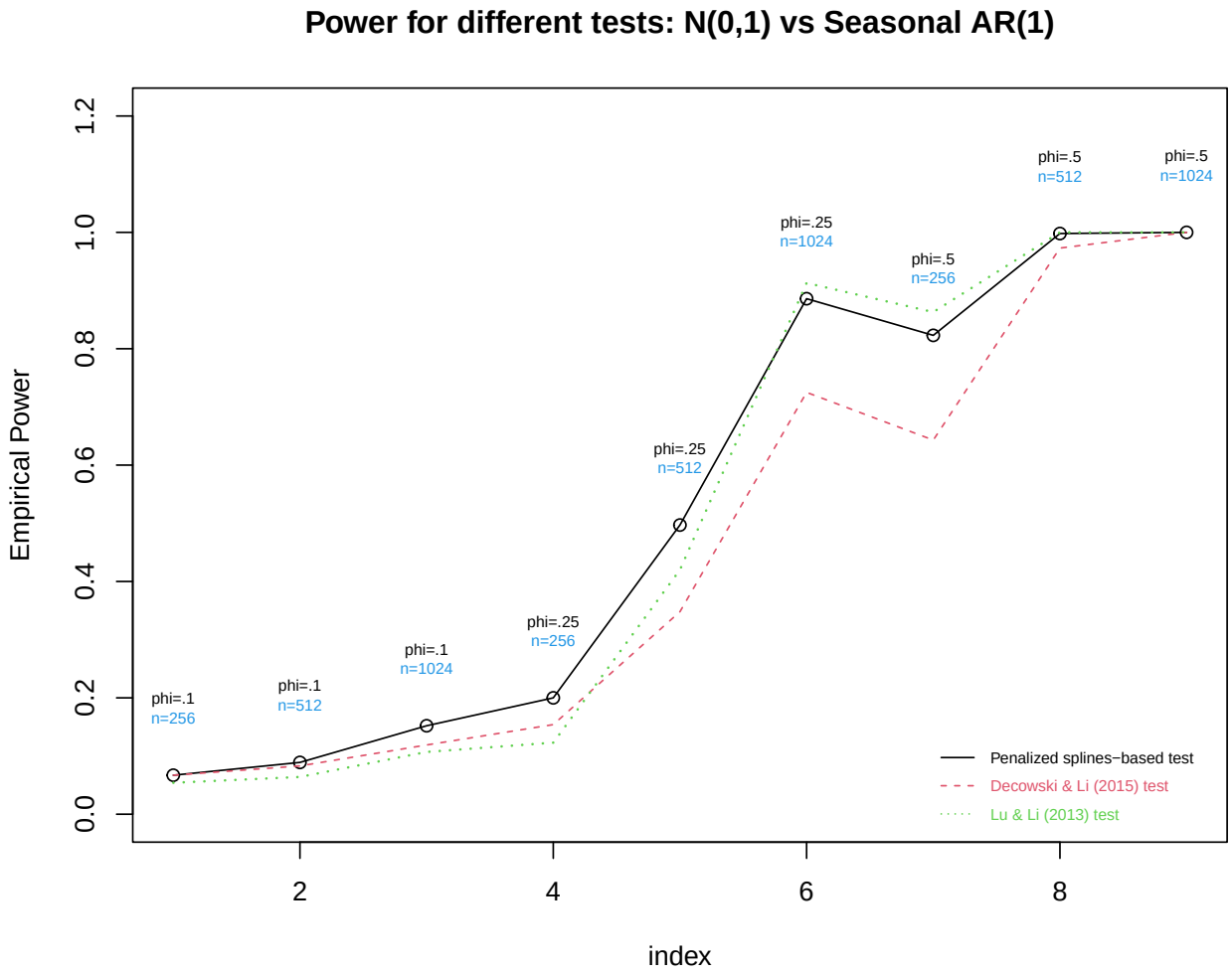

Figure 4.6: Power comparison of our proposed test with previous tests: $X_t \sim N(0,1)$ process and Z_t is a Seasoanl $AR(1)_{12}$ model.

erately large samples of the proposed test $T_{N,\lambda}$, and those of the previous tests quoted from their published papers. For every test, The proportion of rejection tends to 100% as the lengths of the series become larger and(or) the magnitude of the autoregressive parameter ϕ gets away from zero. Our proposed penalized regression spline-based test $T_{N,\lambda}$ yields better results than all previous tests except the $\hat{S}_R$ test. Also, for almost every combination of λ, ϕ, and N, the empirical power of our proposed test is very comparable to that of the $\hat{S}_R$ test. This comparison of power performances is also pictured in Figure 4.7.

Table 4.5: Empirical Powers for N(0,1) vs. AR(1) processes in large samples

ϕ	Previous Tests							Our Proposed Test		
	$D_{5,5\%}$	$X^2_{128,1}$	$X^2_{64,1}$	$\hat{S}_R$	X^2_1	X^2_2	X^2_3	$T_{128,.1}$	$T_{64,.1}$	$T_{64,.01}$
	$\mathbf{n_1 = 256, n_2 = 384}$									
0	4.5	4.7	4.7	5.7	4.3	3.6	2.0	5.5	4.5	4.2
0.1	12.5	10.7	11.2	17.8	13.4	9.8	5.6	15.0	14.8	14.3
0.2	36.2	35.1	38.0	58.6	47.0	39.2	23.2	48.4	50.7	48.7
0.4	93.4	94.8	96.2	100.0	98.4	97.8	93.0	98.4	98.9	98.8
0.6	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$\mathbf{n_1 = 256, n_2 = 512}$									
0	6.1	5.1	5.0	4.6	4.7	4.3	3.5	4.8	4.6	4.5
0.1	14.5	11.6	11.7	23.9	15.4	12.3	8.5	16.7	16.5	15.0
0.2	47.9	38.8	41.9	65.3	51.4	45.1	31.1	53.0	57.6	51.7
0.4	97.4	97.0	97.8	99.9	99.3	99.2	97.3	99.4	99.5	99.5
0.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Again the power performances for this particular example in small-sized time series are presented in Figure 4.8. The numerical values of this comparison is provided in Table C.6 in the Appendix along with the Jin (2011) and Decowski & Li (2015) tests. Here, we have seen that our test showed the best performance in terms of its proportion of rejections under the alternative hypothesis in comparison to the existing tests, irrespective of the equal vs. unequal time series.

