DISSERTATION

CONSTRAINED SPLINE REGRESSION AND HYPOTHESIS TESTS IN THE PRESENCE OF CORRELATION

Submitted by

Huan Wang

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2013

Doctoral Committee:

Advisor: Mary C. Meyer
Co-advisor: Jean D. Opsomer

F. Jay Breidt
Robin M. Reich

ABSTRACT

CONSTRAINED SPLINE REGRESSION AND HYPOTHESIS TESTS
IN THE PRESENCE OF CORRELATION

Extracting the trend from the pattern of observations is always difficult, especially when the trend is obscured by correlated errors. Often, prior knowledge of the trend does not include a parametric family, and instead the valid assumption are vague, such as "smooth" or "monotone increasing," Incorrectly specifying the trend as some simple parametric form can lead to overestimation of the correlation, and conversely, misspecifying or ignoring the correlation leads to erroneous inference for the trend. In this dissertation, we explore spline regression with shape constraints, such as monotonicity or convexity, for estimation and inference in the presence of stationary AR(p) errors. Standard criteria for selection of penalty parameter, such as Akaike information criterion (AIC), cross-validation and generalized cross-validation, have been shown to behave badly when the errors are correlated and in the absence of shape constraints. In this dissertation, correlation structure and penalty parameter are selected simultaneously using a correlation-adjusted AIC. The asymptotic properties of unpenalized spline regression in the presence of correlation are investigated. It is proved that even if the estimation of the correlation is inconsistent, the corresponding projection estimation of the regression function can still be consistent and have the optimal asymptotic rate, under appropriate conditions. The constrained spline fit attains the convergence rate of unconstrained spline fit in the presence of AR(p) errors. Simulation results show that the constrained estimator typically behaves better than the unconstrained version if the true trend satisfies the constraints.

Traditional statistical tests for the significance of a trend rely on restrictive assumptions on the functional form of the relationship, e.g. linearity. In this dissertation, we develop testing procedures that incorporate shape restrictions on the trend and can account for correlated errors. These tests can be used in checking whether the trend is constant versus monotone, linear versus convex/concave and any combinations such as, constant versus increase and convex. The proposed likelihood ratio test statistics have an exact null distribution if the covariance matrix of errors is known. Theorems are developed for the asymptotic distributions of test statistics if the covariance matrix is unknown but the test statistics use a consistent estimator of correlation into their estimation. The comparisons of the proposed test with the F-test with the unconstrained alternative fit and the one-sided t-test with simple regression alternative fit are conducted through intensive simulations. Both test size and power of the proposed test are favorable, smaller test size and greater power in general, comparing to the F-test and t-test.

ACKNOWLEDGEMENTS

Except for my hometown, I have never lived any other place a so long time. Fort Collins, a five-year harbor, my second hometown, and the most beautiful mountainside college town I have ever seen, it is hard for me to say goodbye to you.

I would like to thank one of my advisors, Dr. Mary C. Meyer, who is not only my academic advisor, but also my mentor and role model on how to be a woman statistician. Dr. Meyer is an earnest researcher, knowledgeable scholar, patient teacher and elegant lady. Research needs long-term effect. Whenever I got stuck and struggled aimlessly, your intelligence and expertise guided me out. Without your endless patience and encouragement, I can never grow up into a woman Ph.D. statistician. I would also like to thank the other advisor of me, Dr. Jean D. Opsomer, a smart, actually very smart, and fun mentor, a great and popular department chair and also a Tae Kwon Do master. It is so lucky for me to be advised and educated by a knowledgeable and affectionate couple as Dr. Meyer and you.

I am also grateful to have Dr. F. Jay Breidt and Dr. Robin Reich as my Ph.D. committee members. Dr. Breidt is a brilliant statistician, an enthusiastic rock climber and a fun guy always looks like in his twenties.

Living five years far from my hometown made me feel lonely sometimes, but I am fortunate to meet a lot of friends here, particularly, Yajing Liu and Haonan Chen. Your companion and taking care of me made me feel like being at home. We are rather like family than friends.

I would also like to thank the department for the financial support during my Master and Ph.D. studies in the past five years. Moreover, I would like to thank faculties for the statistical educations. I have greatly enjoyed my study in our department with my fellow students, particularly the girls fun time with Lulu Wang, Xiyue Liao, Jiwen Wu, Ran Fu and Yuan Wang. I am also lucky to study with those smart and fun classmates in 2008 class, Bruce Bugbee, Grant Weller, Wade Herndon, Brett Hunter and Stacy Edmondson.

Finally, I want to thank my parents and my sister. They are always encouraging me to devote my passion and hard work to my interest. They try their best to help me realize my every little or big dream. As a little kid in my family, I always feel living happily under extra care and love. Because of the love you gave me, I know how to give my love to others.

DEDICATION

To my family and my love

You encourage me to live a life that I want, companion me when I feel lack of confidence of my future, fond me and tolerate all my bad temper and infirmity.

TABLE OF CONTENTS

1 Introduction **1**

1.1 Constrained Spline Regression 1

1.2 Shape Restricted Hypothesis Tests 3

2 Constrained Spline Regression in the Presence of Correlation **5**

2.1 Literature Review 5

2.2 Model Setup and Proposed Estimator 8

2.3 Large Sample Theory 13

2.3.1 Rates of Convergence in the Presence of Correlation 13

2.3.2 Fixed-Knot Unpenalized Unconstrained Spline Regression 21

2.3.3 Fixed Knot Constrained Unpenalized Spline Regression 22

2.4 Simulation 25

2.4.1 Data Scenarios 25

2.4.2 The selection of order p and penalty parameter by AIC 26

2.4.3 Three Performance Measures 28

2.5 Global Temperature Data 32

2.6 Conclusion 33

3 Shape Restricted Hypothesis Tests in the Presence of Correlation **36**

3.1 Literature Review 36

3.2 The Proposed Tests 39

3.2.1 Review of Independent Normal Errors Case 40

3.2.2 Stationary AR(p) Errors: Known Correlation 45

3.2.3 Test Statistics with Unknown Correlation 47

3.3 Approximate Null Distributions of Test Statistics with Unknown Correlation 49

3.4 Simulation 58

3.4.1 Test Size 59

3.4.2 Power Study 60

3.5 Real Data 61

3.5.1 Argentina Rainfall Data 61

3.5.2 Price of Liquefied U.S. Natural Gas Exports Data 67

4 Conclusion and Future Work 71

4.1 Conclusion 71

4.2 Future Work 72

4.2.1 Tests of the Monotonicity and Convexity in the presence of correlation 72

References 74

LIST OF TABLES

2.1 Performance of Correlation-adjusted AIC in Selection of p 27

2.2 Performance of Three Measures for Data with AR(1) Errors 29

2.3 Performance of Three Measures for Data with AR(p) Errors 30

2.4 Improvement of Estimation When incorporating into Correlation 31

3.1 Comparisons of Simulated Test Size 60

3.2 Comparisons of Simulated Power 66

LIST OF FIGURES

2.1 Fit Global Temperature Data 1: Constrained Spline Fit, Unconstrained Spline Fit and Linear Fit . . . 34

2.2 Fit Global Temperature Data 2: Constrained Spline Fit, Unconstrained Spline Fit and Linear Fit . . . 34

3.1 Monotonicity Tests of Argentina Rainfall Data . . . 62

3.2 Sample autocorrelation function of residuals from constrained spline fit of Argentina rainfall data . . . 63

3.3 Sample autocorrelation function of residuals from unconstrained spline fit of Argentina rainfall data . . . 63

3.4 Sample autocorrelation function of residuals from linear fit of Argentina rainfall data . . . 64

3.5 Plot of standard residuals from constrained spline fit of Argentina rainfall data 64

3.6 Plot of standard residuals from unconstrained spline fit of Argentina rainfall data . . . 65

3.7 Plot of standard residuals from linear fit of Argentina rainfall data . . . 65

3.8 Convexity Tests of Price of Liquefied U.S. Natural Gas Exports Data . . . 67

3.9 Sample autocorrelation function of residuals from constrained spline fit of price of liquefied U.S. natural gas exports data . . . 68

3.10 Sample autocorrelation function of residuals from unconstrained spline fit of price of liquefied U.S. natural gas exports data . . . 68

3.11 Plot of standard residuals from constrained spline fit of price of liquefied U.S. natural gas exports data . . . 69

3.12 Plot of standard residuals from unconstrained spline fit of price of liquefied U.S. natural gas exports data . . . 69

CHAPTER 1

INTRODUCTION

1.1 Constrained Spline Regression

Regression splines are a popular nonparametric function estimation method, but they are known to be sensitive to knot number and placement. However, if there is more information about the shape of the regression function, like monotonicity or convexity, the shape-restricted splines are robust to knot choices. In this dissertation, we use shape restricted inference for both regression spline estimation and hypothesis tests of the trend. The shape restricted estimation can be transformed into a cone projection problem. Instead of projections onto the linear space spanned by the columns of the design matrix in ordinary least-squares estimation for linear model, constrained estimation involves the projection onto a convex cone determined by shape restriction. Different types of shape restriction, such as monotonicity and convexity, are written into different constraint matrix, based on which the constrained cone is constructed. But the theories and algorithms for different types of shape restriction are similar. Comparing to the unconstrained estimation, we need cone projection algorithm to identify which face of the cone projection falls on, therefore the projection matrices are random.

Extracting the trend from the pattern of observations is always difficult, especially when the trend is obscured by correlated errors. Often, prior knowledge of the trend does not include a parametric family, and instead the valid assumption being vague, such as "smooth" or "monotone increasing." Incorrectly specifying the trend as some simple parametric form can lead to overestimation of the correlation. In this dissertation, we derive the shape restricted nonparametric methods for data with autocorrelated errors. A Cochrane-Orcutt type iteration procedure is developed for estimation of both trend and correlation.

Nonparametric regression estimators are often sensitive to the presence of correlation in the errors. The selection of smoothing parameter for nonparametric methods is a challenging problem, especially when the data are correlated. Most of the popular data-driven methods, such as generalized cross-validation (GCV) and Akaike information criterion (AIC), have been developed under the assumption of independent observations and will be broken down if the correlation is not accounted for. They tend to select a small tuning parameter and the fits become progressively more under-smoothed as the correlation increases. We generalized the AIC to correlated data and develop a correlation-adjusted AIC to select smoothing parameter and the order of autocorrelation, simultaneously.

In order to derive the consistency and convergence rate for the constrained spline regression estimator, several general theorems on the convergence rate of projection estimation were developed for unconstrained and unpenalized regression in the presence of stationary AR(p) errors at first. Then by using those results, the theorems on the convergence rate for constrained unpenalized spline regression were presented. Let $\boldsymbol{\mu}$ be the true mean of data. We developed the consistency and convergence rate for three types of estimators: ordinary least-squares estimator $\hat{\boldsymbol{\mu}}_{\mathbf{I}}$, weighted least-squares estimator $\hat{\boldsymbol{\mu}}_{\mathbf{R}}$ when correlation $\mathbf{R}$ is known and weighted least-squares estimator $\hat{\boldsymbol{\mu}}_{\mathbf{S}}$ when correlation is unknown and there is an $n \times n$ symmetric positive-definite matrix $\mathbf{S}$, which can be used as a substitute of $\mathbf{R}$. For unconstrained spline regression, we proved that if $\mathbf{S}$ satisfies several conditions, the three types of estimators have the same convergence rate. We also proved that the constrained spline estimators attain the convergence rate of the corresponding unconstrained spline estimators. Through the simulations, we also demonstrated that by incorporating into the correlation, estimation of the trend is substantially improved for moderate-sized samples for the Cochrane-Orcutt type iterated estimator.

1.2 Shape Restricted Hypothesis Tests

Researchers often want to check whether the shape of the trend f has some sort of pattern, such as monotonicity or convexity. A test for trend in time series data uses constant for null hypothesis; an alternative hypothesis must be chosen as a way to provide a valid alternate scenario against which to perform a test. Usually, a linear regression model is chosen for alternative. If a linear relation between $\mathbf{y}$ and $\mathbf{x}$ is reasonable, we can fit the data with simple linear regression and test whether the coefficient is positive or not by the one-sided t-test. But, often prior information of the relation between $\mathbf{y}$ and $\mathbf{x}$ does not include a parametric form to model with, therefore we need to switch to nonparametric approach. Because of the flexibility of the nonparametric approach, the fit rarely satisfies the shape constraint. In this dissertation, we developed the tests with shape-restricted inference for correlated data. These tests can be used in checking whether the trend is constant versus monotone, linear versus convex/concave and any combinations such as, constant versus increase and convex.

The likelihood test statistics have exact distributions, a mixture of χ^2 distributions for known variance σ^2 and a mixture of *Beta* distributions for unknown σ^2, if the correlation matrix is known. The mixing distribution can be determined by numerical computation under $\mathbf{H}_0$ as precisely as desired. Theorems were developed for the asymptotic distributions of test statistics if the covariance matrix is unknown but the test statistics use a consistent estimator of correlation into their estimation. The comparisons of the proposed test with the F-test with the unconstrained alternative fit and the one-sided t-test with simple regression alternative fit were conducted through intensive simulations. Both test size and power of the proposed test are favorable.

The rest of the dissertation is organized as follows. In Chapter 2, we propose a constrained spline estimator for data with stationary AR(p) errors with unknown order and unknown correlation parameters. Because of the popularity of the penalized splines, we include a penalty into our estimator. A new correlation-adjusted AIC is given for the selection of the penalty parameter and autoregressive parameters simultaneously. We prove the asymptotic properties of the constrained unpenalized spline estimator in the presence of stationary AR(p) errors. The proposed estimator and the method to select the order of correlation and the penalty parameter are presented in Section 2.2. In Section 2.3, the convergence rate of the estimator in the presence of correlation is derived in a general setting of both parametric and nonparametric regression, and also the specific application of constrained spline regression. The comparison of the convergence rate of the constrained spline regression and the unconstrained spline regression is also discussed in Section 2.3. Simulations evaluating the selection method of the order of AR(p) process and comparing the proposed method with the other two alternatives are conducted in Section 2.4. In Section 2.5, we analyze global temperature data with the proposed method and compare with other methods. In Chapter 3, we propose several types of shape-restricted hypothesis tests using cone projection theories and algorithms. The observations are correlated and the correlation is restricted to a stationary AR(p) process with unknown p and unknown correlation. The alternative fit of the proposed tests is constrained nonparametric regression spline estimator. The proposed tests making use of theories of cone projection for both independent data and correlated data are derived in Section 3.2. In Section 3.3, the approximate distributions for the test statistics with unknown p and unknown correlation are investigated. In Section 3.4, simulations of comparing test size and power of the proposed test with the F-test with unconstrained alternative fit and one-sided t-test with simple regression alternative fit are conducted. In Section 3.5, we apply the proposed tests to test the monotonicity of Argentina rainfall data and the convexity of price of liquefied U.S. natural gas exports data. Conclusions and future work are included in Chapter 4.

CHAPTER 2

CONSTRAINED SPLINE REGRESSION IN THE PRESENCE OF CORRELATION

2.1 Literature Review

Regression splines are a popular nonparametric function estimator method, but they are known to be sensitive to knot number and placement. However, if there is more information about the shape of the regression function, like monotonicity or convexity, the shape-restricted splines are robust to knot choices.

For shape-restricted regression, Brunk (1955, 1958) proposed unsmoothed monotone regression estimation and studied its asymptotic behavior. See Robertson et al. (1988) for details about estimation inference. Ramsay (1998) proposed a device to estimate a smooth strictly monotone function of arbitrary flexibility. Tantiyaswasdikul and Woodroofe (1994) proposed the monotone smoothing splines with penalty on the integrated first derivative. Mammen and Thomas-Agnan (1999) showed that the monotone smoothing splines have an optimal $n^{-p/(2p+1)}$ convergence rate, where $p = \max\{k, r\}$, k is the order of spline and r is the order of derivative. Hall and Huang (2001) developed a biased-bootstrap method for monotonizing general linear, kernel-typed estimators. Meyer (2008) proposed an algorithm for the cubic monotone case, and also extended the method to convex constraints and variants such as increasing-concave.

Penalized splines, introduced by Eilers and Marx (1996), use a large number of knots compared to regression splines, but fewer than in smoothing splines, and hence are less computationally cumbersome. The penalization shrinks the coefficients towards zero, constraining their influence and resulting in a less variable fit than regression splines. Penalized splines are increasingly popular in handling a wide range of nonparametric and semiparametric problems. Ruppert et al. (2003) provided details of this method. Hall and Opsomer (2005) used a white-noise process representation of the penalized spline estimator to obtain the mean squared error and consistency of the estimator. This representation treats the data as being generated from a continuously varying set of basis functions, subject to a penalty, so the complicating effect of the finite set of basis functions is removed. This enabled them to explore the role of the penalty and its relationship with the sample size in ways that are not possible in the discrete-data, finite-basis setting. Li and Ruppert (2008) showed that penalized splines behave similarly to Nadaraya-Watson kernel estimators with equivalent kernels. By this equivalent kernel representation, they developed an asymptotic theory of penalized splines for the cases of piecewise-constant or linear splines, with a first- or second-order difference penalty. Claeskens et al. (2009) developed a general theory of the asymptotic properties of penalized spline estimators for any order of spline and general penalty. They demonstrated that the theoretical properties of penalized spline estimators are either similar to those of regression splines or to those of smoothing splines, with a clear breakpoint distinguishing the cases. Kauermann et al. (2009) used a Bayesian viewpoint by imposing a priori distribution on all parameters and coefficients, arguing that with the postulated rate at which the spine basis dimension increases with the sample size the posterior distribution of the spline coefficients is approximately normal.

Nonparametric regression estimators are often sensitive to the presence of correlation in the errors. Most of the data-driven smoothing parameter selection methods, such as cross-validation, general cross-validation and AIC, will break down if the correlation is ignored. Diggle and Hutchinson (1989) presented an extension of generalized cross-validation which accommodates a known correlation matrix for the errors. Altman (1990) suggested two methods, a direct method and an indirect method, for correcting the selection criteria when the correlation function is known. Hart (1991) used a risk estimation procedure to select the bandwidth in the kernel regression with correlated errors. Hart (1994) proposed time series cross-validation to estimate the bandwidth and gave a time series model for the errors simultaneously. Wang (1998) extended the generalized maximum likelihood, generalized cross-validation and unbiased risk methods to estimate the smoothing parameters and the correlation parameters simultaneously, when the correlation matrix is assumed to depend on a parsimonious set of parameters. Opsomer et al. (2001) gave a general review of the literature in kernel regression, smoothing splines and wavelet regression under correlation. Hall and Keilegom (2003) used difference-based methods to construct estimators of error variance and autoregressive parameters in nonparametric regression with time series errors. They proved that the difference-based estimators can be used to produce a simplified version of time series cross-validation. Francisco-Fernandez and Opsomer (2005) proposed to adjust the generalized cross-validation (GCV) criterion for the spatial correlation and showed that it leads to improved smoothing parameter selection results even when the covariance model is misspecified. Kim et al. (2009) investigated a bandwidth selector based on the use of a bimodal kernel for nonparametric regression with fixed design and proved that the proposed selector is quite effective when the errors are severely correlated.

In this article, we propose a constrained spline estimator for data with stationary AR(p) errors with unknown order and unknown correlation parameters. Because of the popularity of the penalized splines, we include a penalty into our estimator. A new correlation-adjusted AIC is given for the selection of the penalty parameter and autoregressive parameters simultaneously. We prove the asymptotic properties of the constrained unpenalized spline estimator in the presence of stationary AR(p) errors. The asymptotic properties of constrained penalized splines regression, due to the complexity of proofs for penalized splines, are still being studied and not included in this paper.

The proposed estimator and the method to select the order of correlation and the penalty parameter are presented in Section 2. In Section 3, the convergence rate of the estimator in the presence of correlation is derived in a general setting of both parametric and nonparametric regression, and also the specific application of constrained spline regression. The comparison of the convergence rate of the constrained spline regression and the unconstrained spline regression is also discussed in Section 3. Simulations evaluating the selection method of the order of AR(p) process and comparing the proposed method with the other two alternatives are conducted in Section 4. In Section 5, we analyze the global temperature data with the proposed method and compare with other methods.

2.2 Model Setup and Proposed Estimator

Assume that the observed data $\{(x_i, y_i)\}$, for $1 \le i \le n$, are generated by the model

$$y_i = f(x_i) + \sigma\varepsilon_i,$$

where f is a smooth function. Suppose that $x_i \in [0, 1]$ and equally spaced. The errors $\varepsilon_1, \cdots, \varepsilon_n$ come from a segment of a mean zero autoregressive process with order p, i.e. an $AR(p)$ process. Specifically, for some integer $p \ge 1$,

$$\varepsilon_i = \sum_{j=1}^{p} \theta_j \varepsilon_{i-j} + e_i,$$

where e_i are independent standard normal random variables.

The function f is approximated by a linear combination of spline basis functions. Given a set of knots $0 = t_1 < \cdots < t_k = 1$, a set of $m = k + d - 1$ basis functions $b_1(x), \cdots, b_m(x)$ are defined, where $d = 2$ for quadratic splines and $d = 3$ for cubic splines. The standard B-spline basis is used in this article, but another basis spanning the same space can be used instead. Let $\mathbf{b}_1, \cdots, \mathbf{b}_m$ be basis vectors, where $b_{ij} = b_j(x_i)$, so that the basis functions span the space of smooth piecewise polynomial regression functions with the given knots, and the basis vectors span an m-dimensional subspace of $\mathbb{R}^n$.

For the independent-error case, the penalized sum of squares of Eilers and Marx (1996) is:

$$\sum_{i=1}^{n}[y_i - \sum_{j=1}^{m}\alpha_j b_j(x_i)]^2 + \lambda \sum_{j=q+1}^{m}(\Delta^q \alpha_j)^2,$$

where $\Delta^1\alpha_j = \alpha_j - \alpha_{j-1}$ and $\Delta^q\alpha_j = \Delta^{q-1}\Delta\alpha_j$ for $q > 1$. Let $\mathbf{B}$ be the $n \times m$ matrix with the $\mathbf{b}_j$ vectors as columns, let $\mathbf{D}$ be the qth order difference matrix and let $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_m)^{'}$. The penalty parameter $\lambda \geq 0$ controls the smoothness. Minimizing the penalized sum of squares is equivalent to minimizing the vector expression:

$$\psi(\boldsymbol{\alpha}; \mathbf{y}) = \boldsymbol{\alpha}^{'}(\mathbf{B}^{'}\mathbf{B} + \lambda\mathbf{D}^{'}\mathbf{D})\boldsymbol{\alpha} - 2\mathbf{y}^{'}\mathbf{B}\boldsymbol{\alpha}.$$

For the monotone case, we use quadratic splines and define the $k \times m$ matrix $\mathbf{T}$ of the slopes at the knots by $T_{ij} = b_j^{'}(t_i)$. Then the linear combination $\sum_{j=1}^{m}\alpha_j b_j(x)$ is non-decreasing if and only if the coefficient vector is in the set

$$\mathcal{C} = \{\boldsymbol{\alpha} : \mathbf{T}\boldsymbol{\alpha} \geq \mathbf{0}\} \subseteq \mathbb{R}^m.$$

For the convex case, we use cubic splines and $T_{ij} = b_j^{''}(t_i)$; then the linear combination is convex if and only if $\mathbf{T}\boldsymbol{\alpha} \geq \mathbf{0}$.

It is straightforward to find the appropriate spline degree and a constraint matrix $\mathbf{T}$ for constraint such as increasing and concave, or sigmoidal (convex or concave) with known inflation point.

When errors are correlated, let $cor(\boldsymbol{\varepsilon}) = \mathbf{R}$, and first suppose $\mathbf{R}$ is known. Let $\mathbf{R} = \mathbf{L}\mathbf{L}^{'}$ be the Cholesky decomposition, and use the weighted least-squares method to estimate the coefficients. This is equivalent to transforming $\tilde{\mathbf{y}} = \mathbf{L}^{-1}\mathbf{y}, \tilde{\mathbf{B}} = \mathbf{L}^{-1}\mathbf{B}$, $\tilde{\boldsymbol{\varepsilon}} = \mathbf{L}^{-1}\boldsymbol{\varepsilon}$, which has correlation matrix $\mathbf{I}$. The weighted least-squares criterion is

$$\psi(\boldsymbol{\alpha}; \tilde{\mathbf{y}}) = \boldsymbol{\alpha}^{'}(\tilde{\mathbf{B}}^{'}\tilde{\mathbf{B}} + \lambda\mathbf{D}^{'}\mathbf{D})\boldsymbol{\alpha} - 2\tilde{\mathbf{y}}^{'}\tilde{\mathbf{B}}\boldsymbol{\alpha},$$

where $\boldsymbol{\alpha}$ is again restricted to $\mathcal{C}$.

Let $\tilde{\mathbf{L}}\tilde{\mathbf{L}}^{'} = (\tilde{\mathbf{B}}^{'}\tilde{\mathbf{B}} + \lambda\mathbf{D}^{'}\mathbf{D})$, then $\boldsymbol{\phi} = \tilde{\mathbf{L}}^{'}\boldsymbol{\alpha}, \mathbf{z} = \tilde{\mathbf{L}}^{-1}\tilde{\mathbf{B}}^{'}\tilde{\mathbf{y}}$, then

$$\psi(\boldsymbol{\alpha}; \tilde{\mathbf{y}}) = \psi(\boldsymbol{\phi}; \mathbf{z}) = \parallel \boldsymbol{\phi} - \mathbf{z} \parallel^2,$$

where $\boldsymbol{\phi}$ is restricted to $\tilde{\mathcal{C}} = \{\boldsymbol{\phi} : \mathbf{A}\boldsymbol{\phi} \geq \mathbf{0}\} \subseteq \mathbb{R}^m$, a polyhedral cone, where the $k \times m$ $\mathbf{A} = \mathbf{T}(\tilde{\mathbf{L}}^{'})^{-1}$ is full row-rank. Referring to the setup in Meyer (2013), let $\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_{m-k}$ span the null space $\mathcal{V}$ of $\mathbf{A}$, and let $\tilde{\mathbf{A}}$ be the square, nonsingular matrix with the rows of $\mathbf{A}$ as first k rows and $\boldsymbol{\nu}$ vectors as the last rows. The first k columns of $\tilde{\mathbf{A}}^{-1}$ are the edges $\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_k$ of the cone, therefore the cone can be written as

$$\tilde{\mathcal{C}} = \left\{ \boldsymbol{\phi} : \boldsymbol{\phi} = \boldsymbol{\nu} + \sum_{j=1}^{k} \beta_j \boldsymbol{\delta}_j, \quad \boldsymbol{\nu} \in \mathcal{V}, \quad \beta_j \geq 0, j = 1, \cdots, k \right\}.$$

The minimizer $\hat{\boldsymbol{\phi}}$ is the projection of $\mathbf{z}$ onto the cone $\tilde{\mathcal{C}}$ and lands on a face of the cone. The 2^k faces, which partition $\tilde{\mathcal{C}}$, are indexed by the collection of sets $J \subseteq \{1, \ldots, m\}$, and are defined by

$$\mathcal{F}_J = \left\{ \boldsymbol{\phi} : \boldsymbol{\phi} = \boldsymbol{\nu} + \sum_{j \in J} \beta_j \boldsymbol{\delta}_j, \quad \boldsymbol{\nu} \in \mathcal{V}, \quad \beta_j > 0, j \in J \right\}.$$

The interior of the cone is a face with $J = \{1, \ldots, m\}$, and the origin is the face with $J = \emptyset$. We use the hinge algorithm from Meyer (2013) to determine the face $\mathcal{F}_J$ on which the projection falls, so that the estimate coincides with the ordinary least-squares projection onto the linear space spanned by the edges of the chosen face. Let Δ_J be the matrix whose columns are those edges indexed by J, where $J \subseteq \{1, \ldots, m\}$. The projection is

$\hat{\boldsymbol{\phi}} = \Delta_J(\Delta_J'\Delta_J)^{-1}\Delta_J'\mathbf{z}$, and the estimated coefficient vector is

$$\hat{\boldsymbol{\alpha}}_c = (\tilde{\mathbf{L}}')^{-1}\hat{\boldsymbol{\phi}} = (\tilde{\mathbf{L}}')^{-1}\Delta_J(\Delta_J'\Delta_J)^{-1}\Delta_J'\tilde{\mathbf{L}}^{-1}\tilde{\mathbf{B}}'\tilde{\mathbf{y}}.$$

For $\boldsymbol{\mu} \in \mathbb{R}^n$, where $\mu_i = f(x_i)$, the constrained estimated mean with the known $\mathbf{R}$ is $\hat{\boldsymbol{\mu}}_{\mathbf{R}}^c = \mathbf{B}\hat{\boldsymbol{\alpha}}_c$. The matrix

$$\mathbf{P}_{\mathbf{R}}^c = \mathbf{B}(\tilde{\mathbf{L}}')^{-1}\Delta_J(\Delta_J'\Delta_J)^{-1}\Delta_J'\tilde{\mathbf{L}}^{-1}\mathbf{B}'\mathbf{R}^{-1},$$

such that $\hat{\boldsymbol{\mu}}_{\mathbf{R}}^c = \mathbf{P}_{\mathbf{R}}^c\mathbf{y}$, is used to calculate effective degrees of freedom, i.e. $edf = tr(\mathbf{P}_{\mathbf{R}}^c)$.

If $J = \{1, \ldots, m\}$, that is, all edges are used, then $\Delta_J(\Delta_J'\Delta_J)^{-1}\Delta_J' = \mathbf{I}$, and the unconstrained spline satisfies the constraints and is identical to the constrained fit. The unconstrained estimated coefficient vector is

$$\hat{\boldsymbol{\alpha}}_u = (\tilde{\mathbf{L}}')^{-1}(\tilde{\mathbf{L}})^{-1}\tilde{\mathbf{B}}'\tilde{\mathbf{y}} = (\mathbf{B}'\mathbf{R}^{-1}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D})^{-1}\mathbf{B}'\mathbf{R}^{-1}\mathbf{y}.$$

and the unconstrained estimated mean with known $\mathbf{R}$ is $\hat{\boldsymbol{\mu}}_{\mathbf{R}}^u = \mathbf{B}\hat{\boldsymbol{\alpha}}_u$. The trace of $\mathbf{P}_{\mathbf{R}}^u = \mathbf{B}(\mathbf{B}'\mathbf{R}^{-1}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D})^{-1}\mathbf{B}'\mathbf{R}^{-1}$ is the unconstrained edf.

The edf for the constrained fit is a random quantity with $m + 1$ possible values, the largest of which is that of the unconstrained version. Meyer (2012) discussed this for the independent error case.

However, typically $cor(\boldsymbol{\varepsilon}) = \mathbf{R}$ is unknown. Here, we assume AR(p) and use Cochrane-Orcutt type iterations to estimate the matrix $\mathbf{R}$. Altman (1992) introduces a similar iteration procedure for kernel regression in the presence of correlation and also discuss the behaviors of the procedure for different types of kernels.

For our method, given p and λ, the iteration procedure for either constrained or unconstrained trend estimation is

1. Pilot fit: ignoring the correlation, obtain $\hat{\boldsymbol{\mu}}_{\mathbf{I}}^c$ and residuals $\hat{\epsilon}_i = y_i - \hat{\mu}_{\mathbf{I}i}^c$.

2. Use the Yule-Walker method in Chapter 8 of Brockwell and Davis (2009) and residual vector $\hat{\boldsymbol{\epsilon}}$ to estimate coefficients $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$ and the error variance. If $\hat{\gamma}(0)$ and $\hat{\boldsymbol{\gamma}}_p = (\hat{\gamma}_1, \cdots, \hat{\gamma}_p)'$ are the estimates of correlation function values; then obtain $\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\theta}}' \hat{\boldsymbol{\gamma}}_p$;

3. Use Cholesky decompositon $\hat{\mathbf{R}} = \hat{\boldsymbol{\Sigma}}/\hat{\gamma}(0) = \mathbf{L}\mathbf{L}'$, to transform data and basis into $\tilde{\mathbf{y}} = \mathbf{L}^{-1}\mathbf{y}, \tilde{\mathbf{B}} = \mathbf{L}^{-1}\mathbf{B}$. Using data $\tilde{\mathbf{y}}$ and spline basis matrix $\tilde{\mathbf{B}}$, obtain adjusted estimators $\hat{\boldsymbol{\theta}}, \hat{\sigma}^2, \hat{\boldsymbol{\mu}}^c_{\hat{\mathbf{R}}}$.

4. Iterate (2)-(3) twice more, obtaining the final estimators $\hat{\boldsymbol{\theta}}, \hat{\sigma}^2, \hat{\boldsymbol{\mu}}^c_{\hat{\mathbf{R}}}$.

By the result obtained from the iteration procedure, we can compute the correlation-adjusted Akaike information criterion (AIC)

$$AIC = n \log(\hat{\sigma}^2) + 2(p + edf). \tag{1}$$

We use this criterion to choose p and λ simultaneously, that is, we compute fits for a grid of p and λ values and choose the pair to minimize the AIC.

Most commonly used data-driven selection methods for tuning parameter such as generalized cross-validation (GCV) and AIC, have been developed under the assumption of independent observations. When the regression is attempted in the presence of correlated errors, those automated methods will break down if the correlation is ignored. They tend to select a small tuning parameter and the fits become progressively more under-smoothed as the correlation increases. Opsomer et al. (2001) gave an overview of these problems. We will see that these problems are alleviated if the trend is constrained to be monotone or convex.

2.3 Large Sample Theory

2.3.1 Rates of Convergence in the Presence of Correlation

We derive several general theorems on the convergence rate of projection estimation for unconstrained and unpenalized regression in the presence of stationary AR(p) errors. By using those results, the theorems on the convergence rate for constrained unpenalized spline regression are presented.

Without loss of generality, assume $\sigma = 1$. We model the regression function f as being a member of some linear function space $\mathbf{H}$, which is a subspace of all square-integrable, real-valued functions on $[0, 1]$. The least-squares estimation is a projection onto a finite-dimensional approximating subspace $\mathbf{G}_n$, which will be defined explicitly in Condition 2. If $\mathbf{H}$ is finite-dimensional, then we can choose $\mathbf{G}_n = \mathbf{H}$, leading to classical linear regression. Let $\hat{\boldsymbol{\mu}}_{\mathbf{I}}$ be the ordinary least-squares estimator of $\boldsymbol{\mu}$ and $\hat{\boldsymbol{\mu}}_{\mathbf{R}}$ be the weighted least-squares estimator of $\boldsymbol{\mu}$, when $\mathbf{R}$ is known. If $\mathbf{R}$ is unknown, suppose there is an $n \times n$ symmetric positive-definite matrix $\mathbf{S}$, which can be used as an estimator of $\mathbf{R}$; let $\hat{\boldsymbol{\mu}}_{\mathbf{S}}$ be the weighted least-squares estimator with the given matrix $\mathbf{S}$.

It is well known that $\hat{\boldsymbol{\mu}}_{\mathbf{R}}$ is superior to $\hat{\boldsymbol{\mu}}_{\mathbf{I}}$ in that the variance of any linear contrast $\boldsymbol{\lambda}'\hat{\boldsymbol{\mu}}_{\mathbf{R}}$ is no larger than the variance of the corresponding linear contrast of $\boldsymbol{\lambda}'\hat{\boldsymbol{\mu}}_{\mathbf{I}}$. However, the construction of $\hat{\boldsymbol{\mu}}_{\mathbf{R}}$ requires the knowledge of $\mathbf{R}$ and generally $\mathbf{R}$ is not known. In fact, one may wish to estimate the mean function prior to investigating the covariance structure of the errors. Therefore, the properties of the ordinary least-squares estimator $\hat{\boldsymbol{\mu}}_{\mathbf{I}}$ are of interest. Furthermore, if the mean function is estimated with an arbitrary positive-definite symmetric non-random matrix $\mathbf{S}$, it is of interest to check whether this $\hat{\boldsymbol{\mu}}_{\mathbf{S}}$ can still attain the same rate of convergence under some appropriate conditions.

Huang (1998) developed a general theory on rates of convergence for independent observations in a more general setting in which the predictor variable can be random or fixed. We will extend Huang's theory to correlated observations in the case of equally spaced x_i.

For $\boldsymbol{\mu} \in \mathbb{R}^n$, define the norm as $\|\boldsymbol{\mu}\|^2 = \frac{1}{n}\langle \boldsymbol{\mu}, \boldsymbol{\mu} \rangle$, where $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^n a_i b_i$. Let $\mathbf{P}$ be the orthogonal projection matrix onto $\mathbf{G}_n$. Let $\hat{\boldsymbol{\mu}} = \mathbf{P}\mathbf{y}$ and $\tilde{\boldsymbol{\mu}} = \mathbf{P}\boldsymbol{\mu}$, which is called the best approximation in $\mathbf{G}_n$ to $\boldsymbol{\mu}$. The total error can be decomposed as

$$\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} = (\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}) + (\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}).$$

We refer $\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}$ as the estimation error, and $\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}$ as the approximation error.

By the triangle inequality,

$$\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| \leq \|\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}\| + \|\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}\|.$$

Therefore, we can examine separately the contributions of the two parts in this decomposition to the integrated squared error. The contribution to the integrated squared error from the first part is bounded in probability by N_n/n, where N_n is the dimension of $\mathbf{G}_n$, while the contribution from the second part is governed by ρ_n, the approximation power of $\mathbf{G}_n$. The convergence rates for the two parts equal the corresponding rates for the independent scenario under some conditions.

First we state the conditions for the main results. The first two conditions, following Huang (1998), are on the approximating spaces. The first condition requires that the approximating space satisfies a stability constraint. This condition is satisfied by polynomials, trigonometric polynomials and splines. The second condition says that the approximating space must grow so that its distance from any function in $\mathbf{H}$ approaches zero.

For any function f on $[0, 1]$, set $\|f\|_\infty = \max_{x \in [0,1]} |f(x)|$.

Condition 2.1. *There are positive constants A_n such that, $\|f\|_\infty \leq A_n \|f\|$ for all $f \in \mathbf{G}_n$ and $\lim_n A_n^2 N_n / n = 0$.*

Condition 2.2. *There are nonnegative numbers $\rho_n = \rho_n(\mathbf{G}_n)$ such that for $\boldsymbol{\mu} \in \mathbf{H}$,*

$$\inf_{\mathbf{g} \in \mathbf{G}_n} \|\mathbf{g} - \boldsymbol{\mu}\|_\infty \leq \rho_n \to 0 \quad as \quad n \to \infty.$$

and $\limsup_n A_n \rho_n < \infty$.

If $\mathbf{H}$ is finite-dimensional, then we choose $\mathbf{G}_n = \mathbf{H}$, for all n. Condition 1 is automatically satisfied with A_n independent of n, and Condition 2 is satisfied with $\rho_n = 0$.

For the third condition, we require short-term dependence of errors.

Condition 2.3. *Let $\gamma_{|i-j|} = E\varepsilon_i\varepsilon_j$, then there is a positive constant $M \in \mathbb{R}^1$, such that $\sum_{i=1}^{\infty} |\gamma_i| \leq M$.*

This condition implies that the row or column sum of correlation matrix $\mathbf{R}$ is bounded by a constant.

Theorem 1. *Let $\mathbf{P_I}$ be the projection matrix of the ordinary least-squares estimation, then $\hat{\boldsymbol{\mu}}_\mathbf{I} = \mathbf{P_I y}$ and $\tilde{\boldsymbol{\mu}}_\mathbf{I} = \mathbf{P_I}\boldsymbol{\mu}$. If Conditions 2.1, 2.2 and 2.3 hold, then*

$$\|\hat{\boldsymbol{\mu}}_\mathbf{I} - \tilde{\boldsymbol{\mu}}_\mathbf{I}\|^2 = O_p(N_n/n), \quad \|\tilde{\boldsymbol{\mu}}_\mathbf{I} - \boldsymbol{\mu}\|^2 = O(\rho_n^2).$$

Consequently, $\|\hat{\boldsymbol{\mu}}_\mathbf{I} - \boldsymbol{\mu}\|^2 = O_p(N_n/n + \rho_n^2)$.

Huang (1998) derived the convergence rate of the least-squares estimate for independent observations in this general setting of both classical regression and nonparametric regression. Chapter 9 in Fuller (2009) derives the convergence rate for the least-squares estimate for correlated observations in linear regression. We propose a proof for this general setting with AR(p) errors.

Proof. Let $\{\boldsymbol{\psi}_j, 1 \le j \le N_n\}$ be an orthonormal basis of $\mathbf{G}_n$.

$$\hat{\boldsymbol{\mu}}_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}} = \sum_j \langle \hat{\boldsymbol{\mu}}_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}, \boldsymbol{\psi}_j \rangle \boldsymbol{\psi}_j = \sum_j \langle \mathbf{y} - \boldsymbol{\mu}, \boldsymbol{\psi}_j \rangle \boldsymbol{\psi}_j = \sum_j \langle \boldsymbol{\varepsilon}, \boldsymbol{\psi}_j \rangle \boldsymbol{\psi}_j.$$

Then, $\|\hat{\boldsymbol{\mu}}_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 = \frac{1}{n}\sum_j \langle \boldsymbol{\varepsilon}, \boldsymbol{\psi}_j \rangle^2$, and

$$\begin{aligned}
E\|\hat{\boldsymbol{\mu}}_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 &= \frac{1}{n}\sum_{j=1}^{N_n} E\langle \boldsymbol{\varepsilon}, \boldsymbol{\psi}_j \rangle^2 \\
&= \frac{1}{n}\sum_{j=1}^{N_n} \boldsymbol{\psi}_j^{'} \mathbf{R} \boldsymbol{\psi}_j \\
&= \frac{1}{n}\sum_{j=1}^{N_n}\sum_{l=1}^{n}\sum_{k=1}^{n} R_{lk}\psi_{lj}\psi_{kj} \\
&= \frac{1}{n}\sum_{j=1}^{N_n}\left[\sum_{l=1}^{n} R_{ll}\psi_{lj}^2 + 2\sum_{l=1}^{n}\sum_{k>l} R_{lk}\psi_{lj}\psi_{kj}\right] \\
&\le \frac{1}{n}\sum_{j=1}^{N_n}\left[\sum_{l=1}^{n} R_{ll}\psi_{lj}^2 + \sum_{l=1}^{n}\sum_{k>l} R_{lk}(\psi_{lj}^2 + \psi_{kj}^2)\right] \\
&= \frac{1}{n}\sum_{j=1}^{N_n}\left(\sum_{l=1}^{n}\sum_{k=l} R_{kl}\psi_{lj}^2 + \sum_{l=1}^{n}\sum_{k>l} R_{lk}\psi_{lj}^2 + \sum_{l=1}^{n}\sum_{k<l} R_{lk}\psi_{lj}^2\right) \\
&= \frac{1}{n}\sum_{j=1}^{N_n}\sum_{l=1}^{n}\left(\sum_{k=1}^{n} R_{kl}\right)\psi_{lj}^2 \\
&\le \frac{1}{n}\sum_{j=1}^{N_n}\sum_{l=1}^{n} M\psi_{lj}^2 \\
&= \frac{N_n}{n} M,
\end{aligned}$$

where R_{ij} is the i, jth element of $\mathbf{R}$, for $i, j = 1, \ldots, n$. So, $\|\hat{\boldsymbol{\mu}}_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 = O_p(N_n/n)$. That $\|\tilde{\boldsymbol{\mu}}_{\mathbf{I}} - \boldsymbol{\mu}\|^2 = O(\rho_n^2)$ is proved by Huang (1998). From Condition 2.2, we can find $\mathbf{g} \in \mathbf{G}_n$ such that $\|\boldsymbol{\mu} - \mathbf{g}\|_\infty \le 2\rho_n$ and hence $\|\boldsymbol{\mu} - \mathbf{g}\| \le 2\rho_n$. Then we have that $\|\tilde{\boldsymbol{\mu}}_{\mathbf{I}} - \mathbf{g}\|^2 =$

$\|\mathbf{P}(\boldsymbol{\mu}-\mathbf{g})\|^2 \le \|\boldsymbol{\mu}-\mathbf{g}\|^2$. Hence, by the triangle inequality,

$$\|\tilde{\boldsymbol{\mu}}_{\mathbf{I}}-\boldsymbol{\mu}\|^2 \le 2\|\tilde{\boldsymbol{\mu}}_{\mathbf{I}}-\mathbf{g}\|^2 + 2\|\boldsymbol{\mu}-\mathbf{g}\|^2 \le 4\|\boldsymbol{\mu}-\mathbf{g}\| = O(\rho_n^2).$$

Then, we have $\|\hat{\boldsymbol{\mu}}_{\mathbf{I}}-\boldsymbol{\mu}\|^2 = O_p(N_n/n+\rho_n^2)$. □

We need another condition to prove the next results.

Condition 2.4. *The error vector $\boldsymbol{\varepsilon}$ comes from a stationary AR(p) process, for an integrer $p \ge 1$.*

Theorem 2. *Let $\mathbf{P}_{\mathbf{R}}$ be the projection matrix of the weighted least-squares estimation with the known correlation matrix $\mathbf{R}$, then $\hat{\boldsymbol{\mu}}_{\mathbf{R}} = \mathbf{P}_{\mathbf{R}}\mathbf{y}$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{R}} = \mathbf{P}_{\mathbf{R}}\boldsymbol{\mu}$. If Conditions 2.1, 2.2, 2.3 and 2.4 hold, then*

$$\|\hat{\boldsymbol{\mu}}_{\mathbf{R}}-\tilde{\boldsymbol{\mu}}_{\mathbf{R}}\|^2 = O_p(N_n/n), \quad \|\tilde{\boldsymbol{\mu}}_{\mathbf{R}}-\boldsymbol{\mu}\|^2 = O(\rho_n^2).$$

Consequently, $\|\hat{\boldsymbol{\mu}}_{\mathbf{R}}-\boldsymbol{\mu}\|^2 = O_p(N_n/n+\rho_n^2)$.