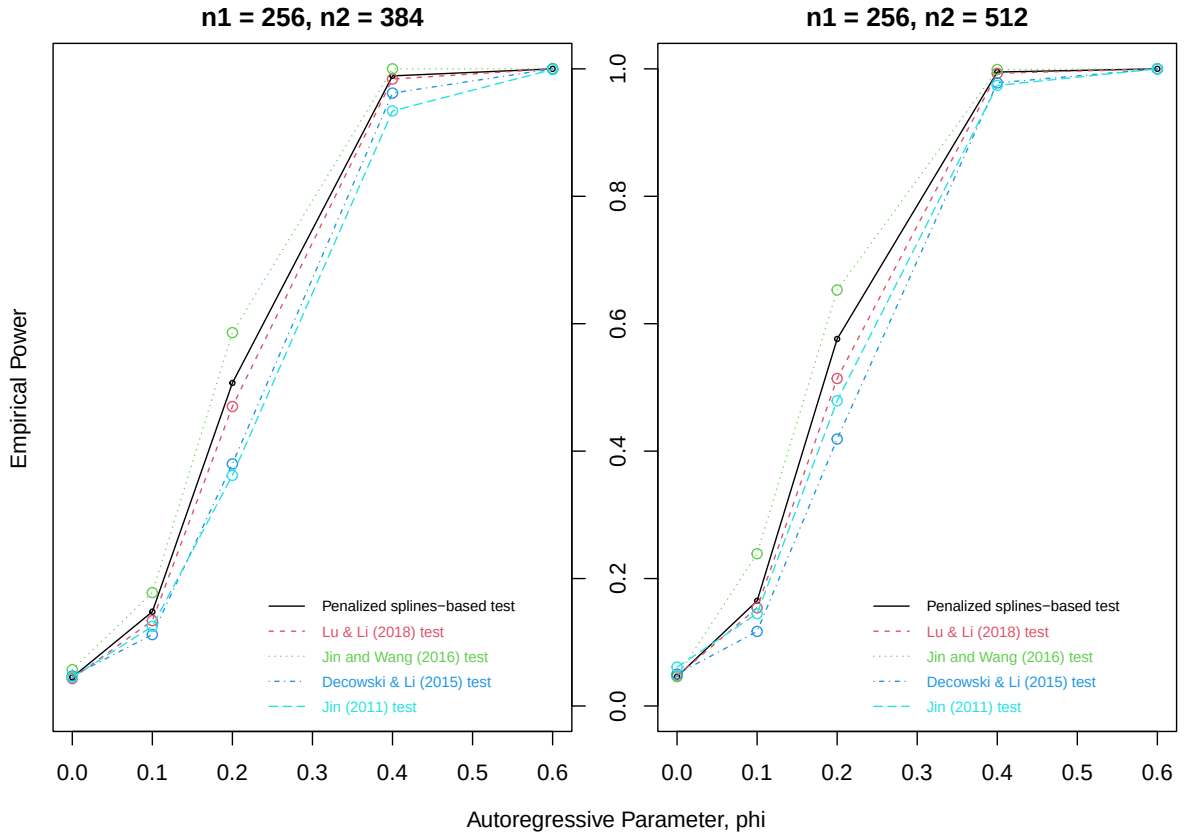

Figure 4.7: Power comparison of our proposed test with previous tests: $X_t \sim N(0,1)$ process and Z_t is a $AR(1)$ model.

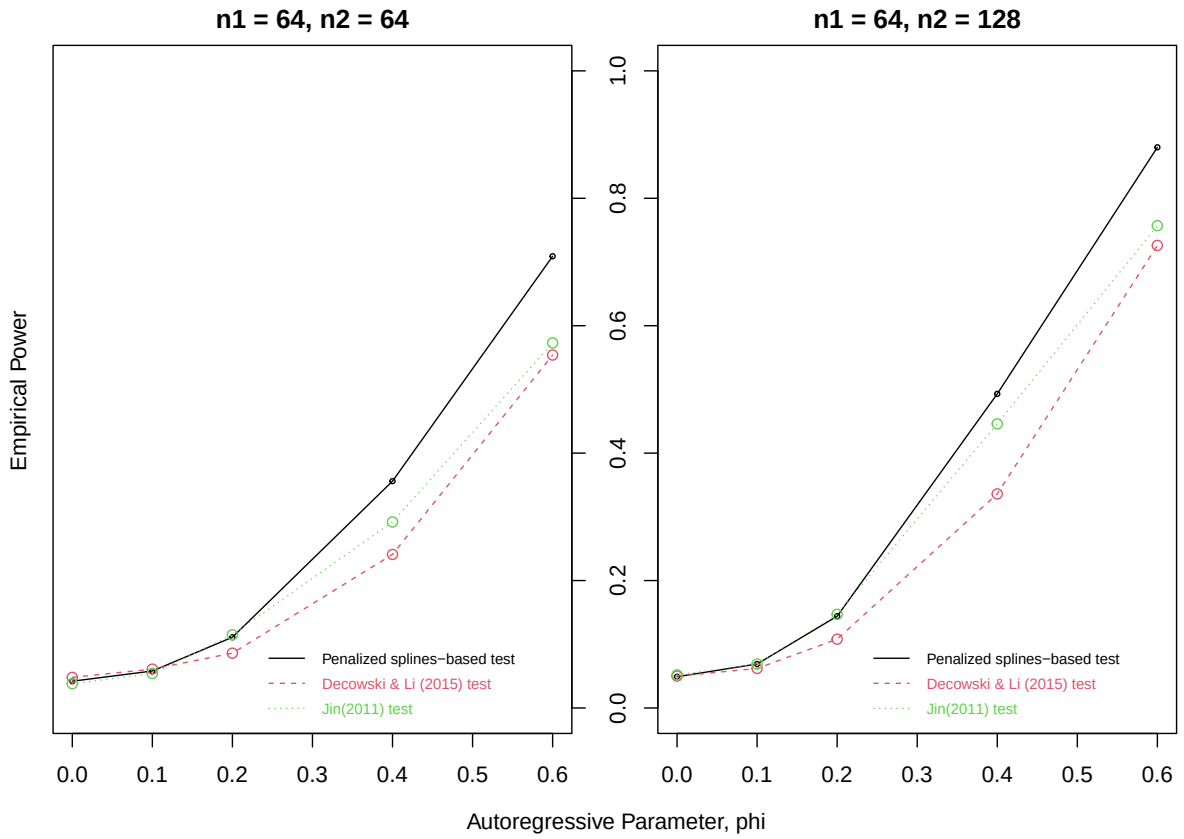

Figure 4.8: Power comparison of our proposed test with previous tests: $X_t \sim N(0,1)$ process and Z_t is a $AR(1)$ model.

4.3.4 **Example 4.4**

Finally, besides the above three specific examples, we were curious to check the performance of our proposed test in terms of the empirical power for some other time-series models, which were not listed in the previous tests. We choose an $AR(1)$ model for the series X_t against a seasonal moving average model with frequency 12 for the series Z_t. The value of the parameters ϕ, and the sample sizes n_1 and n_2 were considered the same as in the example 4.3. The value of the seasonal moving average parameter is determined by the expression $\theta = \frac{\phi}{|\phi|}\sqrt{\frac{\phi^2}{1-\phi^2}}$. We suspected that the observed data for this scenario does not have much more seasonality, because our observed data is the ratio of two spectral estimates, obtained from the respective time series, which reduces the amount of seasonality to some extent. Therefore, for this particular example we considered the value of $\lambda = 1.00, 0.1, 0.01, 0.001, 0.0001$ and 0.00001 as fixed. The empirical power of the proposed test seems to be satisfactorily large and is mostly insensitive to the value of the smoothing parameter. Likewise to the previous examples the test has more power for larger values of the autoregressive parameter and their corresponding moving average parameter. The simulated numerical result of this particular example is presented in Table 4.6.

Table 4.6: Empirical Powers of our proposed test for $AR(1)$ vs. $ARMA(0,1)_{12}$ processes.

ϕ	n	$\lambda = 1.00$	$\lambda = 0.10$	$\lambda = 0.01$	$\lambda = 10^{-3}$	$\lambda = 10^{-4}$	$\lambda = 10^{-5}$
				$\mathbf{N = n/2}$			
0.10	256	0.126	0.125	0.114	0.105	0.089	0.088
0.10	512	0.212	0.223	0.189	0.150	0.139	0.163
0.10	1024	0.397	0.395	0.335	0.281	0.295	0.311
0.258	256	0.568	0.563	0.506	0.457	0.408	0.484
0.258	512	0.877	0.871	0.837	0.776	0.804	0.886
0.258	1024	0.996	0.994	0.990	0.983	0.997	0.999
0.578	256	0.995	0.995	0.994	0.988	0.987	0.999
0.578	512	1.000	1.000	1.000	1.000	1.000	1.000
0.578	1024	1.000	1.000	1.000	1.000	1.000	1.000
				$\mathbf{N = n/4}$			
0.10	256	0.122	0.118	0.117	0.114	0.084	0.085
0.10	512	0.222	0.240	0.186	0.158	0.153	0.172
0.10	1024	0.437	0.422	0.369	0.326	0.303	0.349
0.258	256	0.607	0.619	0.559	0.515	0.451	0.471
0.258	512	0.904	0.914	0.878	0.843	0.802	0.924
0.258	1024	0.998	0.997	0.995	0.994	0.997	0.999
0.578	256	0.998	0.998	0.998	0.995	0.990	0.998
0.578	512	1.000	1.000	1.000	1.000	1.000	1.000
0.578	1024	1.000	1.000	1.000	1.000	1.000	1.000