Proof. Let $\mathbf{L}$ be the Cholesky decomposition of $\mathbf{R}$, then $\mathbf{R} = \mathbf{L}\mathbf{L}'$. Let $\mathbf{y}^* = \mathbf{L}^{-1}\mathbf{y}, \boldsymbol{\mu}^* = \mathbf{L}^{-1}\boldsymbol{\mu}, \boldsymbol{\varepsilon}^* = \mathbf{L}^{-1}\boldsymbol{\varepsilon}$, then the model can be transformed into $\mathbf{y}^* = \boldsymbol{\mu}^* + \boldsymbol{\varepsilon}^*$ and $E(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*'}) = \mathbf{I}$.

Let $\mathbf{G}_n^*$ be the transformed approximating subspace, spanned by $\mathbf{L}^{-1}\boldsymbol{\Psi}$, where the columns of $\boldsymbol{\Psi}$ span $\mathbf{G}_n$. Let $\hat{\boldsymbol{\mu}}^*$ be the orthogonal projection of $\mathbf{y}^*$ onto $\mathbf{G}_n^*$. Let $\tilde{\boldsymbol{\mu}}^*$ be the projection of $\boldsymbol{\mu}^*$ onto $\mathbf{G}_n^*$. By Theorem 2.1 in Huang (1998), we have $\|\hat{\boldsymbol{\mu}}^*-\tilde{\boldsymbol{\mu}}^*\|^2 = O_p(N_n/n)$, and $\|\tilde{\boldsymbol{\mu}}^*-\boldsymbol{\mu}^*\|^2 = O(\rho_n^2)$. Then $\|\hat{\boldsymbol{\mu}}^*-\tilde{\boldsymbol{\mu}}^*\|^2 = \|\mathbf{L}^{-1}(\hat{\boldsymbol{\mu}}_{\mathbf{R}}-\tilde{\boldsymbol{\mu}}_{\mathbf{R}})\|^2 = \frac{1}{n}(\hat{\boldsymbol{\mu}}_{\mathbf{R}}-\tilde{\boldsymbol{\mu}}_{\mathbf{R}})'\mathbf{R}^{-1}(\hat{\boldsymbol{\mu}}_{\mathbf{R}}-\tilde{\boldsymbol{\mu}}_{\mathbf{R}})$. Since $\mathbf{R}^{-1}$ is a Hermitian matrix, its eigenvalues are all real. By the Rayleigh-Ritz Theorem, the Rayleigh-Ritz ratio is bounded by the largest and smallest eigenvalues of $\mathbf{R}^{-1}$,

$$\lambda_{min} \le \frac{(\hat{\boldsymbol{\mu}}_{\mathbf{R}}-\tilde{\boldsymbol{\mu}}_{\mathbf{R}})'\mathbf{R}^{-1}(\hat{\boldsymbol{\mu}}_{\mathbf{R}}-\tilde{\boldsymbol{\mu}}_{\mathbf{R}})}{(\hat{\boldsymbol{\mu}}_{\mathbf{R}}-\tilde{\boldsymbol{\mu}}_{\mathbf{R}})'(\hat{\boldsymbol{\mu}}_{\mathbf{R}}-\tilde{\boldsymbol{\mu}}_{\mathbf{R}})} \le \lambda_{max},$$

where λ_{min} and λ_{max} are the smallest and largest eigenvalues of $\mathbf{R}^{-1}$. For $\mathbf{R}$ positive definite, it is easy to prove that $\mathbf{R}^{-1}$ is also positive definite. So, there exist two constant sequences m_n and M_n, where $0 < m_n \leq M_n < \infty$, for each specific n, such that, $m_n \leq \lambda_{min} \leq \lambda_{max} \leq M_n$, for each n. By Proposition 4.5.3 in Brockwell and Davis (2009), for a stationary AR(p) process, the eigenvalues of its covariance matrix are bounded away from zero and ∞ uniformly in n. Hence, for any n, there exist two constants M and m, such that $m \leq \lambda_{min} \leq \lambda_{max} \leq M$, for each n. Then $\frac{1}{M}\|\hat{\boldsymbol{\mu}}^* - \tilde{\boldsymbol{\mu}}^*\|^2 \leq \|\hat{\boldsymbol{\mu}}_{\mathbf{R}} - \tilde{\boldsymbol{\mu}}_{\mathbf{R}}\|^2 \leq \frac{1}{m}\|\hat{\boldsymbol{\mu}}^* - \tilde{\boldsymbol{\mu}}^*\|^2$. Therefore, $\|\hat{\boldsymbol{\mu}}_{\mathbf{R}} - \tilde{\boldsymbol{\mu}}_{\mathbf{R}}\|^2 = O_p(N_n/n)$.

By the same method used in the proof of $\|\hat{\boldsymbol{\mu}}_{\mathbf{R}} - \tilde{\boldsymbol{\mu}}_{\mathbf{R}}\|^2 = O_p(N_n/n)$, we can prove $\frac{1}{M}\|\tilde{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\|^2 \leq \|\tilde{\boldsymbol{\mu}}_{\mathbf{R}} - \boldsymbol{\mu}\|^2 \leq \frac{1}{m}\|\tilde{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\|^2$. Therefore, $\|\tilde{\boldsymbol{\mu}}_{\mathbf{R}} - \boldsymbol{\mu}\|^2 = O_p(\rho_n^2)$. So, we have $\|\hat{\boldsymbol{\mu}}_{\mathbf{R}} - \boldsymbol{\mu}\|^2 = O_p(N_n/n + \rho_n^2)$. □

For the case that $\mathbf{R}$ is unknown, we show that the estimation of trend is consistent and attains the same asymptotic rate as $\hat{\boldsymbol{\mu}}_{\mathbf{R}}$ for any suitable fixed $n \times n$ matrix $\mathbf{S}$ and argue that therefore the trend is estimated consistently with an estimator of $\mathbf{R}$ based on data.

Theorem 3. *If the correlation matrix* $\mathbf{R}$ *is unknown, and choose a sequence of matries* $\mathbf{S}$ *satisfying the following conditions:*

A1: $\mathbf{S}$ *is symmetric and positive-definite;*

A2: All the eigenvalues of $\mathbf{S}$ *are bounded from zero and* ∞*, uniformly in* n*;*

A3: Let $\mathbf{L}_{\mathbf{S}}$ *be the Cholesky decomposition of* $\mathbf{S}$*, then* $\mathbf{L}_{\mathbf{S}}^{-1}\mathbf{R}(\mathbf{L}_{\mathbf{S}}^{-1})'$ *satisfies Condition 3, which means the sum of the absolute value of its first row is bounded by a constant;*

Let $\mathbf{P}_{\mathbf{S}}$ *be the projection matrix of the weighted least-squares estimation with the substitute correlation matrix* $\mathbf{S}$*, then* $\hat{\boldsymbol{\mu}}_{\mathbf{S}} = \mathbf{P}_{\mathbf{S}}\mathbf{y}$ *and* $\tilde{\boldsymbol{\mu}}_{\mathbf{S}} = \mathbf{P}_{\mathbf{S}}\boldsymbol{\mu}$*. If conditions 2.1 and 2.2 hold, then*

$$\|\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}}\|^2 = O_p(N_n/n), \quad \|\tilde{\boldsymbol{\mu}}_{\mathbf{S}} - \boldsymbol{\mu}\|^2 = O(\rho_n^2).$$

Consequently, $\|\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \boldsymbol{\mu}\|^2 = O_p(N_n/n + \rho_n^2)$.

Proof. Let $\mathbf{y}_{\mathbf{S}}^* = \mathbf{L}_{\mathbf{S}}^{-1}\mathbf{y}, \boldsymbol{\mu}_{\mathbf{S}}^* = \mathbf{L}_{\mathbf{S}}^{-1}\boldsymbol{\mu}, \boldsymbol{\varepsilon}_{\mathbf{S}}^* = \mathbf{L}_{\mathbf{S}}^{-1}\boldsymbol{\varepsilon}$, then the model can be transformed into $\mathbf{y}_{\mathbf{S}}^* = \boldsymbol{\mu}_{\mathbf{S}}^* + \boldsymbol{\varepsilon}_{\mathbf{S}}^*, \quad E(\boldsymbol{\varepsilon}_{\mathbf{S}}^*\boldsymbol{\varepsilon}_{\mathbf{S}}^{*'}) = \mathbf{L}_{\mathbf{S}}^{-1}\mathbf{R}(\mathbf{L}_{\mathbf{S}}^{-1})'$. Let $\mathbf{G}_n^{\mathbf{S}}$ be the transformed approximating subspace. Let $\hat{\boldsymbol{\mu}}_{\mathbf{S}}^*$ be the projection of $\mathbf{y}_{\mathbf{S}}^*$ onto $\mathbf{G}_n^{\mathbf{S}}$. Let $\tilde{\boldsymbol{\mu}}_{\mathbf{S}}^*$ be the projection of $\boldsymbol{\mu}_{\mathbf{S}}^*$ onto $\mathbf{G}_n^{\mathbf{S}}$. For $\mathbf{L}_{\mathbf{S}}^{-1}\mathbf{R}(\mathbf{L}_{\mathbf{S}}^{-1})'$ satisfying the Condition A3, then we have $\|\hat{\boldsymbol{\mu}}_{\mathbf{S}}^* - \tilde{\boldsymbol{\mu}}_{\mathbf{S}}^*\|^2 = O_p(N_n/n)$, and $\|\tilde{\boldsymbol{\mu}}_{\mathbf{S}}^* - \boldsymbol{\mu}_{\mathbf{S}}^*\|^2 = O(\rho_n^2)$, by Theorem 8. Therefore, $\|\hat{\boldsymbol{\mu}}_{\mathbf{S}}^* - \tilde{\boldsymbol{\mu}}_{\mathbf{S}}^*\|^2 = \|\mathbf{L}_{\mathbf{S}}^{-1}(\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}})\|^2 = \frac{1}{n}(\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}})'\mathbf{S}_n^{-1}(\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}})$.

Then, by Rayleigh-Ritz Theorem, we have,

$$\lambda_{min}^{\mathbf{S}} \leq \frac{(\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}})'\mathbf{S}_n^{-1}(\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}})}{(\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}})'(\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}})} \leq \lambda_{max}^{\mathbf{S}},$$

where, $\lambda_{min}^{\mathbf{S}}$ and $\lambda_{max}^{\mathbf{S}}$ are the smallest and largest eigenvalues of $\mathbf{S}_n^{-1}$. By the condition A2 that the eigenvalues of $\mathbf{S}$ are bounded away from zero and ∞ uniformly in n, the eigenvalues of $\mathbf{S}^{-1}$ are also bounded from zero and ∞, which means that there exist two constants m and M, where $0 < m \leq M < \infty$, such that, $m \leq \lambda_{min}^{\mathbf{S}} \leq \lambda_{max}^{\mathbf{S}} \leq M$, for each n. Then $\frac{1}{M}\|\hat{\boldsymbol{\mu}}_{\mathbf{S}}^* - \tilde{\boldsymbol{\mu}}_{\mathbf{S}}^*\|^2 \leq \|\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}}\|^2 \leq \frac{1}{m}\|\hat{\boldsymbol{\mu}}_{\mathbf{S}}^* - \tilde{\boldsymbol{\mu}}_{\mathbf{S}}^*\|^2$. Therefore, $\|\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{S}}\|^2 = O_p(N_n/n)$. As in the proof of Theorem 3.6, we have $\frac{1}{M}\|\tilde{\boldsymbol{\mu}}_{\mathbf{S}}^* - \boldsymbol{\mu}_{\mathbf{S}}^*\|^2 \leq \|\tilde{\boldsymbol{\mu}}_{\mathbf{S}} - \boldsymbol{\mu}\|^2 \leq \frac{1}{m}\|\tilde{\boldsymbol{\mu}}_{\mathbf{S}}^* - \boldsymbol{\mu}_{\mathbf{S}}^*\|^2$. Thus, we have $\|\tilde{\boldsymbol{\mu}}_{\mathbf{S}} - \boldsymbol{\mu}\|^2 = O_p(\rho_n^2)$, and $\|\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \boldsymbol{\mu}\|^2 = O_p(N_n/n + \rho_n^2)$. □

Remark 1. *Theorems 8, 2 and 3 can readily be applied to classical linear regression. Let* $\mathbf{G}_n$ *be the linear space spanned by the columns of* $\mathbf{X}$*, where* $\mathbf{X}$ *is an* $n \times p$ *full row-rank matrix with fixed values, so that* $\mathbf{G}_n = \mathbf{H}$*,* $N_n = p$ *and* $\rho_n = 0$*. The three estimators of* $\boldsymbol{\mu}$ *achieve the same convergence rate of* $n^{-1/2}$*.*

The next theorem proves the consistency of the estimates of the correlation function.

Theorem 4. *Let* $\hat{\boldsymbol{\mu}}$ *be a consistent estimator of* $\boldsymbol{\mu}$*. Let* $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\boldsymbol{\mu}}$*,* $\hat{\gamma}_{\hat{\boldsymbol{\varepsilon}}}(h) = \frac{1}{n}\sum_{i=1}^{n-h}\hat{\varepsilon}_i\hat{\varepsilon}_{i+h}$*,* $\hat{\gamma}_{\boldsymbol{\varepsilon}}(h) = \frac{1}{n}\sum_{i=1}^{n-h}\varepsilon_i\varepsilon_{i+h}$*,* $\gamma(h) = \mathbf{E}\varepsilon_i\varepsilon_{i+h}$*, where* $i = 1, \ldots, n-h; h = 0, 1, \ldots, n-1$*. Let*

$\boldsymbol{\gamma}_{\hat{\boldsymbol{\varepsilon}}} = (\gamma_{\hat{\boldsymbol{\varepsilon}}}(1), \ldots, \gamma_{\hat{\boldsymbol{\varepsilon}}}(n-1))'$ *and* $\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}} = (\gamma_{\boldsymbol{\varepsilon}}(1), \ldots, \gamma_{\boldsymbol{\varepsilon}}(n-1))'$, *then*

$$\|\hat{\boldsymbol{\gamma}}_{\hat{\boldsymbol{\varepsilon}}} - \hat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}}\|^2 = O_p(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2), \quad \|\hat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}} - \boldsymbol{\gamma}\|^2 = O_p(1/n).$$

Consequently, $\|\hat{\boldsymbol{\gamma}}_{\hat{\boldsymbol{\varepsilon}}} - \boldsymbol{\gamma}\|^2 = O_p\left(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2 + \frac{1}{n}\right)$.

Proof.

$$\begin{aligned}
\hat{\boldsymbol{\gamma}}_{\hat{\boldsymbol{\varepsilon}}}(h) - \hat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}}(h) &= \frac{1}{n}\sum_{i=1}^{n-h}(\hat{\varepsilon}_i\hat{\varepsilon}_{i+h} - \varepsilon_i\varepsilon_{i+h}) \\
&= \frac{1}{n}\sum_{i=1}^{n-h}[\varepsilon_i(\mu_{i+h} - \hat{\mu}_{i+h}) + \varepsilon_{i+h}(\mu_i - \hat{\mu}_i) + (\mu_i - \hat{\mu}_i)(\mu_{i+h} - \hat{\mu}_{i+h})] \\
&\leq (\frac{1}{n}\sum_{i=1}^{n-h}\varepsilon_i^2)^{\frac{1}{2}}[\frac{1}{n}\sum_{i=1}^{n-h}(\mu_{i+h} - \hat{\mu}_{i+h})^2]^{\frac{1}{2}} + (\frac{1}{n}\sum_{i=1}^{n-h}\varepsilon_{i+h}^2)^{\frac{1}{2}}[\frac{1}{n}\sum_{i=1}^{n-h}(\mu_i - \hat{\mu}_i)^2]^{\frac{1}{2}} + \\
&\quad [\frac{1}{n}\sum_{i=1}^{n-h}(\mu_{i+h} - \hat{\mu}_{i+h})^2]^{\frac{1}{2}}[\frac{1}{n}\sum_{i=1}^{n-h}(\mu_i - \hat{\mu}_i)^2]^{\frac{1}{2}} \\
&= O_p(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\| + \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2).
\end{aligned}$$

If $\hat{\boldsymbol{\mu}}$ is consistent, $\|\hat{\boldsymbol{\gamma}}_{\hat{\boldsymbol{\varepsilon}}} - \hat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}}\|^2 = \frac{1}{n-1}\sum_{h=1}^{n-1}[\hat{\boldsymbol{\gamma}}_{\hat{\boldsymbol{\varepsilon}}}(h) - \hat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}}(h)]^2 = O_p(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2)$. By Theorem 7.2.1 in Brockwell and Davis (2009), $\sqrt{n}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}}(h))$ is asymptotic normally distributed, for $h = 0, 1, \ldots, n-1$. Thus $\|\hat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}} - \boldsymbol{\gamma}\|^2 = O_p(1/n)$. Therefore, we have $\|\hat{\boldsymbol{\gamma}}_{\hat{\boldsymbol{\varepsilon}}} - \boldsymbol{\gamma}\|^2 = O_p(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|^2 + \frac{1}{n})$. □

In the proposed iteration procedure, the pilot fit is the estimation ignoring correlation. By Theorem 3.8, the estimated correlation based on the pilot fit is consistent. Therefore it satisfies Conditions A1, A2 and A3. By Theorem 3.7, the renewed estimation of trend based on this consistent estimator of correlation is consistent and attains the optimal asymptotic rate.

Remark 2. *For classical linear regression,* $\boldsymbol{\mu} - \hat{\boldsymbol{\mu}} = O_p(n^{-1/2})$, *by Theorem 9.3.1 in Fuller (2009),* $\hat{\boldsymbol{\gamma}}_{\hat{\boldsymbol{\varepsilon}}} - \hat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}} = O_p(1/n)$.

2.3.2 Fixed-Knot Unpenalized Unconstrained Spline Regression

Theorems 8, 2 and 3 can also be applied to fixed-knot spline estimates when the knot positions are pre-specified but the number of knots is allowed to increase with the sample size. In this section, we investigate the large sample theory for only unpenalized situation, i.e. $\lambda = 0$. Suppose f is $p - smooth$ for a specified positive number p, that is, f is $[p]$ times continuously differentiable on $\mathbf{H}$, where $[p]$ is the greatest integer less than p, and all the $[p]th - order$ mixed partial derivatives of f satisfy a Hölder condition with exponent $p - [p]$, referring to Huang (1998). Let $\mathbf{G}_n$ be the linear space of regression splines with degree $d \geq p - 1$. Suppose the knots have bounded mesh ratio, that is, the ratio of the largest inter-knot interval to the smallest is bounded from zero and infinity, uniformly in n. Let a_n denote the smallest distance between two consecutive knots. For the two sequences of positive numbers a_{1n} and a_{2n}, let $a_{1n} \asymp a_{2n}$ mean that the ratio a_{1n}/a_{2n} is bounded away form zero and ∞. Then we have $N_n \asymp 1/a_n$ and $\rho_n \asymp a_n^p \asymp N_n^{-p}$. Hence, the convergence rate for the three estimators, i.e. $\hat{\boldsymbol{\mu}}_{\mathbf{I}}$, $\hat{\boldsymbol{\mu}}_{\mathbf{R}}$ and $\hat{\boldsymbol{\mu}}_{\mathbf{S}}$, is $O_p(\frac{a_n}{n} + a_n^{2p})$. In order to let the rate of convergence be optimal, which means no estimate has a faster rate of convergence uniformly over the class of p-smooth functions, referring to Stone (1982), choose $a_n \asymp n^{-1/(2p+1)}$. This balances the estimation error and the approximation error, that is, $\frac{a_n}{n} \asymp a_n^{2p}$. Applying Theorems 8, 2 and 3 to this setting, we obtain the following results.

Corollary 1. *Suppose conditions 1, 2 and 3 hold and the knots have bounded mesh ratio. If we choose $a_n \asymp n^{-1/(2p+1)}$, then*

$$\|\hat{\boldsymbol{\mu}}_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 = O_p(n^{-2p/(2p+1)}), \quad \|\tilde{\boldsymbol{\mu}}_{\mathbf{I}} - \boldsymbol{\mu}\|^2 = O(n^{-2p/(2p+1)});$$

$$\|\hat{\boldsymbol{\mu}}_{\mathbf{R}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 = O_p(n^{-2p/(2p+1)}), \quad \|\tilde{\boldsymbol{\mu}}_{\mathbf{R}} - \boldsymbol{\mu}\|^2 = O(n^{-2p/(2p+1)});$$

$$\|\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 = O_p(n^{-2p/(2p+1)}), \quad \|\tilde{\boldsymbol{\mu}}_{\mathbf{S}} - \boldsymbol{\mu}\|^2 = O(n^{-2p/(2p+1)}).$$

Consequently,

$$\|\hat{\boldsymbol{\mu}}_{\mathbf{I}} - \boldsymbol{\mu}\|^2 = O_p(n^{-2p/(2p+1)});$$

$$\|\hat{\boldsymbol{\mu}}_{\mathbf{R}} - \boldsymbol{\mu}\|^2 = O_p(n^{-2p/(2p+1)});$$

$$\|\hat{\boldsymbol{\mu}}_{\mathbf{S}} - \boldsymbol{\mu}\|^2 = O_p(n^{-2p/(2p+1)}).$$

2.3.3 Fixed Knot Constrained Unpenalized Spline Regression

In Theorems 8, 2 and 3, we derived the convergence rate in a general setting for both classical regression and nonparametric regression. The next theorem will compare the convergence rate of constrained estimator and the corresponding unconstrained estimator in spline regression in the presence of correlated errors. Let $\mathbf{G}_n^u = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \mathbf{Bb}\}$, which is a finite-dimensional approximating subspace to $\mathbf{H}$ spanned by spline basis. Assume $f \in \mathbf{H}_c$, a subset of all square-integrable, real-valued, constrained functions on $\mathcal{X}$; $\boldsymbol{\mu} \in \mathbb{R}^n$, where $\mu_i = f(x_i)$. Let $\mathbf{G}_n^c = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \mathbf{Bb}, \mathbf{Tb} \geq 0\}$, which is a finite-dimensional approximating subset of $\mathbf{H}_c$.

As before, we consider three kinds of estimators: ordinary least-squares, the weighted least-squares with known $\mathbf{R}$ and the weighted least-squares using a given matrix $\mathbf{S}$ as an substitute of correlation, for both constrained spline regression and unconstrained spline regression, and compare their convergence rates. Let $\mathbf{P}_{\mathbf{I}}^c$ be the projection matrix of the ordinary least-squares estimator in the constrained spline regression. It is a random matrix and depends on J, the index of the face identified by cone projection algorithm for a specific $\mathbf{y}$. Let $\hat{\boldsymbol{\mu}}_{\mathbf{I}}^c = \mathbf{P}_{\mathbf{I}}^c\mathbf{y}$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{I}}^c = \mathbf{P}_{\mathbf{I}}^c\boldsymbol{\mu}$. Let $\mathbf{P}_{\mathbf{I}}^u$ be the projection matrix of the ordinary least-squares estimator in the unconstrained spline estimator. It is a fixed matrix, and corresponds to $\mathbf{P}_{\mathbf{I}}^c$ with $J = \{1, \cdots, m\}$. Let $\hat{\boldsymbol{\mu}}_{\mathbf{I}}^u = \mathbf{P}_{\mathbf{I}}^u\mathbf{y}$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{I}}^u = \mathbf{P}_{\mathbf{I}}^u\boldsymbol{\mu}$. Let $\mathbf{P}_{\mathbf{R}}^c$ be the projection matrix of the weighted least-squares estimator in the constrained spline regression with the known $\mathbf{R}$, then $\hat{\boldsymbol{\mu}}_{\mathbf{R}}^c = \mathbf{P}_{\mathbf{R}}^c\mathbf{y}$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{R}}^c = \mathbf{P}_{\mathbf{R}}^c\boldsymbol{\mu}$. Let $\mathbf{P}_{\mathbf{R}}^u$ be the projection matrix of the weighted least-squares estimator in the unconstrained spline estimator, then $\hat{\boldsymbol{\mu}}_{\mathbf{R}}^u = \mathbf{P}_{\mathbf{R}}^u\mathbf{y}$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{R}}^u = \mathbf{P}_{\mathbf{R}}^u\boldsymbol{\mu}$. Let $\mathbf{P}_{\mathbf{S}}^c$ be the projection matrix of the weighted least-squares estimator in the constrained spline regression with the given matrix $\mathbf{S}$ as an estimator of the unknown $\mathbf{R}$, then $\hat{\boldsymbol{\mu}}_{\mathbf{S}}^c = \mathbf{P}_{\mathbf{S}}^c\mathbf{y}$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{S}}^c = \mathbf{P}_{\mathbf{S}}^c\boldsymbol{\mu}$. Let $\mathbf{P}_{\mathbf{S}}^u$ be the projection matrix of the weighted least-squares estimator in the unconstrained spline estimator with the given matrix $\mathbf{S}$ as an estimator of the unknown $\mathbf{R}$, then $\hat{\boldsymbol{\mu}}_{\mathbf{S}}^u = \mathbf{P}_{\mathbf{S}}^u\mathbf{y}$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{S}}^u = \mathbf{P}_{\mathbf{S}}^u\boldsymbol{\mu}$.

Assume that $\tilde{\boldsymbol{\mu}}_{\mathbf{I}}^u \in \mathbf{G}_n^c$, so that the shape restrictions hold; otherwise, $\tilde{\boldsymbol{\mu}}_{\mathbf{I}}^u$ cannot be consistent for $\boldsymbol{\mu}$, which is assumed to follow the given shape restrictions. Under this assumption, it is easy to prove that $\tilde{\boldsymbol{\mu}}_{\mathbf{I}}^u = \tilde{\boldsymbol{\mu}}_{\mathbf{I}}^c$. The same assumption and reasoning also apply to the other two estimators, therefore $\tilde{\boldsymbol{\mu}}_{\mathbf{R}}^u = \tilde{\boldsymbol{\mu}}_{\mathbf{R}}^c$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{S}}^u = \tilde{\boldsymbol{\mu}}_{\mathbf{S}}^c$. In this context, we use $\tilde{\boldsymbol{\mu}}_{\mathbf{I}}$ instead of $\tilde{\boldsymbol{\mu}}_{\mathbf{I}}^u$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{I}}^c$. The same treatment is used for $\tilde{\boldsymbol{\mu}}_{\mathbf{R}}$ and $\tilde{\boldsymbol{\mu}}_{\mathbf{S}}$. Therefore, the approximation error for the constrained estimators and unconstrained estimators in the same setting are the same, and the comparison of the total error is reduced to the comparison of the estimation error.

Theorem 5. *Let the knots $t_1, \ldots, t_k$ have bounded mesh ratio, then*

$$\|\hat{\boldsymbol{\mu}}_{\mathbf{I}}^c - \boldsymbol{\mu}\|^2 \leq \|\hat{\boldsymbol{\mu}}_{\mathbf{I}}^u - \boldsymbol{\mu}\|^2.$$

Hence the convergence rate of the ordinary least-squares estimator in constrained spline regression attains that of the corresponding unconstrained spline regression, in the presence of correlation.

Proof. The decomposition of errors is $\hat{\boldsymbol{\mu}}^c_{\mathbf{I}} - \boldsymbol{\mu} = (\hat{\boldsymbol{\mu}}^c_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}) + (\tilde{\boldsymbol{\mu}}_{\mathbf{I}} - \boldsymbol{\mu})$ and $\hat{\boldsymbol{\mu}}^u_{\mathbf{I}} - \boldsymbol{\mu} = (\hat{\boldsymbol{\mu}}^u_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}) + (\tilde{\boldsymbol{\mu}}_{\mathbf{I}} - \boldsymbol{\mu})$. So we only need to prove $\|\hat{\boldsymbol{\mu}}^c_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 \leq \|\hat{\boldsymbol{\mu}}^u_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2$. We have

$$\begin{aligned}
\|\hat{\boldsymbol{\mu}}^u_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 &= \|\hat{\boldsymbol{\mu}}^c_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 + \|\hat{\boldsymbol{\mu}}^u_{\mathbf{I}} - \hat{\boldsymbol{\mu}}^c_{\mathbf{I}}\|^2 + 2(\hat{\boldsymbol{\mu}}^u_{\mathbf{I}} - \hat{\boldsymbol{\mu}}^c_{\mathbf{I}})^t(\hat{\boldsymbol{\mu}}^c_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}) \\
&= \|\hat{\boldsymbol{\mu}}^c_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 + \|\hat{\boldsymbol{\mu}}^u_{\mathbf{I}} - \hat{\boldsymbol{\mu}}^c_{\mathbf{I}}\|^2 \\
&\quad -2(\mathbf{y} - \hat{\boldsymbol{\mu}}^u_{\mathbf{I}})^t(\hat{\boldsymbol{\mu}}^c_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}) + 2(\mathbf{y} - \hat{\boldsymbol{\mu}}^c_{\mathbf{I}})^t(\hat{\boldsymbol{\mu}}^c_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}).
\end{aligned}$$

The Karush-Kuhn-Tucker conditions (see Silvapulle and Sen (2004a) Appendix 1) imply, $\langle \mathbf{y} - \hat{\boldsymbol{\mu}}^c_{\mathbf{I}}, \hat{\boldsymbol{\mu}}^c_{\mathbf{I}} \rangle = 0$ and $\langle \mathbf{y} - \hat{\boldsymbol{\mu}}^c_{\mathbf{I}}, \tilde{\boldsymbol{\mu}}^c_{\mathbf{I}} \rangle \leq 0$. Therefore, $\|\hat{\boldsymbol{\mu}}^u_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2 \geq \|\hat{\boldsymbol{\mu}}^c_{\mathbf{I}} - \tilde{\boldsymbol{\mu}}_{\mathbf{I}}\|^2$. □

Theorem 6. *Let the knots $t_1, \ldots, t_k$ have bounded mesh ratio. Then there exists a constant $C \in \mathbb{R}^1$, bounded away from zero and ∞, such that,*

$$\|\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \boldsymbol{\mu}\|^2 \leq C\|\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \boldsymbol{\mu}\|^2.$$

Hence the convergence rate of the weighted least-squares estimator with known correlation in constrained spline regression attains that of the corresponding unconstrained spline regression.

Proof. Let $\mathbf{L}$ be the Cholesky decomposition of $\mathbf{R}$, then $\mathbf{R} = \mathbf{L}\mathbf{L}'$. Let $\mathbf{y}^* = \mathbf{L}^{-1}\mathbf{y}, \boldsymbol{\mu}^* = \mathbf{L}^{-1}\boldsymbol{\mu}$, and $\boldsymbol{\varepsilon}^* = \mathbf{L}^{-1}\boldsymbol{\varepsilon}$, then the model can be transformed into $\mathbf{y}^* = \boldsymbol{\mu}^* + \boldsymbol{\varepsilon}^*$ and $E(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*'}) = \mathbf{I}$. Using the result in Theorem 5, we have $\|\hat{\boldsymbol{\mu}}^*_c - \tilde{\boldsymbol{\mu}}^*\|^2 \leq \|\hat{\boldsymbol{\mu}}^*_u - \tilde{\boldsymbol{\mu}}^*\|^2$, and when transformed back, we get $\|\mathbf{L}^{-1}\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \mathbf{L}^{-1}\tilde{\boldsymbol{\mu}}_{\mathbf{R}}\|^2 \leq \|\mathbf{L}^{-1}\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \mathbf{L}^{-1}\tilde{\boldsymbol{\mu}}_{\mathbf{R}}\|^2$, that is $(\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})'\mathbf{R}^{-1}(\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \tilde{\boldsymbol{\mu}}) \leq (\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})'\mathbf{R}^{-1}(\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})$. Since $\mathbf{R}^{-1}$ is Hermitian matrix, its eigenvalues are all real. By the Rayleigh-Ritz Theorem, the Rayleigh-Ritz ratio is bounded by the largest and smallest eigenvalues of $\mathbf{R}^{-1}$. Then we have,

$$\frac{(\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})'\mathbf{R}^{-1}(\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})}{(\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})'(\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})} \geq \lambda_{min} \quad \text{and} \quad \frac{(\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})'\mathbf{R}^{-1}(\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})}{(\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})'(\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \tilde{\boldsymbol{\mu}})} \leq \lambda_{max};$$

where, λ_{min} and λ_{max} are the smallest and largest eigenvalues of $\mathbf{R}^{-1}$. There exist two constant sequences m_n and M_n, where $0 < m_n \leq M_n < \infty$, for each specific n, such that, $m_n \leq \lambda_{min} \leq \lambda_{max} \leq M_n$, for each n.As in the proof of Theorem 2 for any n, there exist two constants M and n, such that $\|\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \tilde{\boldsymbol{\mu}}\|^2 \leq \frac{M}{m}\|\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \tilde{\boldsymbol{\mu}}\|^2$. So, there exist a constant $C \in \mathbb{R}^1$, bounded away from zero and ∞, such that, $\|\hat{\boldsymbol{\mu}}^c_{\mathbf{R}} - \boldsymbol{\mu}\|^2 \leq C\|\hat{\boldsymbol{\mu}}^u_{\mathbf{R}} - \boldsymbol{\mu}\|^2$. □

Theorem 7. *Let the knots $t_1, \ldots, t_k$ have "bounded mesh ratio". The correlation matrix $\mathbf{R}$ is unknown, so choose any matrix $\mathbf{S}$ satisfying the conditions A1, A2 and A3. Then there exists a constant $C \in \mathbb{R}^1$, bounded away from zero and ∞, such that,*

$$\|\hat{\boldsymbol{\mu}}^c_{\mathbf{S}} - \boldsymbol{\mu}\|^2 \leq C\|\hat{\boldsymbol{\mu}}^u_{\mathbf{S}} - \boldsymbol{\mu}\|^2, \textit{with probability approaching one.}$$

Hence the convergence rate of the weighted least-squares estimator with the given matrix as an substitue of correlation in constrained spline regression attains that in the corresponding unconstrained spline regression.

Using Theorems 5 and 6, the proof is entirely analogous to that of Theorem 3, and is omitted here.

2.4 Simulation

2.4.1 Data Scenarios

Simulations were carried out to examine the performance of the constrained penalized spline estimator, and to compare it with the unconstrained penalized spline estimator and the classical linear regression estimator. Data with different scenarios of trend and noise were generated. For the trend, linear, sigmoid and truncated cubic are used. For the noise, a series of AR(p) errors, where $p = 1, 2, 3, 4$, with gradually increasing correlation were generated, then $y_i = f(i/n) + \epsilon_i$. For the mean function $f(x)$, we used

1. linear: $f(x) = x$
2. sigmoid: $f(x) = \frac{e^{10x-5}}{1+e^{10x-5}}$
3. truncated cubic: $f(x) = 4(x - 1/2)^3 I_{x>1/2}$.

Two series of noise were used in simulations:

- $\epsilon_i = \theta\epsilon_{i-1} + z_i, \theta = 0.1, 0.3, 0.5, 0.7$
- $\epsilon_i = 0.3\epsilon_{i-p} + z_i, p = 1, 2, 3, 4$

where $z_i^{'}s$ are independent and identically normally distributed with mean zero and standard deviation 0.2. The sample size for the simulation is 250 and the number of replications is 1000.

2.4.2 The selection of order p and penalty parameter by AIC

Many authors have studied the effects of correlation on the selection of the smoothing parameter and derived correlation-adjusted selection methods, see Diggle and Hutchinson (1989), Altman (1990) and Wang (1998). None of these selects the order of correlation and the smoothing parameter simultaneously for penalized spline regression. In this article, we use a correlation-adjusted AIC (1) criterion to select the penalty parameter and the order p simultaneously.

For each simulated data set, we compute 60 AIC values using $p = (0, 1, 2, 3, 4, 5)$ and ten values of λ as candidates. The effective degrees of freedom for both constrained and unconstrained estimators with unknown correlation are random. We choose the candidate λ by letting the corresponding effective degrees of freedom for unconstrained penalized spline estimator for independent data be $(4, 5, 6, 8, 10, 12, 16, 20, 25, 30)$. We choose p and λ as the

Table 2.1: The proportion of datasets for which the correlation-adjusted AIC criterion selects the true p, for the proposed estimator, the unconstrained penalized estimator and the classical linear regression estimator. The simulated data are generated by three different mean functions, linear, truncated cubic and sigmoid, with AR(1) errors, where $\theta = 0, 0.1, 0.3, 0.5, 0.7$ and AR(p) errors, where $\epsilon_i = 0.3\epsilon_{i-p} + z_i, p = 0, 2, 3, 4$.

		AR(1)				AR(p)		
f	$\boldsymbol{\theta}$	constrained	unconstrained	linear	p	constrained	unconstrained	linear
linear	0	0.617	0.619	0.717	0	0.617	0.619	0.717
	0.3	0.617	0.581	0.740	2	0.631	0.599	0.756
	0.5	0.663	0.632	0.760	3	0.645	0.599	0.799
	0.7	0.677	0.620	0.761	4	0.695	0.613	0.834
cubic	0	0.673	0.592	0.078	0	0.673	0.592	0.078
	0.3	0.686	0.601	0.541	2	0.709	0.561	0.595
	0.5	0.697	0.587	0.700	3	0.748	0.562	0.691
	0.7	0.704	0.595	0.750	4	0.785	0.593	0.797
sigmoid	0	0.575	0.532	0.019	0	0.575	0.532	0.019
	0.3	0.605	0.546	0.426	2	0.625	0.549	0.521
	0.5	0.658	0.611	0.669	3	0.627	0.520	0.649
	0.7	0.663	0.599	0.727	4	0.720	0.582	0.744

joint minimizer of $AIC_{p,\lambda}$. We repeat this procedure for $N = 1000$ times, and calculate the fraction of times that AIC chooses the true p.

In Table 1, for truncated cubic data and sigmoid data, the values in first and third columns are always greater than the values in the corresponding second and fourth columns, which is just as we expected when the assumptions on the shape are correct. Linear regression behaves poorly for truncated cubic data and sigmoid data when the correlation is zero or small because it is more likely to choose a larger p. When the data are generated by a linear trend, the linear regression does the best job but the behavior of the proposed estimator is still acceptable. We also conducted the simulations with larger p and higher degree of correlated errors, such as, AR(4) with $\theta = (0.4, 0.3, 0.15, 0.1)$. If the correlation is large enough, it can cause the failure of the AIC to select the true p for all three methods.

2.4.3 Three Performance Measures

To compare the performance of the proposed estimator with the unconstrained penalized spline estimator and classical linear regression estimator, the following three measures are constructed. The first and the second measures are used to compare the estimations of the correlation. The third one is used to compare the estimations of the trend.

$$\Delta_{\boldsymbol{\theta}} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\sum_{k=1}^{\mathtt{K}}(\hat{\theta}_{i\hat{p}k} - \theta_k)^2}; \quad \Delta_{\boldsymbol{\gamma}} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\sum_{h=1}^{20}(\hat{\gamma}_{i\hat{p}h} - \gamma_h)^2};$$

and

$$\Delta_{\mathbf{f}} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{\sum_{l=1}^{n}(\hat{f}_{\hat{p}i}(x_l) - f(x_l))^2}{n}}.$$

Here, $\mathtt{K}$ is the largest length of θ, $\theta = (\theta_1, ..., \theta_K)$, $\hat{f}_{\hat{p}i}(x_l)$ is the estimated mean at a specific value x_l under selected order $\hat{p}$ in ith repetition, and $\hat{\gamma}_{i\hat{p}h}$ is the estimator of γ_h with $\hat{p}$ in ith repetition.

In Tables 2 and 3, the values in the first, third and fifth columns are all positive, except for the linear trend with independent error for measure 2. These results show that the constrained penalized spline estimator behaves better than the unconstrained penalized spline estimator in the estimation of both the trend and correlation. For linear data, comparing with unconstrained spline estimator, the proposed estimator still improves around $5\% - 10\%$ for measure 1, $3\% - 19\%$ for measure 2, and $1\% - 13\%$ for measure 3. For both cubic data and sigmoid data, the superiority of the proposed estimator in the estimation of both the trend and correlation is quite evident, where the improvement is around $26\% - 57\%$ for cubic data and $11\% - 31\%$ for sigmoid data. The improvements have an increasing trend with the increase of the correlation, because the constrained estimator is less sensitive than the unconstrained estimator to the increase of correlation.

Table 2.2: The simulated percentage of the proposed estimator's relative improvement in the three measures, comparing with the unconstrained spline estimator and the classical linear regression estimator. Each value is calculated as a ratio. The numerator is the difference of measures, i.e. measure of constrained estimator minus that of the unconstrained estimator or the linear estimator. The denominator is the corresponding measure of the constrained estimator. The datasets are generated by three different kinds of true mean functions, linear, truncated cubic and sigmoid, with AR(1) errors, where $\theta = 0, 0.3, 0.5, 0.7$.

f	$\boldsymbol{\theta}$	$\Delta_{\boldsymbol{\theta}}$		$\Delta_{\boldsymbol{\gamma}}$		$\Delta_{\mathbf{f}}$	
		uncon	linear	uncon	linear	uncon	linear
linear	0	5.2	-41.4	-0.41	-26.5	1.0	-40.8
	0.3	8.8	-26.2	8.06	-34.5	5.0	-40.3
	0.5	5.8	-26.2	9.18	-34.0	5.9	-40.8
	0.7	13.3	-25.5	19.34	-33.8	13.8	-36.8
cubic	0	60.6	284.6	25.5	192.1	25.5	224.0
	0.3	29.1	37.4	34.7	82.7	30.1	145.0
	0.5	29.7	1.5	46.6	24.9	30.2	89.1
	0.7	35.9	-15.5	57.2	-12.7	35.4	37.4
sigmoid	0	27.8	204.4	10.5	143.8	15.0	195.1
	0.3	15.8	30.0	19.9	91.3	13.2	115.4
	0.5	14.9	-4.7	21.1	20.2	15.4	71.0
	0.7	18.1	-21.3	29.9	-20.6	18.9	25.1

Table 2.3: The simulated percentage of the proposed estimator's relative improvement in the three measures, comparing with the unconstrained spline estimator and the classical linear regression estimator. Each value is calculated as a ratio. The numerator is the difference of measures, i.e. measure of constrained estimator minus that of the unconstrained estimator or the linear estimator. The denominator is the corresponding measure of the constrained estimator. The datasets are generated by three different kinds of true mean functions, linear, truncated cubic and sigmoid, with AR(p) errors, where $\epsilon_i = 0.3\epsilon_{i-p} + z_i, p = 0, 1, 2, 3, 4$.

		$\Delta_{\boldsymbol{\theta}}$		$\Delta_{\boldsymbol{\gamma}}$		$\Delta_{\mathbf{f}}$	
f	p	uncon	linear	uncon	linear	uncon	linear
linear	0	5.2	-41.4	-0.4	-26.5	1.0	-40.8
	1	8.8	-26.2	8.1	-34.5	5.0	-40.3
	2	7.8	-6.1	3.7	-28.2	4.9	-43.2
	3	10.2	-31.6	3.9	-30.0	8.2	-46.0
	4	16.2	-34.0	6.6	-24.8	12.2	-46.6
cubic	0	60.6	284.6	25.5	192.1	25.5	224.0
	1	29.1	37.4	34.7	82.7	30.1	145.0
	2	38.6	21.6	37.4	63.7	36.2	141.5
	3	46.2	16.8	34.4	71.8	42.9	151.6
	4	58.8	8.5	32.2	60.4	46.1	145.7
sigmoid	0	27.8	204.4	10.5	143.8	15.0	195.1
	1	15.8	30.0	19.9	91.2	13.2	115.4
	2	17.1	12.8	14.2	67.3	16.1	118.1
	3	24.6	0.8	16.1	66.2	18.1	111.4
	4	30.9	-5.8	13.6	67.5	20.9	113.5

Table 2.4: The simulated percentage of increase in measure 3 by ignoring the correlation, comparing with the estimation by proposed iteration. Each value is calculated as a ratio. The numerator is the difference of measures, i.e. measure of fit ignoring correlation minus the measure of fit obtaining from proposed iteration. The denominator is the corresponding measure of the proposed iteration estimator. The datasets are generated by three different kinds of true mean functions, linear, truncated cubic and sigmoid, with AR(1) errors, where $\theta = 0.3, 0.5, 0.7$. Sample size is 500.

f	$\boldsymbol{\theta}$	constrained	unconstrained	linear
linear	0.3	30.23	37.24	-2.54
	0.5	33.59	56.92	-0.12
	0.7	29.67	65.60	4.45
cubic	0.3	4.00	20.78	0.05
	0.5	11.38	38.06	0.02
	0.7	11.27	43.19	-0.40
sigmoid	0.3	8.11	13.94	-0.01
	0.5	12.21	23.94	-0.08
	0.7	14.70	37.25	-1.12

For the linear data, all the values in the first, third and fifth columns in both Tables 2 and 3 are negative. We cannot expect a nonparametric method to perform better than the correct parametric method. For cubic data and sigmoid data, the improvement of proposed method is noteworthy for small amount of correlation. But for the estimation of $\boldsymbol{\theta}$, the proposed method performs about as well as the linear model for large correlation. There are some extreme large positive values for cubic and sigmoid data with independent observations in the first, third and fifth columns in both Tables 2 and 3. These values demonstrate that incorrectly assuming a parametric form would cost a great deviation when there is no prior information of the parametric family of the trend. The deviation is evident when there is no correlation, and would be obscured when the correlation is increasing.