CHAPTER 5

Concluding Remarks and Future Work

In this thesis, particular attention has been focused on penalized regression splines-based tests for assessing whether two stationary and independent time series with different sample sizes have the same spectral densities or autocovariance functions. This test statistic is based on the penalized spline coefficients of the log ratio of the spectral densities of the two time series. In this study, we considered the number of knots at $K = min([n/4] + 5, 40)$. But we also checked the empirical results on a smaller number of knots at $K = min([n/4], 40)$ and found a negligible difference in terms of the power and type I error rate to the former knot size. The performance of our penalized splines-based test performed almost the same as Jin & Wang (2016) test. But, the Jin & Wang (2016) test is based on the time-domain approach and has certain limitations which were discussed in Section 4.3. Our proposed test outperformed all other previous tests with respect to its proportion of rejections under both null and alternative hypotheses. This test is computationally fast and attained stable empirical size for various time series models under small and large scale sample sizes. Simulation studies also revealed that our proposed test statistic offered very competitive empirical powers for both seasonal and non-seasonal time series models with a reasonable choice of smoothing parameter λ.

In this research, we proved the asymptotic behavior of the theoretical properties of our proposed test statistic. The asymptotic case of our test statistic does not have any known form of theoretical distribution, because the test statistic contains singular values, equal to the basis dimension, which are not fixed quantities for different knot locations. It seems under

the null hypothesis, the proposed test statistic follows a mixture of normal distributions, on which we need to work in the near future. However, the critical value of our test statistic is computed very quickly, because it does not depend on the observed value of the log-spectral ratios, obtained from two assumed time series, which partially serve the purposes of the asymptotic known distributional form of the test statistic. Another drawback of our new test statistic is that we did not conduct a simulation study to obtain empirical power on higher-order time series models. However, it is suspected that for independent and stationary time series models whose spectrum is very smooth and is less heterogeneous we discussed in the previous section, our proposed method seems to work very well. Also, in the near future, we will implement our testing procedure to the time series data of applied statistical problems, as we discussed in the first paragraph of Chapter 1.

In conclusion, in terms of simplicity and good empirical results, our proposed penalized regression spline-based test is a better choice for testing two time series of both equal and unequal lengths.

Bibliography

Besse, P. C., Cardot, H. & Ferraty, F. (1997), 'Simultaneous non-parametric regressions of unbalanced longitudinal data', *Computational Statistics & Data Analysis* **24**(3), 255–270.

Brockwell, P. J. & Davis, R. A. (1991), Stationary time series, *in* 'Time Series: Theory and Methods', Springer, pp. 1–41.

Cai, Z., Fan, J. & Li, R. (2000), 'Efficient estimation and inferences for varying-coefficient models', *Journal of the American Statistical Association* **95**(451), 888–902.

Caiado, J., Crato, N. & Peña, D. (2006), 'A periodogram-based metric for time series classification', *Computational Statistics & Data Analysis* **50**(10), 2668–2684.

Caiado, J., Crato, N. & Peña, D. (2009), 'Comparison of times series with unequal length in the frequency domain', *Communications in Statistics—Simulation and Computation®* **38**(3), 527–540.

Caiado, J., Crato, N. & Peña, D. (2012), 'Tests for comparing time series of unequal lengths', *Journal of Statistical Computation and Simulation* **82**(12), 1715–1725.

Cantoni, E. & Hastie, T. (2002), 'Degrees-of-freedom tests for smoothing splines', *Biometrika* **89**(2), 251–263.

Carmona, R. A. & Wang, A. (1996), Comparison tests for the spectra of dependent multivariate time series, *in* 'Stochastic Modelling in Physical Oceanography', Springer, pp. 69–88.

Chen, H. & Wang, Y. (2011), 'A penalized spline approach to functional mixed effects model analysis', *Biometrics* **67**(3), 861–870.

Chen, H., Wang, Y., Li, R. & Shear, K. (2014), 'A note on a nonparametric regression test through penalized splines', *Statistica Sinica* **24**, 1143.

Chen, J.-C. (1994), 'Testing goodness of fit of polynomial models via spline smoothing techniques', *Statistics & Probability Letters* **19**(1), 65–76.

Claeskens, G., Krivobokova, T. & Opsomer, J. D. (2009), 'Asymptotic properties of penalized spline estimators', *Biometrika* **96**(3), 529–544.

Coates, D. & Diggle, P. (1986), 'Tests for comparing two estimated spectral densities', *Journal of Time Series Analysis* **7**, 7–20.

Cooley, J. W. & Tukey, J. W. (1965), 'An algorithm for the machine calculation of complex fourier series', *Mathematics of computation* **19**(90), 297–301.

Cox, D. & Koh, E. (1989), 'A smoothing spline based test of model adequacy in polynomial regression', *Annals of the Institute of Statistical Mathematics* **41**(2), 383–400.

Cox, D., Koh, E., Wahba, G. & Yandell, B. S. (1988), 'Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models', *The Annals of Statistics* pp. 113–119.

Crainiceanu, C. M. & Ruppert, D. (2004), 'Likelihood ratio tests in linear mixed models with one variance component', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(1), 165–185.

Crainiceanu, C., Ruppert, D., Claeskens, G. & Wand, M. P. (2005), 'Exact likelihood ratio tests for penalised splines', *Biometrika* **92**(1), 91–103.

De Boor, C. (1978), *A practical guide to splines*, Vol. 27, springer-verlag New York.

De Boor, C. (2001), 'A practical guide to splines (revised edition)'.

Decowski, J. & Li, L. (2015), 'Wavelet-based tests for comparing two time series with unequal lengths', *Journal of Time Series Analysis* **36**(2), 189–208.

Dette, H., Kinsvater, T. & Vetter, M. (2011), 'Testing non-parametric hypotheses for stationary processes by estimating minimal distances', *Journal of Time Series Analysis* **32**(5), 447–461.

Dette, H. & Paparoditis, E. (2009), 'Bootstrapping frequency domain tests in multivariate time series with an application to comparing spectral densities', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(4), 831–857.

Dette, H., Preuß, P. & Vetter, M. (2011b), 'A measure of stationarity in locally stationary processes with applications to testing', *Journal of the American Statistical Association* **106**(495), 1113–1124.

Diggle, P. J. & Fisher, N. I. (1991), 'Nonparametric comparison of cumulative periodograms', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **40**, 423–434.

Eichler, M. (2008), 'Testing nonparametric and semiparametric hypotheses in vector stationary processes', *Journal of Multivariate Analysis* **99**(5), 968–1009.

Eilers, P. H. & Marx, B. D. (1996), 'Flexible smoothing with b-splines and penalties', *Statistical science* **11**(2), 89–121.

Eilers, P. H. & Marx, B. D. (2010), 'Splines, knots, and penalties', *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(6), 637–653.

Eubank, R. L. & Spiegelman, C. H. (1990), 'Testing the goodness of fit of a linear model via nonparametric regression techniques', *Journal of the American Statistical Association* **85**(410), 387–392.

Fan, J. (1996), 'Test of significance based on wavelet thresholding and neyman's truncation', *Journal of the American Statistical Association* **91**(434), 674–688.

Fan, J., Zhang, C. & Zhang, J. (2001), 'Generalized likelihood ratio statistics and wilks phenomenon', *The Annals of statistics* **29**(1), 153–193.