Finally, we investigate whether incorporating the correlation actually improves the estimation of the mean function, relative to the simpler estimator that completely ignores the correlation. Incorporating the correlation into the estimation procedure improves the estimation of trend, for both constrained spline regression and unconstrained spline regression. Table 4 shows the differences of effects whether estimating the correlation by the proposed iteration or ignoring the correlation for those three regression models. The increase of constrained estimator is smaller than that of the unconstrained estimator. Because of incorporation of qualitative knowledge of trend, the estimation of trend is more robust to the estimation of correlation. The increment of simple linear regression is almost zero and sometimes even negative.

2.5 Global Temperature Data

There has been much interest in the research of the global temperature change. Hansen et al. (2006) have a discussion on the pattern of global warming. In this article, we use the "Global Annual Mean Surface Air Temperature Change Data" from 1882 to 2008 to demonstrate the behavior of the proposed estimator. The data set comes from `http://data.giss.nasa.gov/gistemp/graphs_v3/Fig.A2.txt`. Assume that the global annual temperature is a stationary auto-regressive process with a monotone increasing tendency during the 1882 to 2008. We fit the data with the monotone constrained penalized spline regression and compare the performance with the unconstrained penalized spline estimator and the classical linear regression estimator. The correlated-adjusted AIC in this paper would be used to select the penalty parameter and the order p. We fit this data with 20 knots and 35 knots for a comparison. The results are in Figures 3.1 and 3.2.

For the situation of 20 knots, $\hat{p} = 2$ and $\hat{\boldsymbol{\theta}} = (0.278, -0.138)$ for constrained penalized spline estimator. This looks more reasonable than the results of the unconstrained penalized estimator, where $\hat{p} = 5$ and $\hat{\boldsymbol{\theta}} = (0.251, -0.160, -0.133, 0.114, -0.224)$. The penalty parameters for both the constrained and unconstrained regression are 0.024, so that the corresponding effective degree of freedom of constrained estimator is 8 and that of the unconstrained estimator is 14. Unconstrained spline regression tends to select the smallest value by AIC among all the candidate λ, which easily leads to overfitting and generates a wiggly curve. The linear regression is inclined to overestimate the correlation as expected, selecting

$\hat{p} = 4$ and $\hat{\boldsymbol{\theta}} = (0.507, 0.003, 0.050, 0.222)$. Woodward and Gray (1993) point out that statistical tests based on simple linear model have little or no ability to distinguish the realizations from the ARMA model with high correlation and those from the linear model. If the number of knots is increased to 35, the constrained penalized spline estimator still estimate $\hat{p} = 2$, and $\hat{\boldsymbol{\theta}} = (0.256, -0.158)$, which is quite similar to the case with 20 knots. But the behavior of the unconstrained penalized spline estimator becomes very unstable with $\hat{p} = 9$ and $\hat{\boldsymbol{\theta}} = (0.0247, -0.334, -0.333, -0.116, -0.364, -0.176, -0.154, -0.111, -0.186)$. From Figure 3.2, the unconstrained fit becomes more wiggly, but the constrained fit has little change.

We also compared the proposed estimation with constrained spline estimation with ignoring the correlation on the Global Temperature Data. But the two curves are almost identical to each other. The difference of the fits is quite small if we use the constrained method.

In conclusion, this example illustrates the robustness to knot choice of the constrained spline regression, and the ability of the proposed AIC criterion to select suitable values for the penalty and the order of the correlation automatically. Meyer (2012) has previously found the robustness of constrained spline regression to penalty choice for independent data, so this confirms these good practical properties for situations with correlated data.

2.6 Conclusion

The asymptotic rate for constrained spline estimators with estimation of correlation and ignoring the correlation have been proved to be the same. Even if we have an inconsistent estimator of correlation, as long as it satisfies appropriate conditions, the estimation of trend based on this estimator is still consistent and attains the optimal rate. However, as illustrated in Table 4, estimation of the trend is substantially improved for moderate-sized samples under proposed iteration method. Further, the asymptotic variances of the three estimators are different. In on-going research, we are studying the hypothesis tests of the

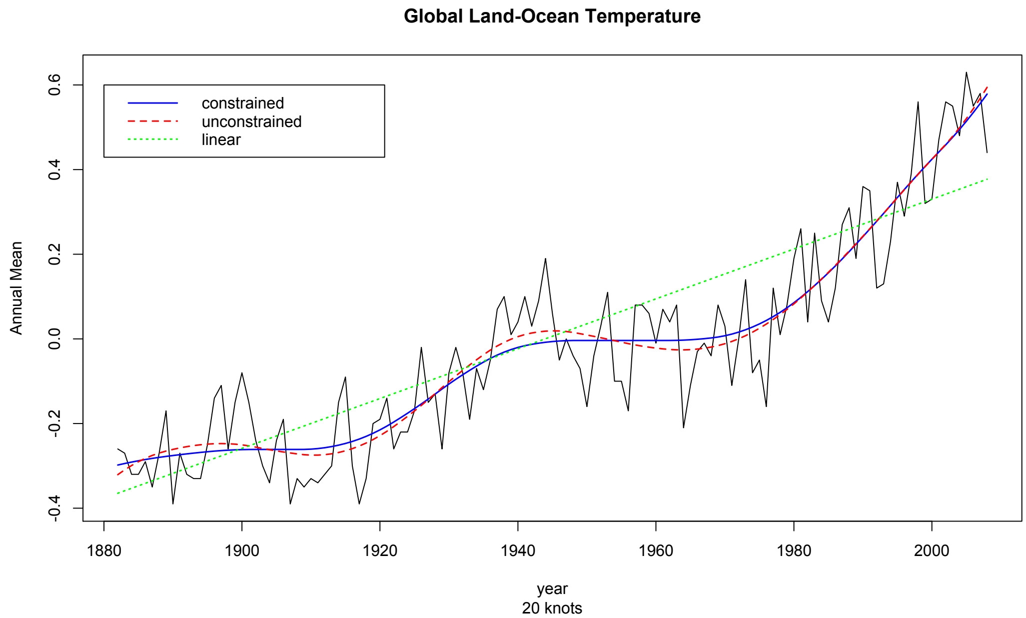

Figure 2.1: Estimated global temperature trends, using constrained penalized spline estimators and unconstrained penalized spline estimator both with 20 knots and penalty and correlation order selected with AIC, and linear regression fit.

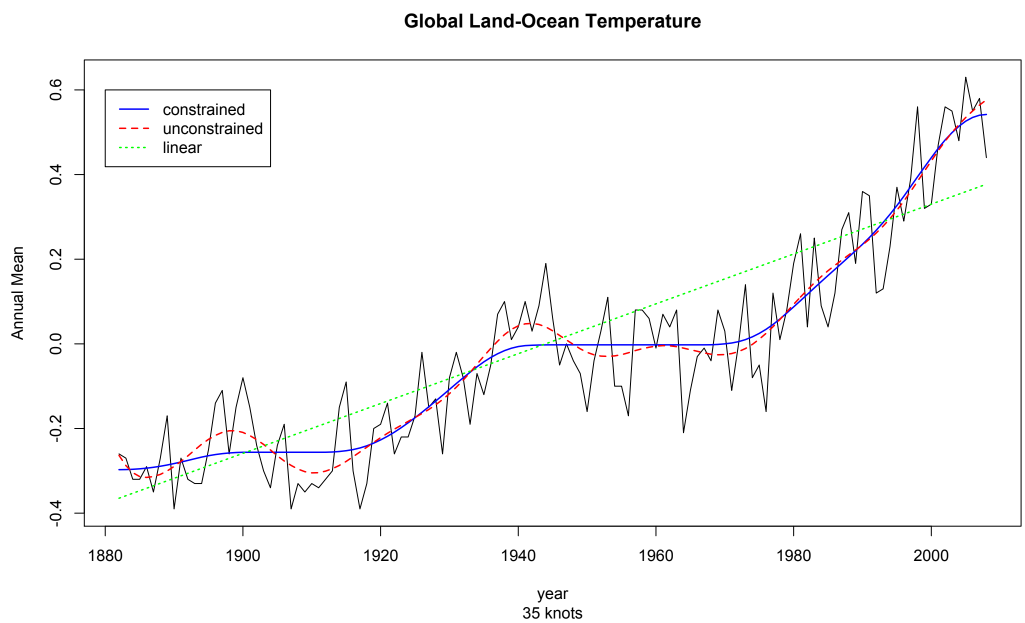

Figure 2.2: Estimated global temperature trends, using constrained penalized spline estimators and unconstrained penalized spline estimator both with 35 knots and penalty and correlation order selected with AIC, and linear regression fit.

trend, such as, constant vs. monotone, in the presence of AR(p) errors. The asymptotic distribution of the test statistic depends on consistent estimation of the correlation.

CHAPTER 3

SHAPE RESTRICTED HYPOTHESIS TESTS IN THE PRESENCE OF CORRELATION

3.1 Literature Review

A test for trend in time series data uses the constant function for the null hypothesis; an alternative hypothesis must be chosen to reflect the context of the problem and provide good power. Often, a linear function is chosen for alternative. Woodward and Gray (1993) perform a test of the existence of a increasing linear trend in global warming data with autoregressive moving average errors.

Raubertas et al. (1986) study the likelihood ratio tests for a set of normal means when the homogeneous linear inequality constraints are imposed on the means for alternative hypotheses using the properties of polyhedral cones. The exact distribution of the likelihood ratio statistic is shown to be the chi-bar-square form. Chapter 2 of Robertson et al. (1988) develops a method for testing whether a regression function is constant or monotone for independent normal data. The alternative estimator is the ordinary unsmoothed isotonic regression. They generate a likelihood ratio test statistic with its exact null distribution that of a mixture of chi-square random variables when the variance is known and a mixture of beta distribution when the variance is unknown, where the mixing distribution is just derived for small sample size. Cohen et al. (1993) offer sufficient conditions for a normal order restricted hypothesis test to be unbiased. The tests include testing homogeneity versus simple order and testing whether the means lie on a line against a convex curve. Cohen et al. (1995) offer sufficient conditions for the tests with order-restricted alternative to be complete and the unbiasedness of tests. The conditions are expressed in terms of cone order monotonicity. They also give a method to construct unbiased and complete hypothesis tests. Meyer (2003) extends the Robertson et al. (1988)'s test for linear versus unsmoothed convex regression function for independent normal data and show that the mixing distribution determined by the convex shape restriction can be calculated from relative volumes of polyhedral cones. Meyer (2008) develops the tests of constant versus increasing and linear versus convex function with smoothed spline regression alternative for independent normal data, which have higher power than the standard version using ordinary shape-restricted regression.

A test for monotone trend with stationary autocorrelated error series is given by Brillinger (1989). The statistic is the ratio of a linear combination of the time series values to an estimate of the standard error of the linear combination. The statistic is shown to have asymptotic power 1 for a broad class of monotonic alternatives. Wu et al. (2001) proposes a test for monotonic trend in short range dependent sequences based on unsmoothed isotonic regression and used this test for changepoint problems. The test is shown to be more powerful than Brillinger (1989)'s test.

For smoothed shape-restricted regression, Ramsay (1998) develop a method to estimate smooth strictly monotone function of arbitrary flexibility. Mukerjee (1988) combines isotonic regression and nonparametric smoothing and generates a hybrid procedure by "isotonizing" the raw data and then smoothing the resulting isotonic regression function by a appropriate kernel. Tantiyaswasdikul and Woodroofe (1994) generalize the isotonic estimate into a piecewise linear isotonic smoothing spline for the estimation of a smooth regression function. Mammen and Thomas-Agnan (1999) prove that the monotone smoothing splines have a optimal convergence rate as $n^{-p/(2p+1)}$, where p is the maximum of the order of spline and the order of derivative. Hall and Huang (2001) develop a method based on maximizing fidelity to the conventional empirical approach subject to monotonicity for monotonizing general linear kernel-typed estimators. Dette et al. (2006) propose a inversion-based method to estimate the monotone regression function. This approach does not require constrained optimization. Dette and Scheder (2006) propose a monotone nonparametric estimate for a regression function of two or more variables. They apply successively one-dimensional isotonization procedures on an initial unconstrained nonparametric regression estimate. Birke and Dette (2006) develop a nonparametric estimate of a convex regression function. The method starts with an unconstrained estimate of the first derivative of the regression function, which is then isotonized and integrated. Meyer (2008) gives details about shape-restricted spline regression and hypothesis tests. The author also proposes an algorithm for the cubic monotone case, and extends the method to convex constraints and variants such as increasing-concave. Wang and Shen (2010) study two monotone univariate regression estimators: a grouped Brunk estimator and a penalized monotone estimator. These estimators are showed to be consistent at the boundary and their mean square errors have optimal convergence rates under some suitable conditions. Wang et al. (2013) propose the constrained spline regression estimator for data with stationary AR(p) errors. The convergence rate for the proposed estimator attains that of the corresponding unconstrained spline regression estimator. This constrained spline regression is used as the alternative fit in this paper.

In this article, we propose several types of hypothesis tests using cone projection theories and algorithms. The observations are serially correlated, having a stationary AR(p) process with unknown p and unknown correlation parameters. The true trend is assumed to be smooth and to have a shape such as monotone or convex. The proposed tests making use of theories of cone projection for both independent data and correlated data are derived in Section 2. In Section 3, the approximate null distributions for the test statistics with unknown p and unknown correlation are investigated. In Section 4, simulations of comparing test size and power of the proposed test with the F-test with unconstrained alternative fit and one-sided t-test with simple regression alternative fit are conducted. In Section 5, we apply the proposed tests to test the monotonicity of Argentina rainfall data and the convexity of price of liquefied U.S. natural gas exports data.

3.2 The Proposed Tests

We assume that the observed data (x_i, y_i), for $i = 1, \cdots, n$, are generated by a smooth trend f with some correlated noise. Let

$$y_i = f(x_i) + \sigma\varepsilon_i, \tag{2}$$

where $x_i \in [0, 1]$, equally spaced. The error term $\boldsymbol{\varepsilon}$ comes from an $AR(p)$ process with zero mean, that is, for some integer $p \geq 1$,

$$\varepsilon_i = \sum_{j=1}^{p} \eta_j \varepsilon_{i-j} + z_i,$$

where z_i, for $i = 1, \cdots, n$ are independent standard normal variables.

Researchers usually want to check if the trend f is constant or has some sort of pattern, such as monotonicity or convexity. A simple and widely used approach to deal with the test of constant versus monotone is the one-sided t-test. If a linear relation between $\mathbf{y}$ and $\mathbf{x}$ is reasonable, we can fit the data with simple linear regression and fit the errors with time series models, such as autoregressive models. Then test whether the coefficient is positive or not by the one-sided t-test. But, often prior information of the relation between $\mathbf{y}$ and $\mathbf{x}$ does not include a parametric form to model with, instead that a vague assumption is more appropriate, such as "smooth", "monotone" and/or "convex". We need to switch to nonparametric approach. But because of the flexibility of the nonparametric approach, the fit rarely satisfies the shape constraint. In this article, we apply the cone projection theory into the regression spline to generate the constrained spline estimate as the alternative fit of the proposed tests. The constraint types includes monotone, convex/concave and combinations of them, such as, increase and convex, etc.

3.2.1 Review of Independent Normal Errors Case

Instead of projections onto the linear space spanned by the columns of design matrix in ordinary least-squares estimation for linear model, constrained estimation involves projections onto the convex cone determined by the shape restriction. Regression splines are used to estimate $f(x)$ in (2). Specify a set of knots $0 = t_1 < \cdots < t_k = 1$. A set of $m = k + d - 1$ basis functions $b_1(x), \cdots, b_m(x)$ are defined, where d is degree of polynomial spline with $d = 2$ for quadratic splines and $d = 3$ for cubic splines. We choose the standard B-spline basis, but other bases spanning the same space can be used instead. Let $\mathbf{b}_1, \cdots, \mathbf{b}_m \in \mathbb{R}^n$ be the basis vectors, where $b_{ij} = b_j(x_i)$. The basis functions span the space of smooth piecewise polynomial regression functions with the given knots, and a m-dimensional subspace of $\mathbb{R}^n$ is spanned by the basis vectors.

First, assume $\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{'}) = \mathbf{I}$, the identity matrix. Let $\mathbf{B}$ be the $n \times m$ matrix with $\mathbf{b}_j$ as columns and let $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_m)^{'}$ be the coefficient vector of spline bases. So, $\mathbf{B}\boldsymbol{\alpha}$ is the best approximation of the true mean vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)^{'}$, where $\theta_i = f(x_i)$ in the space spanned by the spline bases. The vector expression of sum of squared errors is equals to

$$\psi(\boldsymbol{\alpha}; \mathbf{y}) = \boldsymbol{\alpha}^{'}(\mathbf{B}^{'}\mathbf{B})\boldsymbol{\alpha} - 2\mathbf{y}^{'}\mathbf{B}\boldsymbol{\alpha}. \tag{3}$$

For monotone constraint, quadratic splines are used. We define the $k \times m$ matrix $\mathbf{T}$ as the slopes at each knot by $T_{ij} = \partial b_j(t)/\partial t|_{t=t_i}$, for $i = 1, \cdots, k; j = 1, \cdots, m; m = k + 1$. The linear combination $\sum_{j=1}^{m} \alpha_j b_j(x)$ is non-decreasing if and only if its first-order derivative is non-negative. Therefore, the vector of coefficients must satisfy the following constraint:

$$\mathcal{C}_\alpha = \{\boldsymbol{\alpha} : \mathbf{T}\boldsymbol{\alpha} \geq \mathbf{0}\} \subseteq \mathbb{R}^m. \tag{4}$$

The first derivatives of the quadratic spline functions are still piecewise linear between every two consecutive knots, therefore, the constraint on each knot is sufficient to constrain the fit everywhere.

For convex or concave constraints, we use cubic splines and define $T_{ij} = \partial^2 b_j(t)/\partial t^2|_{t=t_i}$, for $i = 1, \cdots, k; j = 1, \cdots, m; m = k + 2$. Then the linear combination $\sum_{j=1}^{m} \alpha_j b_j(x)$ is convex if and only if its second order derivative is non-negative, that is $\mathbf{T}\boldsymbol{\alpha} \geq \mathbf{0}$.

For combinations of monotone and convex/concave, cubic splines are used. Both the first-order and second-order derivatives are needed. We define $\mathbf{T}$ matrix for each different types of combinations, so that the linear combination $\sum_{j=1}^{m} \alpha_j b_j(x)$ satisfies each constraint if and only if the coefficients vector satisfies the linear inequality (4) with corresponding definition of $\mathbf{T}$:

- increasing and convex: $T_{ij} = \frac{\partial^2 b_j(t)}{\partial t^2}|_{t=t_i}, i = 1, \cdots, k; j = 1, \cdots, m; m = k + 2$ and $T_{ij} = \frac{\partial b_j(t)}{\partial t}|_{t=t_1}, i = k + 1; j = 1, \cdots, m; m = k + 2$;
- increasing and concave: $T_{ij} = -\frac{\partial^2 b_j(t)}{\partial t^2}|_{t=t_i}, i = 1, \cdots, k; j = 1, \cdots, m; m = k + 2$ and $T_{ij} = \frac{\partial b_j(t)}{\partial t}|_{t=t_k}, i = k + 1; j = 1, \cdots, m; m = k + 2$;
- decreasing and convex: $T_{ij} = \frac{\partial^2 b_j(t)}{\partial t^2}|_{t=t_i}, i = 1, \cdots, k; j = 1, \cdots, m; m = k + 2$ and $T_{ij} = -\frac{\partial b_j(t)}{\partial t}|_{t=t_k}, i = k + 1; j = 1, \cdots, m; m = k + 2$;
- decreasing and concave: $T_{ij} = -\frac{\partial^2 b_j(t)}{\partial t^2}|_{t=t_i}, i = 1, \cdots, k; j = 1, \cdots, m; m = k + 2$ and $T_{ij} = -\frac{\partial b_j(t)}{\partial t}|_{t=t_1}, i = k + 1; j = 1, \cdots, m; m = k + 2$.

Apply the Cholesky decomposition to $\mathbf{B}'\mathbf{B}$, where $\mathbf{L}\mathbf{L}' = \mathbf{B}'\mathbf{B}$. Define $\boldsymbol{\phi} = \mathbf{L}'\boldsymbol{\alpha}, \mathbf{z} = \mathbf{L}^{-1}\mathbf{B}'\mathbf{y}$, then the expression (3) is transformed into

$$\psi(\boldsymbol{\phi}; \mathbf{z}) = \parallel \boldsymbol{\phi} - \mathbf{z} \parallel^2, \tag{5}$$

and the new constraint set is

$$\mathcal{C}_\phi = \{\boldsymbol{\phi} : \mathbf{A}\boldsymbol{\phi} \geq \mathbf{0}\} \subseteq \mathbb{R}^m, \tag{6}$$

which is a polyhedral cone defined by the constraint matrix $\mathbf{A} = \mathbf{T}(\mathbf{L}')^{-1}$.

Now the problem has been expressed as a cone projection problem: minimize Equation (5) over the cone (6). The definitions of constraint cone, polar cone, faces of the cone, and sectors need to be reviewed, see Meyer (2013) and Silvapulle and Sen (2004b).

The constraint matrix is a $k^* \times m$, full row-rank matrix with $k^* = k, m = k + 1$ for monotone case, $k^* = k, m = k + 2$ for convex/concave case and $k^* = k + 1, m = k + 2$ for the four types of combination. Let $\boldsymbol{\nu}_1, \cdots, \boldsymbol{\nu}_{m-k^*}$ span the null space $\mathcal{V}$ of $\mathbf{A}$. Let $\mathbf{A}_+$ be the square, nonsingular matrix with the rows of $\mathbf{A}$ as first k^* rows and $\boldsymbol{\nu}$ vectors as the rest rows. The first k^* columns $\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_{k^*}$ of $\mathbf{A}_+^{-1}$ are the generators of the cone, therefore the constraint set (6) can be written as

$$\mathcal{C}_\phi = \left\{ \boldsymbol{\phi} : \boldsymbol{\phi} = \sum_{i=1}^{m-k^*} \beta_i \boldsymbol{\nu}_i + \sum_{j=1}^{k^*} \beta_j \boldsymbol{\delta}_j, \quad \boldsymbol{\nu}_i \in \mathcal{V}, \quad \beta_j \geq 0, j = 1, \cdots, k^* \right\}.$$

Let $\Omega = \mathcal{C}_\phi \cap \mathcal{V}^\perp$, then Ω is a polyhedral convex cone that does not contain any linear space. It is the set of nonnegative linear combinations of the $\boldsymbol{\delta}$ vectors. Therefore, Ω can be expressed as,

$$\Omega = \left\{ \boldsymbol{\phi} : \boldsymbol{\phi} = \sum_{j=1}^{m} \beta_j \boldsymbol{\delta}_j, \quad \beta_j \geq 0, j = 1, \cdots, m \right\}.$$

The polar cone is defined as all vectors in the linear space spanned by spline basis making obtuse angles with all vectors in constraint set $\mathcal{C}_\phi$:

$$\Omega_0 = \left\{ \boldsymbol{\rho} : \langle \boldsymbol{\rho}, \boldsymbol{\phi} \rangle \leq 0, \text{for all } \boldsymbol{\phi} \in \mathcal{C}_\phi \right\}. \tag{7}$$

Let $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_m$ are the rows of the matrix $[-(\Delta^{'}\Delta)^{-1}\Delta^{'}]$, where the columns of Δ are the δ_j. Then by the Proposition 3 in Meyer (2008), the vectors $\boldsymbol{\gamma}_j$ are the generators of Ω_0, that is,

$$\Omega_0 = \left\{ \boldsymbol{\rho} : \boldsymbol{\rho} = \sum_{j=1}^{m} \beta_j \boldsymbol{\gamma}_j, \beta_j \geq 0; j = 1, \cdots, m \right\}. \tag{8}$$

We could express the proposed tests into a cone projection problem:

$$\mathbf{H}_0 : \boldsymbol{\phi} \in \mathcal{V} \quad vs. \quad \mathbf{H}_1 : \boldsymbol{\phi} \in \mathcal{C}_\phi.$$

Let $\widehat{\boldsymbol{\phi}}^0$ be the null fit, where $\widehat{\boldsymbol{\phi}}^0 = \Pi(\mathbf{z}|\mathcal{V})$, the projection of $\mathbf{z}$ onto $\mathcal{V}$. For convex/concave case, the null fit is the simple linear regression estimate. Denote the alternative fit as $\widehat{\boldsymbol{\phi}}^1$, where $\widehat{\boldsymbol{\phi}}^1 = \Pi(\mathbf{z}|\mathcal{C}_\phi)$, the projection of $\mathbf{z}$ onto $\mathcal{C}_\phi$. The estimator $\widehat{\boldsymbol{\phi}}^1$ lands on a face of the cone. The faces of the cone are indexed by the collection of sets $J \subseteq \{1, \ldots, m\}$, and defined as

$$\mathcal{F}_J = \left\{ \boldsymbol{\phi} : \boldsymbol{\phi} = \sum_{i=1}^{m-k^*} \beta_i \boldsymbol{\nu}_i + \sum_{j \in J} \beta_j \boldsymbol{\delta}_j, \quad \boldsymbol{\nu}_i \in \mathcal{V}, \quad \beta_j > 0, j \in J \right\}.$$

There are 2^m faces partitioning $\mathcal{C}_\phi$. The interior of the cone is the face with $J = \{1, \ldots, m\}$, and the origin is the face with $J = \emptyset$. Algorithms on cone projections can be used to determine the face $\mathcal{F}_J$ on which the projection falls. Fraser and Massam (1989) propose the mixed primal-dual bases algorithm for cone projections and use it for concave non-parametric regression. Chapter 23 of Lawson and Hanson (1995) states the non-negative least-squares algorithm to solve the quadratic problems with linear inequality constraints. Meyer (2013) develops the Hinge Algorithm for quadratic programming and states its applications in statistics. Then the null fit of the data for the independent case is $\widehat{\mathbf{y}}_{\mathbf{I}}^0 = \mathbf{B}(\mathbf{L}')^{-1}\widehat{\boldsymbol{\phi}}^0$ and the alternative fit is $\widehat{\mathbf{y}}_{\mathbf{I}}^1 = \mathbf{B}(\mathbf{L}')^{-1}\widehat{\boldsymbol{\phi}}^1$.

Let the corresponding sum of squared errors under null hypothesis and alternative hypothesis be $SSE_I^0 = \|\mathbf{y} - \widehat{\mathbf{y}}_{\mathbf{I}}^0\|^2$ and $SSE_I^1 = \|\mathbf{y} - \widehat{\mathbf{y}}_{\mathbf{I}}^1\|$. By Meyer (2008), a likelihood ratio test statistic for known σ^2 is

$$\chi_I^2 = \frac{1}{\sigma^2}(SSE_I^0 - SSE_I^1).$$

If the σ^2 is unknown, the test statistic is

$$B_I = \frac{\chi_I^2}{\chi_I^2 + SSE_I^1/\sigma^2} = \frac{SSE_I^0 - SSE_I^1}{SSE_I^0}.$$

To derive the null distribution of χ_I^2, we need the definition of sectors $\mathbf{C_J}$. A similar definition is developed by Meyer (2008). For all $\mathbf{J} \subseteq \{1, \cdots, m\}$, we have

$$\mathbf{C_J} = \{\mathbf{z} : \mathbf{z} = \boldsymbol{\nu} + \sum_{j \in \mathbf{J}} \beta_j \boldsymbol{\delta}_j + \sum_{j \notin \mathbf{J}} \beta_j \boldsymbol{\gamma}_j\}, \tag{9}$$

where $\boldsymbol{\nu} \in \mathcal{V}$; $\beta_j > 0$ for $j \in \mathbf{J}$, $\beta_j \geq 0$ for $j \notin \mathbf{J}$. The 2^m sectors are disjoint and cover $\mathbb{R}^m$, and

$$\mathbf{P}(\chi_I^2 \leq a) = \sum_{subsets \mathbf{J}} \mathbf{P}(\chi_I^2 \leq a | \mathbf{z} \in \mathbf{C_J}) \mathbf{P}(\mathbf{z} \in \mathbf{C_J}).$$

Meyer (2003, 2008) discuss this and showed that under $\mathbf{H}_0$, $\mathbf{P}(\chi_I^2 \leq a | \mathbf{z} \in \mathbf{C_J}) = \mathbf{P}(\chi^2(d) \leq a)$, where $\chi_I^2(k)$ is a χ^2 random variable with k degrees of freedom and d is the number of indices in $\mathbf{J}$. Therefore, given the number of chosen edges of the convex cone, test statistic χ_I^2 has a χ^2-distribution. Since the number of chosen edges of the cone is a random quantity, the null distribution of the test statistic χ_I^2 has a mixture of χ^2-distribution. Let D be random variable indicating the size of $\mathbf{J}$, the number of generators corresponding to the sector in which $\mathbf{z}$ falls.

$$\mathbf{P}(\chi_I^2 \le a) = \sum_{d=0}^{m} \mathbf{P}(\chi^2(d) \le a)\mathbf{P}(D = d).$$

Since χ_I^2 and SSE_I^1 are independent, we have

$$\mathbf{P}(B_I^2 \le a) = \sum_{d=0}^{m} \mathbf{P}\left[Beta\left(\frac{d}{2}, \frac{n-d-r}{2}\right) \le a\right]\mathbf{P}(D = d),$$

where $\chi^2(0) \equiv 0$, and $Beta(\alpha, \beta)$ is a $Beta$ random variable with parameters α and β. Let $Beta(0, \beta) \equiv 0$ and $Beta(\alpha, 0) \equiv 1$. The mixing probabilities $\mathbf{P}(D = d)$ can be determined by numerical computation under $\mathbf{H}_0$ as precisely as desired. So, the distribution of the test statistic is exact, when errors are independent.

3.2.2 Stationary AR(p) Errors: Known Correlation

If the real data do not behave independently, there is the need to develop an hypothesis test for correlated data. In this article, the correlation is restricted to be stationary AR(p). Let $cor(\boldsymbol{\varepsilon}) = \mathbf{R}$, and first suppose $\mathbf{R}$ is known. Let $\mathbf{R} = \mathbf{W}_R\mathbf{W}_R^{'}$, where $\mathbf{W}_R$ is the Cholesky decomposition of $\mathbf{R}$. Applying the weighted least-squares method, so that $\mathbf{y}_R = \mathbf{W}_R^{-1}\mathbf{y}, \mathbf{B}_R = \mathbf{W}_R^{-1}\mathbf{B}$, and $\boldsymbol{\varepsilon}_R = \mathbf{W}_R^{-1}\boldsymbol{\varepsilon}$, which has correlation matrix $\mathbf{I}$. The expression to be minimized becomes

$$\psi(\boldsymbol{\alpha}; \mathbf{y}_R) = \boldsymbol{\alpha}^{'}(\mathbf{B}_R^{'}\mathbf{B}_R)\boldsymbol{\alpha} - 2\mathbf{y}_R^{'}\mathbf{B}_R\boldsymbol{\alpha}. \tag{10}$$

Let $\mathbf{L}_R$ be the Cholesky decomposition of $\mathbf{B}_R^{'}\mathbf{B}_R$, then $\mathbf{L}_R\mathbf{L}_R^{'} = \mathbf{B}_R^{'}\mathbf{B}_R$. Define $\boldsymbol{\phi}_R = \mathbf{L}_R^{'}\boldsymbol{\alpha}, \mathbf{z}_R = \mathbf{L}_R^{-1}\mathbf{B}_R^{'}\mathbf{y}_R$, then the expression (10) equals to

$$\psi(\boldsymbol{\phi}_R; \mathbf{z}_R) = \parallel \boldsymbol{\phi}_R - \mathbf{z}_R \parallel^2,$$

where $\boldsymbol{\phi}_R$ is restricted to $\mathcal{C}_\phi^R = \{\boldsymbol{\phi}_R : \mathbf{A}_R\boldsymbol{\phi}_R \ge \mathbf{0}\} \subseteq \mathbb{R}^m$, and $\mathbf{A}_R = \mathbf{T}(\mathbf{L}_R^{'})^{-1}$. With the transformation of the data and basis, the constraint matrix is also transformed into $\mathbf{A}_R$, therefore the polyhedral cone is also transformed. The construction of transformed cone is similar to the Section (3.2.1). Let $\boldsymbol{\nu}_1^R, \cdots, \boldsymbol{\nu}_{m-k^*}^R$ span the null space $\mathcal{V}^R$ of $\mathbf{A}^R$, and let $\mathbf{A}_+^R$ be the square, nonsingular matrix with the rows of $\mathbf{A}^R$ as first k^* rows and $\boldsymbol{\nu}^R$ vectors as the last rows. The first k^* columns of $(\mathbf{A}_+^R)^{-1}$ are the edges $\boldsymbol{\delta}_1^R, \cdots, \boldsymbol{\delta}_{k^*}^R$ of the cone, therefore the transformed constrained set can be written as

$$\mathcal{C}_\phi^R = \left\{ \boldsymbol{\phi} : \boldsymbol{\phi} = \boldsymbol{\nu}^R + \sum_{j=1}^{k^*} \beta_j \boldsymbol{\delta}_j^R, \quad \boldsymbol{\nu} \in \mathcal{V}^R, \quad \beta_j \geq 0, j = 1, \cdots, k^* \right\}.$$

The transformed cone $\Omega^R = \mathcal{C}_\phi^R \cap (\mathcal{V}^R)^\perp$, and is expressed as,

$$\Omega^R = \left\{ \boldsymbol{\phi} : \boldsymbol{\phi} = \sum_{j=1}^{m} \beta_j \boldsymbol{\delta}_j^R, \quad \beta_j \geq 0, j = 1, \cdots, m \right\}.$$

The transformed polar cone is constructed as the independent case, and expressed as

$$\Omega_0^R = \left\{ \boldsymbol{\rho} : \boldsymbol{\rho} = \sum_{j=1}^{m} \beta_j \boldsymbol{\gamma}_j^R, \beta_j \geq 0; j = 1, \cdots, m \right\}. \quad (11)$$

The minimizer $\widehat{\boldsymbol{\phi}}_R^1$ is the projection of $\mathbf{z}_R$ onto the transformed cone $\mathcal{C}_\phi^R$ and lands on a face of the transformed cone. The 2^m faces are defined by

$$\mathcal{F}_J^R = \left\{ \boldsymbol{\phi} : \boldsymbol{\phi} = \boldsymbol{\nu}^R + \sum_{j \in J} \beta_j \boldsymbol{\delta}_j^R, \quad \boldsymbol{\nu} \in \mathcal{V}^R, \quad \beta_j > 0, j \in J \right\}.$$

Wang et al. (2013) give the details about the projection of the transformed data onto the transformed cone when the correlation is known.

Let the minimizer of sum of squared errors under null and alternative hypothesis be $\widehat{\boldsymbol{\alpha}}_R^0$ and $\widehat{\boldsymbol{\alpha}}_R^1$ for known correlation $\mathbf{R}$, where $\widehat{\boldsymbol{\alpha}}_R^0 = (\mathbf{L}_R^{'})^{-1}\widehat{\boldsymbol{\phi}}_R^0$ and $\widehat{\boldsymbol{\alpha}}_R^1 = (\mathbf{L}_R^{'})^{-1}\widehat{\boldsymbol{\phi}}_R^1$. Therefore the null fit and alternative fit are $\widehat{\mathbf{y}}_R^0 = \mathbf{B}\widehat{\boldsymbol{\alpha}}_R^0$ and $\widehat{\mathbf{y}}_R^1 = \mathbf{B}\widehat{\boldsymbol{\alpha}}_R^1$. Let $\widehat{\boldsymbol{\theta}}_R^0 = \mathbf{B}_R\widehat{\boldsymbol{\alpha}}_R^0$ and $\widehat{\boldsymbol{\theta}}_R^1 = \mathbf{B}_R\widehat{\boldsymbol{\alpha}}_R^1$. Then $SSE_R^0 = \|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}_R^0\|^2$ and $SSE_R^1 = \|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}_R^1\|^2$. The likelihood ratio test statistic for known σ^2 is,

$$\chi_R^2 = \frac{1}{\sigma^2}\left(SSE_R^0 - SSE_R^1\right)$$

Define the transformed sector $\mathbf{C}_\mathbf{J}^R$ for all $\mathbf{J} \subseteq \{1, \cdots, m\}$ as

$$\mathbf{C}_\mathbf{J}^R = \{\mathbf{z}_R : \mathbf{z}_R = \boldsymbol{\nu}^R + \sum_{j\in\mathbf{J}} \beta_j \boldsymbol{\delta}_j^R + \sum_{j\notin\mathbf{J}} \beta_j \boldsymbol{\gamma}_j^R\},$$

It is easy to prove that χ_R^2 still has an exact distribution as a mixture of χ^2 random variables, that is,

$$\mathbf{P}(\chi_R^2 \leq a) = \sum_{subsets\mathbf{J}} \mathbf{P}(\chi_R^2 \leq a|\mathbf{z}_R \in \mathbf{C}_\mathbf{J}^R)\mathbf{P}(\mathbf{z}_R \in \mathbf{C}_\mathbf{J}^R) = \sum_{d=0}^{m} \mathbf{P}(\chi^2(d) \leq a)\mathbf{P}(\tilde{D} = d).$$

The likelihood ratio test statistic, if σ^2 is unknown, is

$$B_R = \frac{\chi_R^2}{\chi_R^2 + SSE_R^1/\sigma^2} = \frac{SSE_R^0 - SSE_R^1}{SSE_R^0};$$

and

$$\mathbf{P}(B_R \leq a) = \sum_{d=0}^{m} \mathbf{P}\left[Beta\left(\frac{d}{2}, \frac{n-d-r}{2}\right) \leq a\right]\mathbf{P}(\tilde{D} = d).$$

where, $\tilde{D}$ is a random variable, indicating the number of edges of the transformed cone that the projection algorithm identifies.

3.2.3 Test Statistics with Unknown Correlation

Often the correlation parameters are unknown and we need make inference based on the estimate of the correlation. Let $\widehat{\mathbf{R}}$ be an estimator of $\mathbf{R}$, and $\mathbf{W}_{\widehat{R}}\mathbf{W}_{\widehat{R}}'$ is the Cholesky decomposition of $\widehat{\mathbf{R}}$.

Transforming the data and basis with $\widehat{\mathbf{R}}$, we have $\mathbf{y}_{\widehat{R}} = \mathbf{W}_{\widehat{R}}^{-1}\mathbf{y}, \mathbf{B}_{\widehat{R}} = \mathbf{W}_{\widehat{R}}^{-1}\mathbf{B}$, $\boldsymbol{\varepsilon}_{\widehat{R}} = \mathbf{W}_{\widehat{R}}^{-1}\boldsymbol{\varepsilon}$, where $\mathbf{E}(\boldsymbol{\varepsilon}_{\widehat{R}}\boldsymbol{\varepsilon}_{\widehat{R}}^{'}) = \mathbf{W}_{\widehat{R}}^{-1}\widehat{\mathbf{R}}(\mathbf{W}_{\widehat{R}}^{-1})'$. The transformed sum of squared errors is

$$\psi(\boldsymbol{\alpha};\mathbf{y}_{\widehat{R}}) = \boldsymbol{\alpha}^{'}(\mathbf{B}_{\widehat{R}}^{'}\mathbf{B}_{\widehat{R}})\boldsymbol{\alpha} - 2\mathbf{y}_{\widehat{R}}^{'}\mathbf{B}_{\widehat{R}}\boldsymbol{\alpha}. \tag{12}$$

Let $\mathbf{L}_{\widehat{R}}$ be the Cholesky decomposition of $\mathbf{B}_{\widehat{R}}^{'}\mathbf{B}_{\widehat{R}}$, where $\mathbf{L}_{\widehat{R}}\mathbf{L}_{\widehat{R}}^{'} = \mathbf{B}_{\widehat{R}}^{'}\mathbf{B}_{\widehat{R}}$. Define $\boldsymbol{\phi}_{\widehat{R}} = \mathbf{L}_{\widehat{R}}^{'}\boldsymbol{\alpha}, \mathbf{z}_{\widehat{R}} = \mathbf{L}_{\widehat{R}}^{-1}\mathbf{B}_{\widehat{R}}^{'}\mathbf{y}_{\widehat{R}}$. We have

$$\psi(\boldsymbol{\alpha};\mathbf{y}_{\widehat{R}}) = \psi(\boldsymbol{\phi}_{\widehat{R}};\mathbf{z}_{\widehat{R}}) = \parallel \boldsymbol{\phi}_{\widehat{R}} - \mathbf{z}_{\widehat{R}} \parallel^2,$$

where $\boldsymbol{\phi}_{\widehat{R}}$ is restricted to $\mathcal{C}_{\phi}^{\widehat{R}} = \{\boldsymbol{\phi}_{\widehat{R}} : \mathbf{A}_{\widehat{R}}\boldsymbol{\phi}_{\widehat{R}} \geq \mathbf{0}\} \subseteq \mathbb{R}^m$, and $\mathbf{A}_{\widehat{R}} = \mathbf{T}(\mathbf{L}_{\widehat{R}}^{'})^{-1}$.

Let $\tilde{\nu}_1^{\widehat{R}}, \cdots, \tilde{\nu}_{m-k^*}^{\widehat{R}}$ span the null space $\tilde{V}_{\widehat{R}}$ of the constraint matrix $\mathbf{A}_{\widehat{R}}$. Let $\mathbf{A}_{\widehat{R}}^{+}$ be the square, nonsingular matrix with rows of $\mathbf{A}_{\widehat{R}}$ as first k^* rows and $\boldsymbol{\nu}_{\widehat{R}}$ vectors as the last rows. The first k^* columns of $(\mathbf{A}_{\widehat{R}}^{+})^{-1}$ are the edges $\delta_1^{\widehat{R}}, \cdots, \delta_{k^*}^{\widehat{R}}$ of the cone.

Let $\widehat{\boldsymbol{\alpha}}_{\widehat{R}}^0$ and $\widehat{\boldsymbol{\alpha}}_{\widehat{R}}^1$ be the minimizer of equation (12) under null and alternative hypothesis weighted by the estimator $\widehat{\mathbf{R}}$. Therefore the null fit and alternative fit of data are $\widehat{\mathbf{y}}_{\widehat{R}}^0 = \mathbf{B}\widehat{\boldsymbol{\alpha}}_{\widehat{R}}^0$ and $\widehat{\mathbf{y}}_{\widehat{R}}^1 = \mathbf{B}\widehat{\boldsymbol{\alpha}}_{\widehat{R}}^1$. We use the iteration procedure introduced by Wang et al. (2013) to estimate the alternative fit, where the order p is selected by a correlation-adjusted AIC.

Let $\widehat{\boldsymbol{\varepsilon}}$ be the residual vector of alternative fit, where $\widehat{\varepsilon}_i = y_i - \widehat{y}_{\widehat{R}i}^1$. Let $\gamma_h = \mathbf{E}\varepsilon_i\varepsilon_{i+h}$ and $\widehat{\gamma}_h = \frac{1}{n}\sum_{i=1}^{n-h}\widehat{\varepsilon}_i\widehat{\varepsilon}_{i+h}$, where $i = 1, \ldots, n-h$, and $h = 0, 1, \ldots, n-1$. Since the errors are assumed to come from a segment of AR(p) process, the correlation matrix $\mathbf{R}$ has elements $R_{ij} = \gamma_{|i-j|+1}/\gamma_0$. Let $\widehat{\mathbf{R}}$ be an estimator of $\mathbf{R}$ with each element $\widehat{R}_{ij} = \widehat{\gamma}_{|i-j|+1}/\widehat{\gamma}_0$. So, both $\mathbf{R}$ and $\widehat{\mathbf{R}}$ are symmetric and positive definite.

Let $\widehat{\boldsymbol{\theta}}^0_{\widehat{R}} = \mathbf{B}_{\widehat{R}}\hat{\boldsymbol{\alpha}}^0$ and $\widehat{\boldsymbol{\theta}}^1_{\widehat{R}} = \mathbf{B}_{\widehat{R}}\hat{\boldsymbol{\alpha}}^1$. Then we have $SSE^0_{\widehat{R}} = \|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}\|^2$ and $SSE^1_{\widehat{R}} = \|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^1_{\widehat{R}}\|^2$. The likelihood ratio test statistic for known σ^2 can be expressed as

$$\chi^2_{\widehat{R}} = \frac{1}{\sigma^2}(SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}}). \tag{13}$$

The likelihood ratio test statistic for unknown σ^2 is

$$B_{\widehat{R}} = \frac{\chi^2_{\widehat{R}}}{\chi^2_{\widehat{R}} + SSE^1_{\widehat{R}}/\sigma^2} = \frac{SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}}}{SSE^0_{\widehat{R}}}. \tag{14}$$

In the next section, we derive the limiting distribution of the test statistics when correlation is unknown.

3.3 Approximate Null Distributions of Test Statistics with Unknown Correlation

In this context, we use $\|\mathbf{M}\| \equiv \max\{\sqrt{\lambda} : \lambda \text{ is an eigenvalue of } \mathbf{M}'\mathbf{M}\}$, for any arbitrary $n \times m$ matrix $\mathbf{M}$ with M_{ij} as its element, where $i = 1, \cdots, n; j = 1, \cdots, m$.