Ferguson, T. S. (1996), *A course in large sample theory*, chapman and hall.

Green, P. J. & Silverman, B. W. (1993), *Nonparametric regression and generalized linear models: a roughness penalty approach*, Crc Press.

Gu, C. (2002), *Smoothing spline ANOVA models*, Vol. 297, Springer.

Hastie, T. & Tibshirani, R. (1987), 'Generalized additive models: some applications', *Journal of the American Statistical Association* **82**(398), 371–386.

Hastie, T., Tibshirani, R. & Friedman, J. H. (2001), *The elements of statistical learning: data mining, inference, and prediction*, Vol. 2, Springer Series in Statistics.

Jayasuriya, B. R. (1996), 'Testing for polynomial regression using nonparametric regression techniques', *Journal of the American Statistical Association* **91**(436), 1626–1631.

Jentsch, C. & Pauly, M. (2012), 'A note on using periodogram-based distances for comparing spectral densities', *Statistics & probability letters* **82**(1), 158–164.

Jentsch, C. & Pauly, M. (2015), 'Testing equality of spectral densities using randomization techniques', *Bernoulli* **21**(2), 697–739.

Jin, L. (2011), 'A data-driven test to compare two or multiple time series', *Computational statistics & data analysis* **55**(6), 2183–2196.

Jin, L., Cai, L. & Wang, S. (2019), 'A computational bootstrap procedure to compare two dependent time series', *Journal of Statistical Computation and Simulation* **89**(15), 2831–2847.

Jin, L. & Wang, S. (2016), 'A new test for checking the equality of the correlation structures of two time series', *Journal of Time Series Analysis* **37**(3), 355–368.

Kelly, C. & Rice, J. (1990), 'Monotone smoothing with application to dose-response curves and the assessment of synergism', *Biometrics* pp. 1071–1085.

Li, L. & Lu, K. (2018), 'Tests for the equality of two processes' spectral densities with unequal lengths using wavelet methods', *Journal of Time Series Analysis* **39**(1), 4–27.

Li, R. & Nie, L. (2008), 'Efficient statistical inference procedures for partially nonlinear models and their applications', *Biometrics* **64**(3), 904–911.

Li, Y. & Ruppert, D. (2008), 'On the asymptotics of penalized splines', *Biometrika* **95**(2), 415–436.

Liu, A. & Wang, Y. (2004), 'Hypothesis testing in smoothing spline models', *Journal of Statistical computation and simulation* **74**(8), 581–597.

Lu, K. & Li, L. (2013), 'On fan's adaptive neyman tests for comparing two spectral densities', *Journal of Statistical Computation and Simulation* **83**(9), 1585–1601.

Lund, R., Bassily, H. & Vidakovic, B. (2009), 'Testing equality of stationary autocovariances', *Journal of Time Series Analysis* **30**(3), 332–348.

Mahmoudi, M. R., Heydari, M. H. & Roohi, R. (2019), 'A new method to compare the spectral densities of two independent periodically correlated time series', *Mathematics and Computers in Simulation* **160**, 103–110.

O'Sullivan, F. (1986), 'A statistical perspective on ill-posed inverse problems', *Statistical science* pp. 502–518.

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M. & Schmid, M. (2019), 'A review of spline function procedures in r', *BMC medical research methodology* **19**(1), 1–16.

Piccolo, D. (1990), 'A distance measure for classifying arima models', *Journal of time series analysis* **11**, 153–164.

Pötscher, B. & Reschenhofer, E. (1988), 'Discriminating between two spectral densities in case of replicated observations', *Journal of Time Series Analysis* **9**, 221–224.

Preuß, P. & Hildebrandt, T. (2013), 'Comparing spectral densities of stationary time series with unequal sample sizes', *Statistics & Probability Letters* **83**(4), 1174–1183.

Racine, J. S. (2014), 'A primer on regression splines', *URL: http://cranrprojectorg/web/packages/crs/vignettes/splineprimerpdf* .

Reinsch, C. H. (1967), 'Smoothing by spline functions', *Numerische mathematik* **10**(3), 177–183.

Royston, P. & Sauerbrei, W. (2008), *Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables*, Vol. 777, John Wiley & Sons.

Ruppert, D. (2002), 'Selecting the number of knots for penalized splines', *Journal of computational and graphical statistics* **11**(4), 735–757.

Ruppert, D., Wand, M. P. & Carroll, R. J. (2003), *Semiparametric regression*, number 12, Cambridge university press.

Ruppert, D., Wand, M. P. & Carroll, R. J. (2009), 'Semiparametric regression during 2003–2007', *Electronic journal of statistics* **3**, 1193.

Schwetlick, H. & Kunert, V. (1993), 'Spline smoothing under constraints on derivatives', *BIT Numerical Mathematics* **33**(3), 512–528.

Shumway, R. H., Stoffer, D. S. & Stoffer, D. S. (2000), *Time series analysis and its applications*, Vol. 3, Springer.

Wahba, G. (1990), *Spline models for observational data*, SIAM.

Wand, M. P. (2000), 'A comparison of regression spline smoothing procedures', *Computational Statistics* **15**(4), 443–462.

Wand, M. P. (2003), 'Smoothing and mixed models', *Computational statistics* **18**(2), 223–249.

Wand, M. P. & Ormerod, J. (2008), 'On semiparametric regression with o'sullivan penalized splines', *Australian & New Zealand Journal of Statistics* **50**(2), 179–198.

Wang, X., Shen, J. & Ruppert, D. (2011), 'On the asymptotics of penalized spline smoothing', *Electronic Journal of Statistics* **5**, 1–17.

Wood, S. N. (2006), *Generalized additive models: an introduction with R*, chapman and hall/CRC.

Wood, S. N. (2017*a*), *Generalized additive models: an introduction with R*, second edition edn, chapman and hall/CRC.

Wood, S. N. (2017*b*), 'P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data', *Statistics and Computing* **27**(4), 985–989.

Xiao, L. (2019), 'Asymptotic theory of penalized splines', *Electronic Journal of Statistics* **13**(1), 747–794.

Zhang, C. (2004), 'Assessing the equivalence of nonparametric regression tests based on spline and local polynomial smoothers', *Journal of statistical planning and inference* **126**(1), 73–95.

Zhou, S., Wolfe, D. & Shen, X. (1998), 'Local asymptotics for regression splines and confidence regions', *The annals of statistics* **26**(5), 1760–1782.

APPENDIX A

Supplementary Mathematical Derivations and Theoretical Justifications

Derivation of the model parameter

From the data model (3.4) we see that $v_0 = E(\ln V_0)$, where

$$V_0 \sim \frac{\chi^2_{(2p_1)}/2p_1}{\chi^2_{(2p_2)}/2p_2} = \left(\frac{p_2}{p_1}\right)\frac{\chi^2_{(2p_1)}}{\chi^2_{(2p_2)}}, \text{ and}$$
$$\ln V_0 = \ln\left(\frac{p_2}{p_1}\right) + \ln\left(\chi^2_{(2p_1)}\right) - \ln\left(\chi^2_{(2p_2)}\right)$$

This implies that

$$\begin{aligned} E(\ln V_0) &= \ln\left(\frac{p_2}{p_1}\right) + E\left[\ln\left(\chi^2_{(2p_1)}\right)\right] - E\left[\ln\left(\chi^2_{(2p_2)}\right)\right] \\ &= \ln\left(\frac{p_2}{p_1}\right) + E\left[\ln\left(X_1\right)\right] - E\left[\ln\left(X_2\right)\right], \end{aligned} \tag{A.1}$$

where, $X_i = X^2_{(2p_i)} \approx \text{Gamma}\left(p_i, \frac{1}{2}\right)$; for $i = 1, 2$.