Lemma 1. *If* $\|\widehat{\mathbf{R}} - \mathbf{R}\| \xrightarrow{p} 0$*, then* $\|\mathbf{W}_{\widehat{R}} - \mathbf{W}_R\| \xrightarrow{p} 0$*, as* $n \to \infty$*.*

Since both $\mathbf{R}$ and $\widehat{\mathbf{R}}$ are symmetric and positive definite, they have the unique Cholesky decomposition with positive diagonal entries, respectively. Each element of matrix $\mathbf{W}_{\widehat{R}}$ and $\mathbf{W}_R$ is uniquely determined by the corresponding elements of $\widehat{\mathbf{R}}$ and $\mathbf{R}$. It is easy to derive that $\|\mathbf{W}_{\widehat{R}} - \mathbf{W}_R\| \xrightarrow{p} 0$, as $n \to \infty$.

Lemma 2. *If* $\|\widehat{\mathbf{R}} - \mathbf{R}\| \xrightarrow{p} 0$*, then* $\|\mathbf{W}^{-1}_{\widehat{R}} - \mathbf{W}^{-1}_R\| \xrightarrow{p} 0$*, as* $n \to \infty$*.*

Proof. First, we prove $\|\mathbf{W}^{-1}_R\|$ and $\|\mathbf{W}^{-1}_{\widehat{R}}\|$ are bounded. By Siddiqui (1958), for $AR(p)$ process, both $\mathbf{R}^{-1}$ and $\widehat{\mathbf{R}}^{-1}$ are banded matrices. Let R^{-1}_{ij} and $\widehat{R}^{-1}_{ij}$ be the element of ith row and jth column of $\mathbf{R}^{-1}$ and $\widehat{\mathbf{R}}^{-1}$, correspondingly. Then

$$R_{ij}^{-1} = 0 \quad \text{and} \quad \widehat{R}_{ij}^{-1} = 0, \quad \text{for } p < |i-j| < n.$$

Since $\mathbf{R}^{-1} = (\mathbf{W}_R^{-1})'\mathbf{W}_R^{-1}$ and $\widehat{\mathbf{R}}^{-1} = (\mathbf{W}_{\widehat{R}}^{-1})'\mathbf{W}_{\widehat{R}}^{-1}$, then by Theorem 4.3.1 in Golub and Van Loan (1996), $\mathbf{W}_R^{-1}$ and $\mathbf{W}_{\widehat{R}}^{-1}$ have the same lower bandwidth as $\mathbf{R}^{-1}$ and $\widehat{\mathbf{R}}^{-1}$. Then we have

$$W_{ij} = 0 \quad \text{and} \quad \widehat{W}_{ij} = 0, \quad \text{for } i-j < 0 \text{ and } i-j > p.$$

We have $R_{ij}^{-1} = \sum_{k=i}^{j+p} W_{ki}W_{kj}$ and $\widehat{R}_{ij}^{-1} = \sum_{k=i}^{j+p} \widehat{W}_{ki}\widehat{W}_{kj}$. By Siddiqui (1958), for $AR(p)$ process, R_{ij}^{-1} and $\widehat{R}_{ij}^{-1}$ are both finite summations of polynomials of η_j and $\widehat{\eta}_j$ correspondingly, therefore both R_{ij}^{-1} and $\widehat{R}_{ij}^{-1}$ are bounded. Then both w_{ij} and $\widehat{W}_{ij}$ are bounded for $0 \le i-j \le p$. Let $\mathbf{M}$ be an arbitrary $n \times n$ matrix. Define $\|\mathbf{M}\|_1 = \max_{1 \le j \le n} \sum_{i=1}^n |M_{ij}|$, the maximum absolute column sum of the matrix. Define $\|\mathbf{M}\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^n |M_{ij}|$, the maximum absolute row sum of the matrix. Then $\|\mathbf{W}_{\widehat{R}}^{-1}\|_1$, $\|\mathbf{W}_{\widehat{R}}^{-1}\|_\infty$, $\|\mathbf{W}_R^{-1}\|_1$, and $\|\mathbf{W}_R^{-1}\|_\infty$ are all bounded, since $\mathbf{W}_{\widehat{R}}^{-1}$ and $\mathbf{W}_R^{-1}$ are both banded matrix and each elements are bounded. By Chapter 5 in Horn and Johnson (1990), $\|\mathbf{M}\| \le \sqrt{\|\mathbf{M}\|_1 \|\mathbf{M}\|_\infty}$. Hence $\|\mathbf{W}_{\widehat{R}}^{-1}\|$ and $\|\mathbf{W}_R^{-1}\|$ are bounded. Applying Lemma (1), we have

$$\|\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1}\| = \|\mathbf{W}_{\widehat{R}}^{-1}(\mathbf{W}_R - \mathbf{W}_{\widehat{R}})\mathbf{W}_R^{-1}\| \le \|\mathbf{W}_{\widehat{R}}^{-1}\| \|\mathbf{W}_R - \mathbf{W}_{\widehat{R}}\| \|\mathbf{W}_R^{-1}\| \xrightarrow{p} 0, \text{as } n \to \infty.$$

□

Lemma 3. *Assume* $\|\widehat{\mathbf{R}} - \mathbf{R}\| \xrightarrow{p} 0$, *as* $n \to \infty$. *Let* $\mathbf{e} = (e_1, \cdots, e_n)'$ *be a multivariate normal random vector with mean zero and covariance matrix* Σ. *Then* $(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_{\widehat{R}}^{-1})\mathbf{e} \xrightarrow{p} 0$, *component-wise, as* $n \to \infty$.

Proof. Let $D_W = \mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1}$. By the proof of Lemma (2), we have

$$d_{ij} = W_{ij} - \widehat{W}_{ij} = 0 \quad \text{for } i - j < 0 \text{ and } i - j > p,$$

where d_{ij} is the element of D_W, for $i, j = 1, \cdots, n$. Then, for $i = 1, \cdots, n$,

$$(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\mathbf{e} = (\sum_{j=1}^{n} d_{1j}e_j, \cdots, \sum_{j=1}^{n} d_{nj}e_j)' = (\sum_{1 \le j \le 1+p} d_{1j}e_j, \cdots, \sum_{n \le j \le n+p} d_{nj}e_j).$$

From Lemma 1, we have $d_{ij} \xrightarrow{p} 0$, as $n \to \infty$, for $i, j = 1, \cdots, n$. By Slutsky's theorem from Slutsky (1925), $d_{ij}e_j \xrightarrow{p} 0$, as $n \to \infty$, for any $i, j = 1, \cdots, n$. So,

$$\sum_{i \le j \le i+p} d_{ij}e_j \xrightarrow{p} 0, \quad \text{as } n \to \infty, \text{ for } i = 1, \cdots, n.$$

Therefore, we have $(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\mathbf{e} \xrightarrow{p} 0$, component-wise, as $n \to \infty$. □

Lemma 4. *If* $\|\widehat{\mathbf{R}} - \mathbf{R}\| \xrightarrow{p} 0$, *then* $\|\mathbf{W}_{\widehat{R}}^{-1}\mathbf{R}(\mathbf{W}_{\widehat{R}}^{-1})' - \mathbf{I}\| \xrightarrow{p} 0$, *as* $n \to \infty$, *where* $\mathbf{I}$ *is the* $n \times n$ *identity matrix.*

Proof. Let $\mathbf{W}_{\widehat{R}}^{-1} = \mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1} + \mathbf{W}_R^{-1}$, then

$$\begin{aligned}
\|\mathbf{W}_{\widehat{R}}^{-1}\mathbf{R}(\mathbf{W}_{\widehat{R}}^{-1})' - \mathbf{I}\| &= \|(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\mathbf{R}(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})' + (\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\mathbf{W}_R + \mathbf{W}'_R(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})'\| \\
&\le \|\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1}\|\|\mathbf{R}\|\|(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})'\| + \\
&\quad \|\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1}\|\|\mathbf{W}_R\| + \|\mathbf{W}'_R\|\|(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})'\|.
\end{aligned}$$

By Section 4.2.3 in Golub and Van Loan (1996), $\|\mathbf{W}_R\|$ and $\|\mathbf{W}_{\widehat{R}}\|$ are both bounded. Applying Lemma (2), $\|\mathbf{W}_{\widehat{R}}^{-1}\mathbf{R}(\mathbf{W}_R^{-1})' - \mathbf{I}\| \xrightarrow{p} 0$, as $n \to \infty$. □

Theorem 8. *If* $\|\widehat{\mathbf{R}} - \mathbf{R}\| \xrightarrow{p} 0$ *as* $n \to \infty$, *then for each* k, *the limiting distribution of test statistic (13) is*

$$\mathbf{P}\left(\chi^2_{\widehat{R}} \le a\right) \xrightarrow{D} \sum_{d=0}^{m} \mathbf{P}(\chi^2(d) \le a)\mathbf{P}(\tilde{D} = d), \text{ as } n \to \infty.$$

Proof. First, we prove that given $\mathbf{z} \in \mathbf{C}_J$ and $\text{card}(J) = d$, then $SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}} | \tilde{D} = d \xrightarrow{D} \chi^2(d)$. It is enough to prove

$$\frac{(SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}}) - (SSE^0_R - SSE^1_R)}{SSE^0_R - SSE^1_R} \xrightarrow{P} 0, \text{as } n \to \infty.$$

Under $\mathbf{H}_0$:

$$\begin{aligned}
SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}} &= \|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^1_{\widehat{R}}\|^2 - \|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^1_{\widehat{R}}\|^2 \\
&= \|\widehat{\boldsymbol{\theta}}^1_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}\|^2 \\
&= \|\mathbf{B}_{\widehat{R}}(\mathbf{L}'_{\widehat{R}})^{-1}\widehat{\boldsymbol{\phi}}^1_{\widehat{R}} - \mathbf{B}_{\widehat{R}}(\mathbf{L}'_{\widehat{R}})^{-1}\widehat{\boldsymbol{\phi}}^0_{\widehat{R}}\|^2 \\
&= \|\widehat{\boldsymbol{\phi}}^1_{\widehat{R}} - \widehat{\boldsymbol{\phi}}^0_{\widehat{R}}\|^2 \\
&= \|\pi(\mathbf{z}_{\widehat{R}} | \mathcal{C}^{\widehat{R}}_\phi) - \pi(\mathbf{z}_{\widehat{R}} | \mathcal{V}^{\widehat{R}})\|^2 \\
&= \|\pi(\mathbf{z}_{\widehat{R}} - \boldsymbol{\phi}_{\widehat{R}} + \boldsymbol{\phi}_{\widehat{R}} | \mathcal{C}^{\widehat{R}}_\phi) - \pi(\mathbf{z}_{\widehat{R}} - \boldsymbol{\phi}_{\widehat{R}} + \boldsymbol{\phi}_{\widehat{R}} | \mathcal{V}^{\widehat{R}})\|^2 \\
&= \|\pi(\mathbf{z}_{\widehat{R}} - \boldsymbol{\phi}_{\widehat{R}} | \mathcal{C}^{\widehat{R}}_\phi) - \pi(\mathbf{z}_{\widehat{R}} - \boldsymbol{\phi}_{\widehat{R}} | \mathcal{V}^{\widehat{R}})\|^2 \\
&= \|\pi(\mathbf{z}_{\widehat{R}} - \boldsymbol{\phi}_{\widehat{R}} | \boldsymbol{\Omega}^{\widehat{R}})\|^2.
\end{aligned}$$

Let $\mathbf{e}_{\widehat{R}} = \mathbf{z}_{\widehat{R}} - \boldsymbol{\phi}_{\widehat{R}}$ and $\mathbf{f}_{\widehat{R}} = \widehat{\mathbf{W}}^{-1}_{\widehat{R}}\mathbf{f}$, then

$$\begin{aligned}
\mathbf{e}_{\widehat{R}} &= \mathbf{L}^{-1}_{\widehat{R}}\mathbf{B}'_{\widehat{R}}\mathbf{y}_{\widehat{R}} - \mathbf{L}'_{\widehat{R}}\boldsymbol{\alpha} \\
&= \mathbf{L}^{-1}_{\widehat{R}}\mathbf{B}'_{\widehat{R}}(\mathbf{f}_{\widehat{R}} - \mathbf{B}_{\widehat{R}}\boldsymbol{\alpha} + \mathbf{B}_{\widehat{R}}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}_{\widehat{R}})\mathbf{L}'_{\widehat{R}}\boldsymbol{\alpha} \\
&= \mathbf{L}^{-1}_{\widehat{R}}\mathbf{B}'_{\widehat{R}}(\mathbf{f}_{\widehat{R}} - \mathbf{B}_{\widehat{R}}\boldsymbol{\alpha}) + \mathbf{L}^{-1}_{\widehat{R}}\mathbf{B}'_{\widehat{R}}\boldsymbol{\varepsilon}_{\widehat{R}}.
\end{aligned}$$

Under $\mathbf{H}_0$ that $\mathbf{f}$ is constant, $\mathbf{f}_{\widehat{R}} - \mathbf{B}_{\widehat{R}}\boldsymbol{\alpha} = 0$. Let $\{\mathbf{U}^{\widehat{R}}_j\}^d_{j=1}$ be a set of orthonormal bases of convex cone $\boldsymbol{\Omega}^{\widehat{R}}$, then

$$\|\pi(\mathbf{z}_{\widehat{R}} - \boldsymbol{\phi}_{\widehat{R}} | \boldsymbol{\Omega}^{\widehat{R}})\|^2 = \sum_{j=1}^{d} \mathbf{U}^{\widehat{R}'}_j \mathbf{e}_{\widehat{R}} \mathbf{e}'_{\widehat{R}} \mathbf{U}^{\widehat{R}}_j$$

$$= \sum_{j=1}^{d} (\mathbf{U}_j^{\widehat{R}'} \mathbf{L}_{\widehat{R}}^{-1} \mathbf{B}_{\widehat{R}}' \boldsymbol{\varepsilon}_{\widehat{R}})^2.$$

Therefore,

$$\begin{aligned} \frac{(SSE_{\widehat{R}}^0 - SSE_{\widehat{R}}^1) - (SSE_R^0 - SSE_R^1)}{SSE_R^0 - SSE_R^1} &= \frac{\sum_{j=1}^d \left[(\mathbf{U}_j^{\widehat{R}'} \mathbf{e}_{\widehat{R}})^2 - (\mathbf{U}_j^{R'} \mathbf{e}_R)^2 \right]}{SSE_R^0 - SSE_R^1} \\ &= \frac{\sum_{j=1}^d \left[(\mathbf{U}_j^{\widehat{R}'} \mathbf{e}_{\widehat{R}} - \mathbf{U}_j^{R'} \mathbf{e}_R)^2 - 2\mathbf{U}_j^{R'} \mathbf{e}_R (\mathbf{U}_j^{\widehat{R}'} \mathbf{e}_{\widehat{R}} - \mathbf{U}_j^{R'} \mathbf{e}_R) \right]}{SSE_R^0 - SSE_R^1}. \end{aligned}$$

Given $\mathbf{J}$ and card$(J) = d$, $SSE_R^0 - SSE_R^1|_{\tilde{D}=d} \overset{\mathbf{H}_0}{\sim} \chi^2$, and $\mathbf{U}_j^{R'} \mathbf{e}_R|_{\tilde{D}=d} \overset{\mathbf{H}_0}{\sim} N(0,1)$. Therefore, it is enough to prove

$$\mathbf{U}_j^{\widehat{R}'} \mathbf{e}_{\widehat{R}} - \mathbf{U}_j^{R'} \mathbf{e}_R = \mathbf{U}_j^{\widehat{R}'} \mathbf{L}_{\widehat{R}}^{-1} \mathbf{B}_{\widehat{R}}' \boldsymbol{\varepsilon}_{\widehat{R}} - \mathbf{U}_j^{R'} \mathbf{L}_R^{-1} \mathbf{B}_R' \boldsymbol{\varepsilon}_R \xrightarrow{P} 0.$$

The proof of (3.3) is divided into two steps: in step 1, we prove that the difference of each piece that weighted by $\widehat{\mathbf{R}}$ and that weighted by the true $\mathbf{R}$ will convergent to zero; in step 2, we will prove the convergency of the multiplication of those four pieces.

1. Proof of $\|\mathbf{B}_{\widehat{R}} - \mathbf{B}_R\| \xrightarrow{P} 0$, as $n \to \infty$.

   Applying Lemma 2, we have

   $$\|\mathbf{B}_{\widehat{R}} - \mathbf{B}_R\| = \|(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\mathbf{B}\| \leq \|\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1}\| \|\mathbf{B}\| \xrightarrow{P} 0, \text{as } n \to \infty,$$

   since $\|\mathbf{B}\|$ is bounded.

2. Proof of $\|\mathbf{L}_{\widehat{R}} - \mathbf{L}_R\| \xrightarrow{P} 0$, as $n \to \infty$.

   First, we prove,

$$\|\mathbf{L}_{\widehat{R}}\mathbf{L}_{\widehat{R}}^{'} - \mathbf{L}_R\mathbf{L}_R^{'}\| = \|\mathbf{B}^{'}(\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1})\mathbf{B}\| \leq \|\mathbf{B}^{'}\|\|\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}\|\|\mathbf{B}\| \xrightarrow{P} 0, \text{as } n \to \infty.$$

   Since $\|\mathbf{R}^{-1}\|$ and $\|\widehat{\mathbf{R}}^{-1}\|$ are bounded by Siddiqui (1958), we have

$$\|\widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}\| = \|\widehat{\mathbf{R}}^{-1}(\mathbf{R} - \widehat{\mathbf{R}})\mathbf{R}^{-1}\| \leq \|\widehat{\mathbf{R}}^{-1}\|\|\mathbf{R} - \widehat{\mathbf{R}}\|\|\mathbf{R}^{-1}\| \xrightarrow{p} 0, \text{as}\, n \to \infty.$$

   Therefore $\|\mathbf{L}_{\widehat{R}} - \mathbf{L}_R\| \xrightarrow{P} 0$, as $n \to \infty$. Applying Lemma 1, we derive $\|\mathbf{L}_{\widehat{R}} - \mathbf{L}_R\| \xrightarrow{P} 0$, as $n \to \infty$.

3. Proof of $\|\mathbf{U}_j^{\widehat{R}} - \mathbf{U}_j^R\| \xrightarrow{P} 0$, as $n \to \infty$, for $j = 1, \cdots, d$. Let $\mathbf{\Delta}_R$ and $\mathbf{\Delta}_{\widehat{R}}$ be the matrix with $\boldsymbol{\delta}_j^R$ and $\boldsymbol{\delta}_j^{\widehat{R}}$ as the columns, respectively. Since the edges $\boldsymbol{\delta}_j^R$ of the transformed cone $\mathbf{\Omega}_R$ are generated as the first k columns of the $\mathbf{A}_R^+$ and the edges $\boldsymbol{\delta}_j^R$ of the transformed cone $\mathbf{\Omega}^{\widehat{R}}$ are generated as the first k columns of the $\mathbf{A}_{\widehat{R}}^+$, then we have $\mathbf{T}(\mathbf{L}_R^{'})^{-1}\mathbf{\Delta}_R = \mathbf{I}_{k\times k}$ and $\mathbf{T}(\mathbf{L}_{\widehat{R}}^{'})^{-1}\mathbf{\Delta}_{\widehat{R}} = \mathbf{I}_{k\times k}$. Therefore, for any type of constraint matrix $\mathbf{T}$, we have

$$\mathbf{T}\left[(\mathbf{L}_{\widehat{R}}^{'})^{-1}\mathbf{\Delta}_{\widehat{R}} - (\mathbf{L}_R^{'})^{-1}\mathbf{\Delta}_R\right] = 0.$$

   Then $(\mathbf{L}_R^{'})^{-1}(\mathbf{\Delta}_R - \mathbf{\Delta}_{\widehat{R}}) + (\mathbf{L}_R^{-1} - \mathbf{L}_{\widehat{R}}^{-1})\mathbf{\Delta}_{\widehat{R}} = 0$. So,

$$\|\mathbf{\Delta}_R - \mathbf{\Delta}_{\widehat{R}}\| = \|\mathbf{L}_R^{'}(\mathbf{L}_R^{-1} - \mathbf{L}_{\widehat{R}}^{-1})\mathbf{\Delta}_{\widehat{R}}\| \leq \|\mathbf{L}_R^{'}\|\|\mathbf{L}_R^{-1} - \mathbf{L}_{\widehat{R}}^{-1}\|\|\mathbf{\Delta}_{\widehat{R}}\|$$

   By Lemma 2, we derive that $\|\mathbf{\Delta}_R - \mathbf{\Delta}_{\widehat{R}}\| \xrightarrow{P} 0$, as $n \to \infty$, since $\{\mathbf{U}_j^{\widehat{R}}\}_{j=1}^d$ is a linear combination of $\{\boldsymbol{\delta}_j^{\widehat{R}}\}_{j=1}^d$, therefore, $\|\mathbf{U}_j^{\widehat{R}} - \mathbf{U}_j^R\| \xrightarrow{P} 0$, as $n \to \infty$, for $j = 1, \cdots, d$.

4. Proof of $\|\boldsymbol{\varepsilon}_{\widehat{R}} - \boldsymbol{\varepsilon}_R\| \xrightarrow{P} 0$, as $n \to \infty$.

   By Lemma 3, we prove $\boldsymbol{\varepsilon}_{\widehat{R}} - \boldsymbol{\varepsilon}_R = (\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\boldsymbol{\varepsilon} \xrightarrow{P} 0$, as $n \to \infty$.