For a random variable $X \sim Gamma(\alpha, \beta)$ with α and β being the shape and the rate parameter, respectively,

$$\begin{aligned} f_X(x;\alpha,\beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} \\ &= x\exp[-\log\Gamma(\alpha) + \alpha\log(x) + \alpha\log\beta - \beta x] \\ &= x\exp[\alpha\log(x) - \beta x - \{\log\Gamma(\alpha) - \alpha\log\beta\}] \end{aligned}$$

This satisfies that the gamma distribution with parameters α and β belongs to the exponen-

tial family with natural parameters $\eta_1 = \alpha - 1$ and $\eta_2 = -\beta$. Now, in terms of these natural parameters, the gamma pdf can be written as

$$\begin{aligned} f_x(x;\alpha,\beta) &= x \exp\left[(\eta_1+1)\ln x + \eta_2 x - \{\ln\Gamma(\eta_1+1) - (\eta_1+1)\ln(-\eta)\right] \\ &= h(x)\exp\left[g(\eta_1)\,T_1(x) + g(\eta_2)\,T_2(x) - A(\eta)\right] \end{aligned}$$

which is a form of a natural exponential family distribution with natural parameters η_1 and η_2 and the corresponding natural statistic $T_1(x) = \ln(x),$ and $T_2(x) = x$. Here,

$$A(\eta_1,\eta_2) = \ln\Gamma(\eta_1+1) - (\eta_1+1)\ln(-\eta_2)$$

is called a log-partition function. One can obtain the moments of sufficient statistics simply by differentiating the function $A(\eta_1,\eta_2)$. So, the mean of the sufficient statistic for η_1 is:

$$\begin{aligned} \mathrm{E}[\ln x] &= \frac{\partial A(\eta_1,\eta_2)}{\partial\eta_1} = \frac{\partial}{\partial\eta_1}\left(\ln\Gamma(\eta_1+1) - (\eta_1+1)\ln(-\eta_2)\right) \\ &= \psi(\eta_1+1) - \ln(-\eta_2) \\ &= \psi(\alpha) - \ln\beta \end{aligned}$$

Where $\psi(.)$ is the digamma function (a derivative of log gamma). Now, for η_2 :

$$\begin{aligned} \mathrm{E}[x] &= \frac{\partial A(\eta_1,\eta_2)}{\partial\eta_2} = \frac{\partial}{\partial\eta_2}\left(\ln\Gamma(\eta_1+1) - (\eta_1+1)\ln(-\eta_2)\right) \\ &= -(\eta_1+1)\frac{1}{-\eta_2}(-1) = \frac{\eta_1+1}{-\eta_2} \\ &= \frac{\alpha}{\beta} \end{aligned}$$

Now, to compute the variance of $\log x$, we differentiate the $A(\eta_1, \eta_2)$ twice as:

$$\begin{aligned}\text{Var}(\ln x) &= \frac{\partial^2 A(\eta_1, \eta_2)}{\partial \eta_1^2} = \frac{\partial}{\partial \eta_1}\left(\psi(\eta_1 + 1) - \ln(-\eta_2)\right) \\ &= \psi^{(1)}(\eta_1 + 1) \\ &= \psi^{(1)}(\alpha), \text{ where } \psi^{(1)}(.) \text{ is the trigamma function.}\end{aligned}$$

Now, back into the equation (A.1), we have

$$\begin{aligned}v_0 = E(\ln V_0) &= \ln\left(\frac{p_2}{p_1}\right) + \psi(p_1) - \ln(1/2) - \psi(p_2) + \log(1/2) \\ &= \ln(p_2/p_1) + \psi(p_1) - \psi(p_2).\end{aligned}$$

Similarly, the variance of the data model can be derived as

$$\begin{aligned}Var(\ln V_0) &= Var[\ln(X_1)] + Var[\ln(X_2)] \\ &= \psi^{(1)}(p_1) + \psi^{(1)}(p_2).\end{aligned}$$

APPENDIX B

List of Supplementary Figures

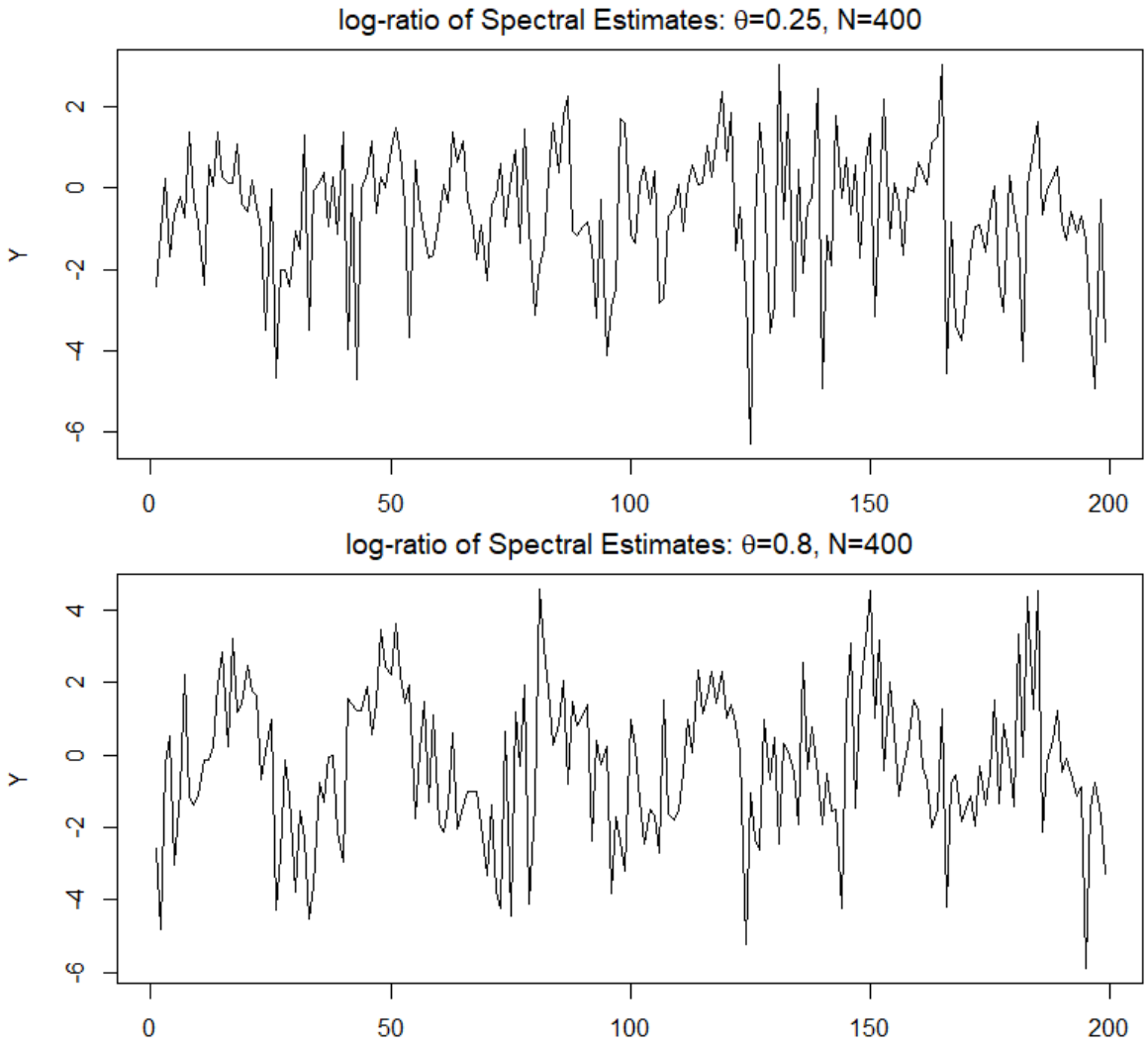

Figure B.1: Time series plot for observed data: $X_t \sim N(0,1)$ whereas, $Z_t \sim ARMA(0,1)_{12}$ with seasonal moving-average parameter parameter $\theta = 0.25$ (*top plot*) and $\theta = 0.80$ (*bottom plot*)

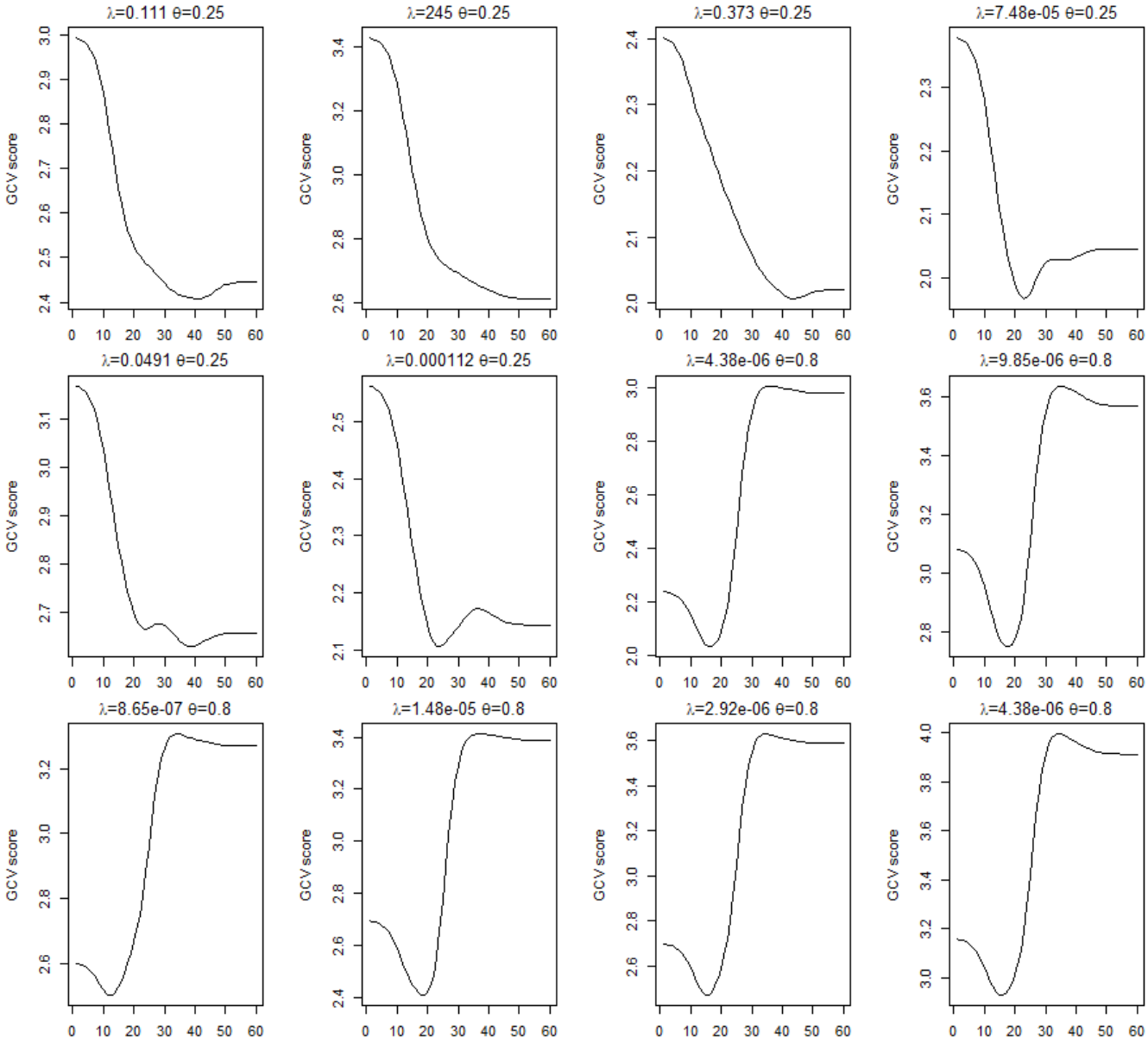

Figure B.2: Generalized Cross Validation for optimal λ from the same time series used in Figure B.1

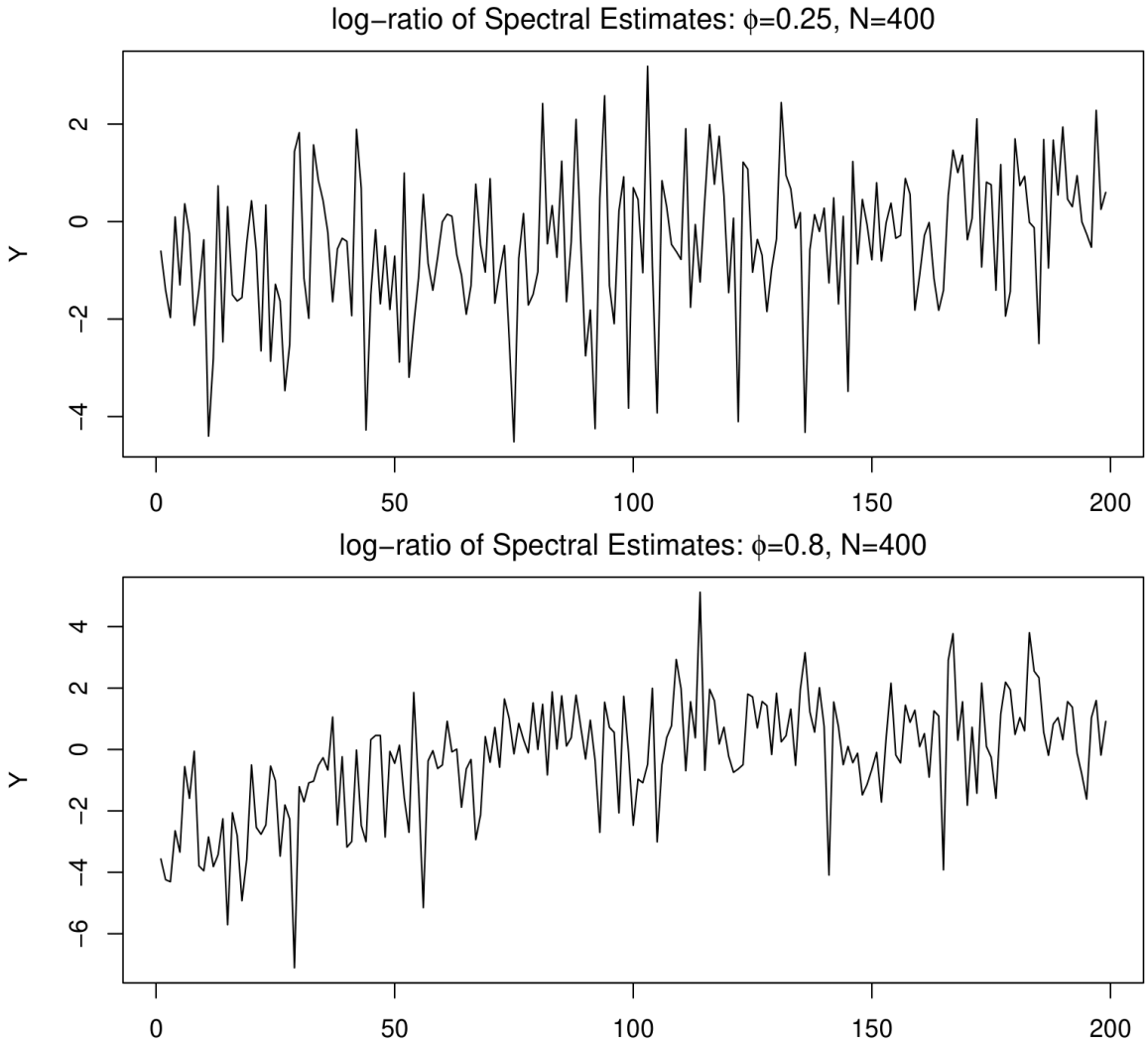

Figure B.3: Time series plot for observed data: $X_t \sim N(0,1)$ and $Z_t \sim AR(1)$ with autoregressive parameter $\phi = 0.25(topplot)$ and $\phi = 0.80(bottomplot)$.

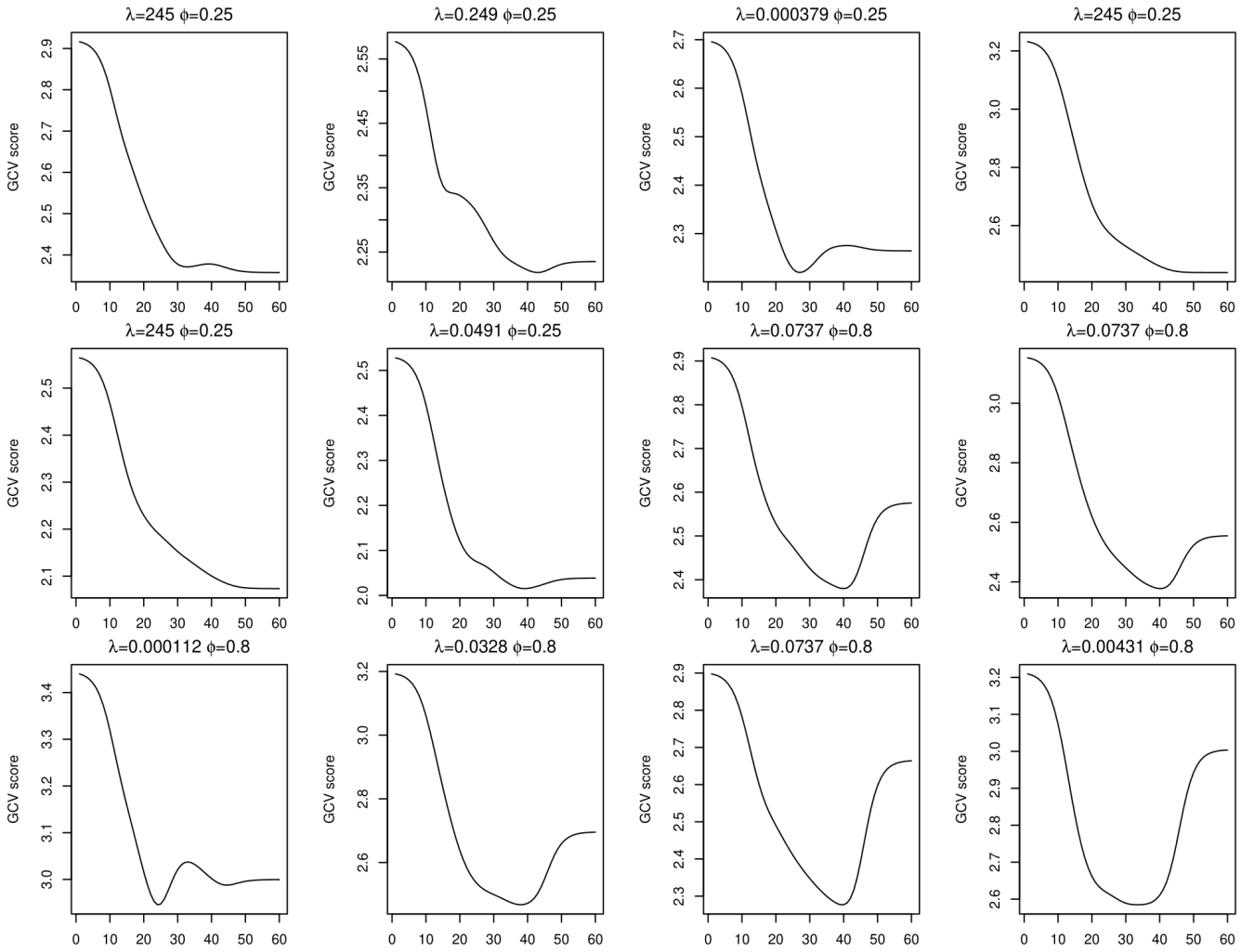

Figure B.4: Generalized Cross Validation for optimal λ from the same time series used in Figure B.3.

APPENDIX C

List of supplementary Tables

Table C.1: Empirical type I errors (in %) for an AR(1) model with $\phi = 0.5$

$p_1, p_2;\ N$	$\lambda = 1.0$	$\lambda = 0.1$	$\lambda = 0.01$	$\lambda = 0.001$	$\lambda = 0.0001$
			$\mathbf{n_1 = n_2 = 300}$		
1,1; 300	4.94	5.70	4.82	4.96	4.88
2,2; 150	5.44	5.64	5.24	4.82	5.40
3,3; 100	4.94	5.04	5.00	4.24	4.48
4,4; 75	4.72	4.52	4.56	4.98	4.72
5,5; 60	4.30	4.06	4.68	4.14	4.36
			$\mathbf{n_1 = n_2 = 1000}$		
1,1; 1000	4.72	5.46	5.24	4.92	5.36
2,2; 500	5.32	5.00	4.50	4.86	5.16
3,3; 333	5.26	4.30	4.42	5.28	4.34
4,4; 250	4.78	5.14	5.34	4.48	5.10
5,5; 200	5.00	4.58	4.84	4.38	4.48
6,6; 166	4.56	4.46	4.84	4.66	4.14
8,8; 125	5.54	5.30	4.34	4.88	4.48
10,10; 100	4.68	5.70	5.44	5.02	4.70
			$\mathbf{n_1 = 200, n_2 = 400}$		
1,2; 200	5.00	5.16	4.64	5.02	5.86
2,4; 100	4.72	4.90	4.80	4.44	4.88
3,6; 66	4.52	4.98	4.64	4.12	4.50
4,8; 50	4.24	4.68	5.20	4.28	4.92
5,10; 40	4.52	4.58	4.08	4.46	4.26
			$\mathbf{n_1 = 600, n_2 = 1000}$		
2,3; 300	4.46	5.10	5.30	5.04	5.26
3,5; 200	4.72	5.24	4.84	4.98	5.12
6,10; 100	4.90	4.54	4.52	4.90	3.84

Table C.2: Empirical type I errors (in %) for an MA(1) model with $\theta = 0.5$

$p_1, p_2;\ N$	$\lambda = 1.0$	$\lambda = 0.1$	$\lambda = 0.01$	$\lambda = 0.001$	$\lambda = 0.0001$
			$\mathbf{n_1 = n_2 = 300}$		
1,1; 300	5.14	4.84	5.72	5.52	5.42
2,2; 150	4.70	5.00	5.72	4.96	4.82
3,3; 100	4.74	4.80	4.26	4.44	4.68
4,4; 75	4.56	4.86	4.92	5.14	5.72
5,5; 60	4.94	4.96	4.84	4.00	4.42
			$\mathbf{n_1 = n_2 = 1000}$		
1,1; 1000	5.28	5.24	5.30	4.84	4.96
2,2; 500	4.76	4.34	4.80	5.06	4.44
3,3; 333	5.30	5.30	5.08	5.14	5.34
4,4; 250	4.82	5.32	5.24	5.20	4.66
5,5; 200	4.34	4.86	4.98	4.96	4.28
6,6; 166	5.22	5.80	4.56	4.90	5.28
8,8; 125	4.58	4.76	4.80	4.68	4.30
10,10; 100	5.04	4.30	4.40	4.20	4.10
			$\mathbf{n_1 = 200, n_2 = 400}$		
1,2; 200	5.42	4.88	5.24	5.10	5.76
2,4; 100	5.54	4.80	5.38	4.84	5.08
3,6; 66	4.40	4.90	4.14	3.82	4.72
4,8; 50	5.14	4.68	4.12	4.68	4.90
5,10; 40	4.12	4.18	4.24	4.76	4.94
			$\mathbf{n_1 = 600, n_2 = 1000}$		
2,3; 300	5.22	5.32	4.90	4.36	4.76
3,5; 200	5.16	4.94	4.72	4.72	4.86
6,10; 100	5.02	4.92	4.64	5.02	4.84

Table C.3: Empirical type I errors (in %) for a seasoanl AR(1) model with frequency 12 and the autoregressive parameter, $\phi = 0.5$.

$p_1, p_2;\ N$	$\lambda = 1.0$	$\lambda = 0.1$	$\lambda = 0.01$	$\lambda = 0.001$	$\lambda = 0.0001$
			$\mathbf{n_1 = n_2 = 300}$		
1,1; 300	5.61	5.25	5.96	5.99	5.79
2,2; 150	4.88	4.96	4.89	5.05	5.93
3,3; 100	5.68	5.28	5.61	6.15	5.56
4,4; 75	6.16	6.38	6.30	6.54	6.58
5,5; 60	5.66	5.00	5.91	6.08	5.99
			$\mathbf{n_1 = n_2 = 1000}$		
1,1; 1000	5.29	5.57	4.49	5.12	5.53
2,2; 500	4.88	5.22	5.40	4.90	5.32
3,3; 333	5.13	4.62	4.97	4.68	5.03
4,4; 250	5.03	5.06	4.78	5.72	5.17
5,5; 200	4.89	5.32	5.09	4.66	4.91
6,6; 166	5.11	5.42	5.36	5.13	5.04
8,8; 125	5.19	5.32	5.27	5.30	5.52
10,10; 100	5.37	4.96	4.93	6.04	5.84
			$\mathbf{n_1 = 200, n_2 = 400}$		
1,2; 200	5.61	5.45	5.66	5.88	6.01
2,4; 100	5.53	5.78	5.60	5.95	6.29
3,6; 66	5.58	5.55	5.77	6.53	6.16
4,8; 50	6.32	6.50	6.59	6.63	7.51
5,10; 40	6.84	6.51	7.04	7.77	9.16
			$\mathbf{n_1 = 600, n_2 = 1000}$		
1,1; 600	5.24	5.30	5.31	5.45	4.68
1,2; 500	5.70	5.70	5.10	6.31	6.05
2,3; 300	4.99	5.24	5.14	4.98	5.30
2,4; 250	5.26	5.21	5.36	5.30	5.74
3,5; 200	4.67	5.16	5.41	6.04	5.28
4,6; 150	4.63	5.21	5.45	5.51	5.30
4,8; 125	5.43	5.20	5.20	5.76	5.91
6,10; 100	5.37	4.82	5.27	5.04	5.69

Table C.4: Empirical type I errors (in %) for an AR(3) model with autoregressive parameters, $\phi_1 = 0.65,\ \phi_2 = -0.85,\ and\ \phi_3 = 0.3$.

$p_1, p_2;\ N$	$\lambda = 1.0$	$\lambda = 0.1$	$\lambda = 0.01$	$\lambda = 0.001$	$\lambda = 0.0001$
			$\mathbf{n_1 = n_2 = 300}$		
1,1; 300	6.10	6.01	6.12	5.93	5.46
2,2; 150	5.93	6.00	6.08	5.15	5.27
3,3; 100	5.64	6.18	5.84	5.73	5.29
4,4; 75	5.75	6.23	6.85	5.68	5.73
5,5; 60	6.15	5.71	5.29	5.67	4.61
			$\mathbf{n_1 = n_2 = 1000}$		
1,1; 1000	5.49	5.34	5.08	5.05	5.28
2,2; 500	4.61	5.52	5.02	5.38	5.40
3,3; 333	4.75	5.13	5.14	5.12	5.51
4,4; 250	5.63	5.10	5.28	4.72	4.86
5,5; 200	5.77	4.58	5.06	4.66	5.16
6,6; 166	4.78	5.68	5.40	5.17	4.79
8,8; 125	5.84	4.68	5.75	5.05	4.78
10,10; 100	5.23	4.82	5.03	4.90	4.69
			$\mathbf{n_1 = 200, n_2 = 400}$		
1,2; 200	6.56	6.40	6.45	5.71	6.05
2,4; 100	6.39	6.08	6.56	5.65	6.16
3,6; 66	6.23	6.79	5.98	5.77	5.11
4,8; 50	6.48	6.44	6.16	5.54	5.58
5,10; 40	5.76	5.73	5.64	5.14	5.58
			$\mathbf{n_1 = 600, n_2 = 1000}$		
1,1; 600	5.69	5.46	5.47	5.62	6.14
1,2; 500	5.73	5.19	5.60	5.54	5.21
2,3; 300	5.51	5.51	4.89	5.77	5.61
2,4; 250	5.43	5.32	5.13	5.33	4.97
3,5; 200	5.37	4.82	5.01	5.33	4.85
4,6; 150	5.36	5.28	5.38	4.71	5.03
4,8; 125	5.49	6.16	6.00	5.27	5.01
6,10; 100	5.05	5.43	5.73	5.17	4.81

Table C.5: Empirical Powers of our proposed test for $N(0,1)$ vs. $ARMA(0,1)_{12}$ Processes.

θ	n	N=n/2			N=n/4		
		$\lambda = 10^{-5}$	$\lambda = 10^{-6}$	$\lambda = 10^{-7}$	$\lambda = 10^{-5}$	$\lambda = 10^{-6}$	$\lambda = 10^{-7}$
0.10	256	0.073	0.066	0.069	0.047	0.058	0.064
0.10	512	0.089	0.089	0.088	0.077	0.092	0.092
0.10	1024	0.154	0.148	0.122	0.148	0.150	0.132
0.258	256	0.155	0.197	0.189	0.095	0.198	0.212
0.258	512	0.415	0.418	0.352	0.388	0.500	0.493
0.258	1024	0.839	0.800	0.705	0.883	0.886	0.814
0.578	256	0.649	0.784	0.769	0.361	0.745	0.769
0.578	512	0.994	0.998	0.989	0.992	0.997	0.998
0.578	1024	1.000	1.000	1.000	1.000	1.000	1.000

Table C.6: Empirical Powers for N(0,1) vs. AR(1) Processes in Small Samples.

ϕ	Previous Tests				Our Proposed Test			
	$D_{5,5\%}$	$X^2_{64,1}$	$X^2_{32,1}$	$X^2_{64,2}$	$T_{64,.1}$	$T_{64,.01}$	$T_{32,.1}$	$T_{32,.01}$
	$\mathbf{n_1 = 64, n_2 = 64}$							
0	3.8	5.3	6.2	4.8	5.3	5.5	4.2	4.6
0.1	5.4	6.2	6.5	6.1	6.9	7.3	5.8	6.0
0.2	11.5	8.6	8.2	8.6	10.9	11.2	11.1	11.1
0.4	29.2	21.9	17.2	24.1	31.4	29.1	35.6	34.1
0.6	57.3	47.7	37.3	55.4	62.7	61.3	70.9	70.9
	$\mathbf{n_1 = 64, n_2 = 128}$							
0	5.2	5.2	6.1	5.0	5.0	5.9	4.9	4.3
0.1	6.9	6.3	7.0	6.2	6.9	6.7	6.9	6.7
0.2	14.7	10.1	8.8	10.8	13.6	13.5	14.4	13.0
0.4	44.6	29.0	21.8	33.6	43.6	40.7	49.3	45.5
0.6	75.7	65.3	53.6	72.6	81.0	80.1	88.0	87.6