Then we have

$$\mathbf{U}_j^{\widehat{R}'}\mathbf{L}_{\widehat{R}}^{-1}\mathbf{B}_{\widehat{R}}^{'}\boldsymbol{\varepsilon}_{\widehat{R}} - \mathbf{U}_j^{R'}\mathbf{L}_{R'}^{-1}\mathbf{B}_{R'}^{'}\boldsymbol{\varepsilon}_{R'} = \left[\mathbf{U}_j^{R'} + (\mathbf{U}_j^{\widehat{R}'} - \mathbf{U}_j^{R'})\right]\left[\mathbf{L}_R^{-1} + (\mathbf{L}_{\widehat{R}}^{-1} - \mathbf{L}_R^{-1})\right]\left[\mathbf{B}_R^{'} + (\mathbf{B}_{\widehat{R}}^{'} - \mathbf{B}_R^{'})\right] \times \left[\boldsymbol{\varepsilon}_R + (\boldsymbol{\varepsilon}_{\widehat{R}} - \boldsymbol{\varepsilon}_R)\right] - \mathbf{U}_j^{R'}\mathbf{L}_{R'}^{-1}\mathbf{B}_{R'}^{'}\boldsymbol{\varepsilon}_{R'}.$$

When we write this multiplication into a summation of 2^4 parts, the first part equals to $\mathbf{U}_j^{R'}\mathbf{L}_{R'}^{-1}\mathbf{B}_{R'}^{'}\boldsymbol{\varepsilon}_{R'}$. In every other part, $\|\mathbf{U}_j^{R'}\|$, $\|\mathbf{L}_R^{-1}\|$, $\|\mathbf{B}_R^{'}\|$ and $\|\boldsymbol{\varepsilon}_R\|$ are all bounded in probability. Applying the above 4 proofs, it is straightforward to prove

$$\|\mathbf{U}_j^{\widehat{R}'}\mathbf{L}_{\widehat{R}}^{-1}\mathbf{B}_{\widehat{R}}^{'}\boldsymbol{\varepsilon}_{\widehat{R}} - \mathbf{U}_j^{R'}\mathbf{L}_R^{-1}\mathbf{B}_R^{'}\boldsymbol{\varepsilon}_R\| \xrightarrow{P} 0, \text{as}\, n \to \infty.$$

Therefore, given $\mathbf{y} \in \mathbf{C}_J$ and $\text{card}(J) = d$, we have

$$\frac{(SSE_{\widehat{R}}^0 - SSE_{\widehat{R}}^1) - (SSE_R^0 - SSE_R^1)}{SSE_R^0 - SSE_R^1} \xrightarrow{P} 0, \text{as } n \to \infty.$$

Then, $SSE_{\widehat{R}}^0 - SSE_{\widehat{R}}^1 | \tilde{D} = d \xrightarrow{D} \chi^2(d)$. So, for each k, we have

$$\mathbf{P}\left(\chi_{\widehat{R}}^2 \le a\right) \xrightarrow{D} \sum_{d=0}^{m} \mathbf{P}(\chi^2(d) \le a)\mathbf{P}(\tilde{D} = d), \text{as } n \to \infty.$$

□

Theorem 9. *If* $\|\widehat{\mathbf{R}} - \mathbf{R}\| \xrightarrow{p} 0$*, as* $n \to \infty$*, then for each* k*, the limiting distribution of test statistic (14) is*

$$\mathbf{P}\left(B_{\widehat{R}} \le a\right) \xrightarrow{D} \sum_{d=0}^{m} \mathbf{P}\left[Beta\left(\frac{d}{2}, \frac{n-d-r}{2}\right) \le a\right]\mathbf{P}(\tilde{D} = d), \text{ as } n \to \infty.$$

Proof. First, we need to prove that given $\mathbf{y} \in \mathbf{C}_J$, and $\text{card}(J) = d$,

$$\left.\frac{SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}}}{SSE^0_{\widehat{R}}}\right|_{\tilde{D}=d} \xrightarrow{D} Beta\left(\frac{d}{2}, \frac{n-d-r}{2}\right), \text{as } n \to \infty.$$

Since

$$\left.\frac{SSE^0_{R} - SSE^1_{R}}{SSE^0_{R}}\right|_{\tilde{D}=d} \overset{\mathbf{H}_0}{\sim} Beta\left(\frac{d}{2}, \frac{n-d-r}{2}\right),$$

it is enough to prove that under $\mathbf{H}_0$, given $\mathbf{y} \in \mathbf{C}_J$, and $\text{card}(J) = d$, $B_{\widehat{R}} - B_R \xrightarrow{P} 0, \text{as } n \to \infty$. That is,

$$\frac{SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}}}{SSE^0_{\widehat{R}}} - \frac{SSE^0_{R} - SSE^1_{R}}{SSE^0_{R}} =$$
$$\frac{(SSE^0_R - SSE^0_{\widehat{R}})(SSE^0_R - SSE^1_R) + SSE^0_R\left[(SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}}) - (SSE^0_R - SSE^1_R)\right]}{(SSE^0_{\widehat{R}} - SSE^0_R)SSE^0_R + (SSE^0_R)^2}$$

Applying Theorem 1, we have

$$(SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}}) - (SSE^0_R - SSE^1_R) \xrightarrow{P} 0, \text{as } n \to \infty.$$

So, it is enough to prove

$$\frac{SSE^0_{\widehat{R}} - SSE^0_{R}}{SSE^0_{R}} \xrightarrow{P} 0, \text{as } n \to \infty.$$

That is,

$$\frac{SSE^0_{\widehat{R}} - SSE^0_{R}}{SSE^0_{R}} = \frac{\|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}\|^2 - \|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}^0_R\|^2}{\|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}^0_R\|^2}$$

$$= \frac{(\|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}\| + \|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}^0_R\|)(\|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}\| - \|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}^0_R\|)}{\|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}^0_R\|^2}.$$

It is enough to prove $\|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}\| - \|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}^0_R\| \xrightarrow{P} 0, \text{as}\, n \to \infty$, where

$$\begin{aligned}
\|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}\| - \|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}^0_R\| &\leq \|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}} - \mathbf{y}_R + \widehat{\boldsymbol{\theta}}^0_R\| \\
&\leq \|\mathbf{y}_{\widehat{R}} - \mathbf{y}_R\| + \|\widehat{\boldsymbol{\theta}}^0_R - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}\| \\
&\leq \|\mathbf{f}_{\widehat{R}} - \mathbf{f}_R\| + \|\boldsymbol{\varepsilon}_{\widehat{R}} - \boldsymbol{\varepsilon}_R\| + \|\mathbf{W}_R^{-1}(\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_R - \mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}})\| \\
&\quad + \|(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}}\|
\end{aligned}$$

1. Proof of $\|\mathbf{f}_{\widehat{R}} - \mathbf{f}_R\| \xrightarrow{P} 0, \text{as}\, n \to \infty$.

   Applying Lemma 2, $\|\mathbf{f}_{\widehat{R}} - \mathbf{f}_R\| \leq \|(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\|\|\mathbf{f}\| \xrightarrow{P} 0, \text{as}\, n \to \infty$.

2. Proof of $\|\mathbf{W}_R^{-1}(\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_R - \mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}})\| \xrightarrow{P} 0, \text{as}\, n \to \infty$.

   In the setup of this paper, under $\mathbf{H}_0$, $\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_R$ and $\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}}$ are parametric regression estimators, specifically, weighted least-squares estimators of simple linear regression model. Applying Theorem 2 and 3 in Wang et al. (2013) to simple linear regression model, we have $\|\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_R - \mathbf{f}\| = O_p(n^{-1/2})$ and $\|\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}} - \mathbf{f}\| = O_p(n^{-1/2})$. Therefore,

$$\begin{aligned}
\|\mathbf{W}_R^{-1}(\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_R - \mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}})\| &\leq \|\mathbf{W}_R^{-1}\|\|(\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_R - \mathbf{f}) + (\mathbf{f} - \mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}})\| \\
&\leq \|\mathbf{W}_R^{-1}\| \left[\|(\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_R - \mathbf{f})\| + \|(\mathbf{f} - \mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}})\|\right] \xrightarrow{P} 0, \text{as}\, n \to \infty.
\end{aligned}$$

3. Proof of $\|(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}}\| \xrightarrow{P} 0, \text{as}\, n \to \infty$.

   Applying Lemma 2 and Theorem 3 in Wang et al. (2013), we have

$$\|(\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1})\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}}\| \leq \|\mathbf{W}_{\widehat{R}}^{-1} - \mathbf{W}_R^{-1}\| \left[\|\mathbf{B}\widehat{\boldsymbol{\alpha}}^0_{\widehat{R}} - \mathbf{f}\| + \|\mathbf{f}\|\right] \xrightarrow{P} 0, \text{as}\, n \to \infty.$$

We have proved that $\|\boldsymbol{\varepsilon}_{\widehat{R}} - \boldsymbol{\varepsilon}_R\| \xrightarrow{P} 0, \text{as}\, n \to \infty$, in the proof of Theorem 1. Then we derive that

$$\|\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}\| - \|\mathbf{y}_R - \widehat{\boldsymbol{\theta}}^0_R\| \xrightarrow{P} 0, \text{as}\, n \to \infty.$$

Under $\mathbf{H}_0$, both $\mathbf{y}_{\widehat{R}} - \widehat{\boldsymbol{\theta}}^0_{\widehat{R}}$ and $\mathbf{y}_R - \widehat{\boldsymbol{\theta}}^0_R$ are the residuals of simple linear regression, therefore their norm are well bounded in probability. So, under $\mathbf{H}_0$, given $\mathbf{y} \in \mathbf{C}_J$, and $\text{card}(J) = d$, we have

$$\frac{SSE^0_{\widehat{R}} - SSE^0_R}{SSE^0_R} \xrightarrow{P} 0, \text{as}\, n \to \infty.$$

It is sufficient to derive that

$$\frac{SSE^0_{\widehat{R}} - SSE^1_{\widehat{R}}}{SSE^0_{\widehat{R}}} - \frac{SSE^0_R - SSE^1_R}{SSE^0_R} \xrightarrow{P} 0, \text{as}\, n \to \infty.$$

So, we have

$$\mathbf{P}\left(B_{\widehat{R}} \le a\right) \xrightarrow{D} \sum_{d=0}^{m} \mathbf{P}\left[Beta\left(\frac{d}{2}, \frac{n-d-r}{2}\right) \le a\right] \mathbf{P}(\tilde{D} = d), \text{as}\, n \to \infty.$$

□

3.4 Simulation

The performance of the proposed test is demonstrated through a series of simulations. The test size and power of the proposed test are compared with those of the F-test with unconstrained regression spline estimator as its alternative fit and those of the one-sided t-test with simple regression estimator as its alternative. The test size used in this section is the simulated proportion of the times the null hypothesis is rejected when the data are generated from the null distributions, constant zero mean or linear mean with AR(p) errors. The power of the test is the simulated proportion of times rejecting the null hypothesis when data are generated by the alternative distribution, any monotone trend or convex/concave trend with AR(p) errors. The sample sizes are 250 and 500. In each setup, 10000 datasets are generated to calculated the proportion of rejection. Two scenarios of assumptions are used: one assumes known p and unknown correlation $\mathbf{R}$, and the other assumes unknown p and unknown $\mathbf{R}$.

3.4.1 Test Size

In order to demonstrate the superiority in test size of the proposed test, we simulate datasets with AR(1) errors and also AR(p) errors with $p > 1$. The performance of the proposed test is favorable if the correlation is not quite large. Some of the simulation results are presented in Table 1. Two series of errors are used in the simulations of Table 1. One series are AR(1) errors with increasing correlation: $\varepsilon_i = \eta\varepsilon_{i-1} + z_i, \eta = 0, 0.2, 0.4, 0.6$; the other are three AR(2) sequences: $\varepsilon_i = \eta_1\varepsilon_{i-1} + \eta_2\varepsilon_{i-2} + z_i, \eta = (0, 0.3); \eta = (0.2, 0.2); \eta = (0.3, 0.3)$. If the correlation is known, the test size is exactly equal to 0.05. If the correlation is unknown, the test sizes are inflated by the estimated correlation and increasing with the expansion of the correlation. But if the correlation is ignored in the inference, the test size will be totally blown up. For AR(1) errors, if the correlation is ignored, the test size is 0.172 with $\eta = 0.3$ and 0.301 for $\eta = 0.5$. These highly inflated test sizes illustrate that the correlation must be taken into account. From Table 1, for AR(1) errors with $\eta = 0.2, 0.4, 0.6$, the sizes grow from 0.05 to around 0.080 for $n = 250$ and around 0.065 for $n = 500$ for the proposed test. But for the F-test, the test sizes are enlarged much more than the proposed test. For the simulations with $n = 250$ and unknown p and $\mathbf{R}$, the test sizes for F-test are generally greater than 0.1. This is what we expected because the constrained fit uses more information of the data than the corresponding unconstrained fit. The one-sided t-test has favorable test size when the correlation is comparatively small, such as, AR(1) errors with $\eta = 0, 0.2$. But the test sizes are exaggerated speedily as the increase of the correlation. When $\eta = 0.6$, the test sizes are greater than 0.1 for both $n = 250$ and $n = 500$.

Table 3.1: Simulated proportions of rejecting the null hypothesis. Data are generated from AR(1) sequences with $\eta = 0, 0.2; 0.4; 0.6$ and AR(2), with $\eta = (0, 0.3); (0.2, 0.2); (0.3, 0.3)$. The columns with title "$p, \widehat{\mathbf{R}}$" are test sizes with known p and unknown $\mathbf{R}$. The columns with title "$\widehat{p}, \widehat{\mathbf{R}}$" contains test sizes with unknown p and $\mathbf{R}$, where $\widehat{p}$ is chosen by AIC. Columns from 2 to 5 are test sizes of the proposed test with constrained spline estimators as their alternative fits. Columns from 6 to 9 are test sizes of the F-test with unconstrained spine estimators as their alternative fits. Columns from 10 to 13 are test sizes of the one-sided t-test with simple linear regression estimators as their alternative fits.

	constrained				unconstrained				linear			
	n=250		n=500		n=250		n=500		n=250		n=500	
	$p, \widehat{\mathbf{R}}$	$\widehat{p}, \widehat{\mathbf{R}}$	$p, \widehat{\mathbf{R}}$	$\widehat{p}, \widehat{\mathbf{R}}$	$p, \widehat{\mathbf{R}}$	$\widehat{p}, \widehat{\mathbf{R}}$	$p, \widehat{\mathbf{R}}$	$\widehat{p}, \widehat{\mathbf{R}}$	$p, \widehat{\mathbf{R}}$	$\widehat{p}, \widehat{\mathbf{R}}$	$p, \widehat{\mathbf{R}}$	$\widehat{p}, \widehat{\mathbf{R}}$
$\eta = 0$	0.054	0.056	0.051	0.059	0.074	0.093	0.063	0.074	0.055	0.054	0.051	0.051
$\eta = 0.2$	0.060	0.067	0.057	0.054	0.080	0.122	0.070	0.097	0.061	0.066	0.055	0.057
$\eta = 0.4$	0.063	0.079	0.056	0.065	0.106	0.148	0.084	0.115	0.070	0.076	0.071	0.071
$\eta = 0.6$	0.077	0.087	0.067	0.061	0.158	0.196	0.122	0.151	0.103	0.112	0.096	0.105
$\eta = (0, 0.3)$	0.070	0.071	0.065	0.069	0.140	0.178	0.110	0.138	0.067	0.068	0.064	0.065
$\eta = (0.2, 0.2)$	0.069	0.081	0.064	0.067	0.151	0.182	0.110	0.133	0.070	0.072	0.065	0.068
$\eta = (0.3, 0.3)$	0.089	0.093	0.076	0.079	0.222	0.263	0.171	0.191	0.093	0.100	0.092	0.089

3.4.2 Power Study

Data are generated with sigmoid, truncated cubic and linear trend:

1. linear: $f(x) = \psi_i x$, $i = 1, 2, 3$;

2. sigmoid: $f(x) = \psi_i \frac{e^{10x-5}}{1+e^{10x-5}}$, $i = 1, 2, 3$;

3. truncated cubic: $f(x) = \psi_i 4(x - 1/2)^3 I_{x>1/2}$, $i = 1, 2, 3$.

The errors are generated from AR(1) series with $\eta = 0, 0.2, 0.4, 0.6$. For each kind of trend, we select slopes ψ_1, ψ_2 and ψ_3, by which the powers of the corresponding one-sided t-tests for independent normal errors are 0.25, 0.5 and 0.75. Sample size is 250 for each dataset. For each setup, 10000 datasets are generated independently. For each dataset, the proportions of times rejecting the null hypothesis are calculated for the proposed test with constrained spline estimator as its alternative fit, F-test with unconstrained spline estimator as its alternative fit and one-sided t-test with simple linear regression estimator as its alternative fit. The collection of simulation output are in Table 2.

When $\eta = 0; 0.2; 0.4$, the powers of the proposed test with constrained alternative are greater than F-test uniformly for all three kinds of trends with different slopes. This is also we we expected, because the proposed test uses the information of the shape of trend, therefore has higher power than the F-test with unconstrained alternative. But when $\eta = 0.6$, the power of F-test is no less than that of the proposed test. However, from Table 1, the test size of the F-test is highly inflated when $\eta = 0.6$, hence the F-test is still ineffective even if the power is favorable. Even for linear data where one-sided t-test does the best among those three tests as expected, the proposed test still behaves better than the F-test. One-sided t-test is competitive in power for linear trend and some non-linear trends closed to linear, such as sigmoid with $f(x) = (e^{10x-5})/(1+e^{10x-5})$, for $\eta = 0, 0.2, 0.4$. But for other non-linear trend, such as truncated cubic with $f(x) = \psi_i 4(x-1/2)^3 I_{x>1/2}$, the superiority in power of the proposed test is obvious for $\eta = 0, 0.2, 0.4$. Because simple linear regression alternative of the t-test fail to catch the curvature of the true trend. When $\eta = 0.6$, just like F-test, from Table 1 we know that the test sizes for one-sided t-test are highly inflated and greater than 0.1. Therefore the t-test is undesirable when the correlation grows greater.

3.5 Real Data

3.5.1 Argentina Rainfall Data

Data of yearly rainfall in Argentina from 1884 to 1996 are used in Wu et al. (2001). They propose an hypothesis test for monotonic trends in short range dependent data. The alternative fit is a piece-wise linear isotonic regression estimator. They also use this test as one perspective for changepoint problems. We use the proposed hypothesis test with the constrained spline regression estimator as its alternative fit and one-sided t-test with simple linear regression estimator as its alternative fit to test whether the yearly rainfall in Argentina from 1884 to 1996 is constant or monotone increasing. We also use F-test with unconstrained spline regression estimator as its alternative fit to test whether the yearly rainfall data is constant or not. The three fits on Argentina rainfall data are in Figure 3.1.

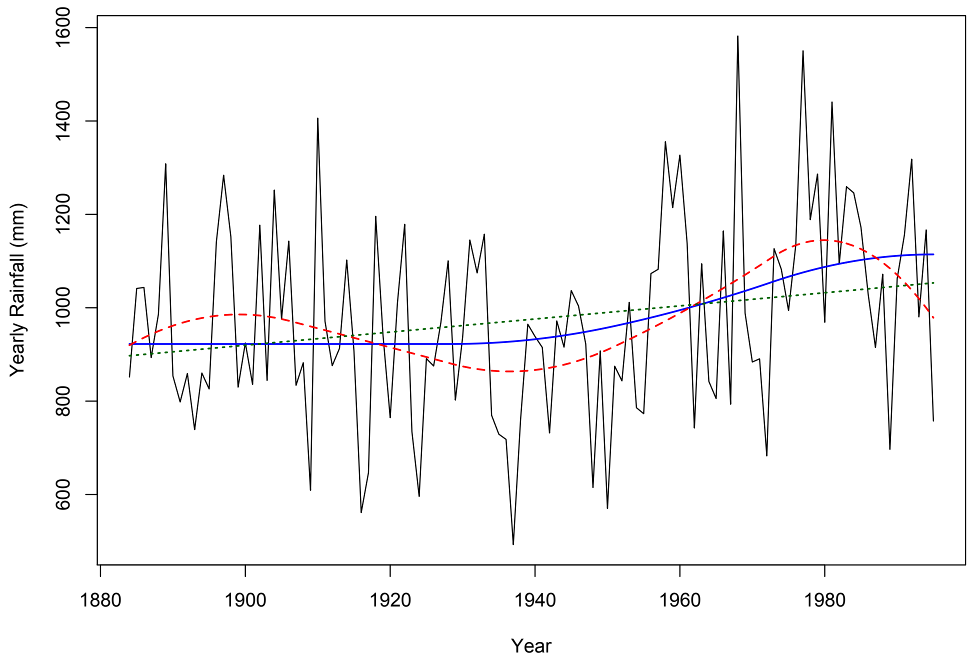

Figure 3.1: Three kinds of alternative fits on Argentina rainfall data: blue solid curve is the constrained regression spline fit, the alternative fit of our proposed test; the red dashed curve is the unconstrained regression spline fit, the alternative fit of F-test; the green dotted line is the simple linear regression fit, the alternative fit of the one-sided t-test. For both constrained spline regression and unconstrained spline regression, 6 knots are used.

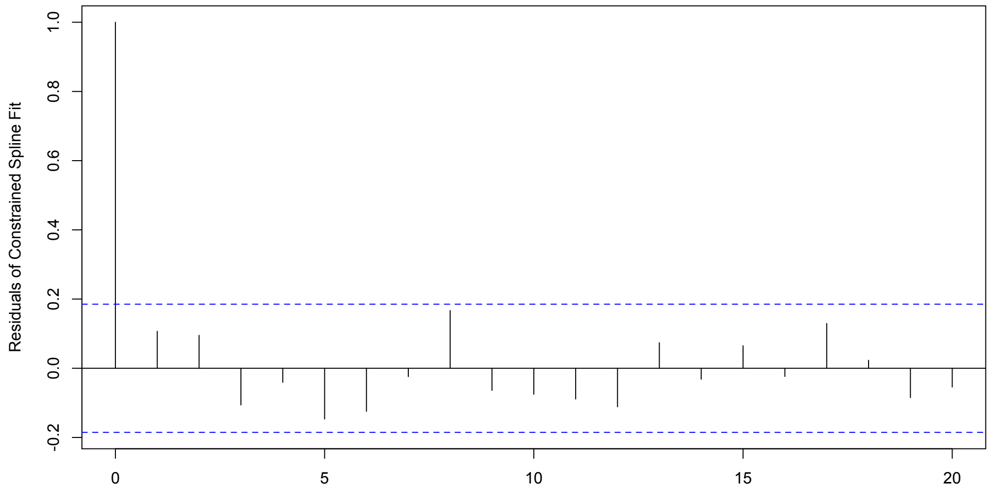

Figure 3.2: Sample autocorrelation function of residuals from constrained spline fit of Argentina rainfall data.

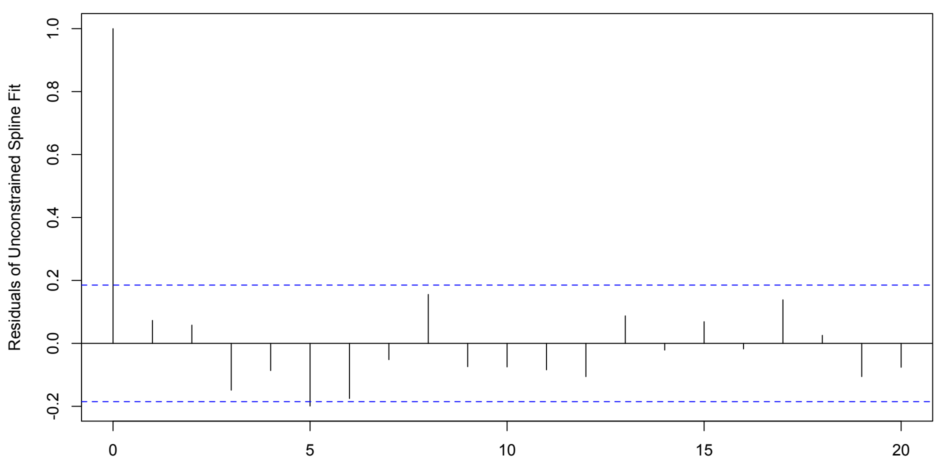

Figure 3.3: Sample autocorrelation function of residuals from unconstrained spline fit of Argentina rainfall data.

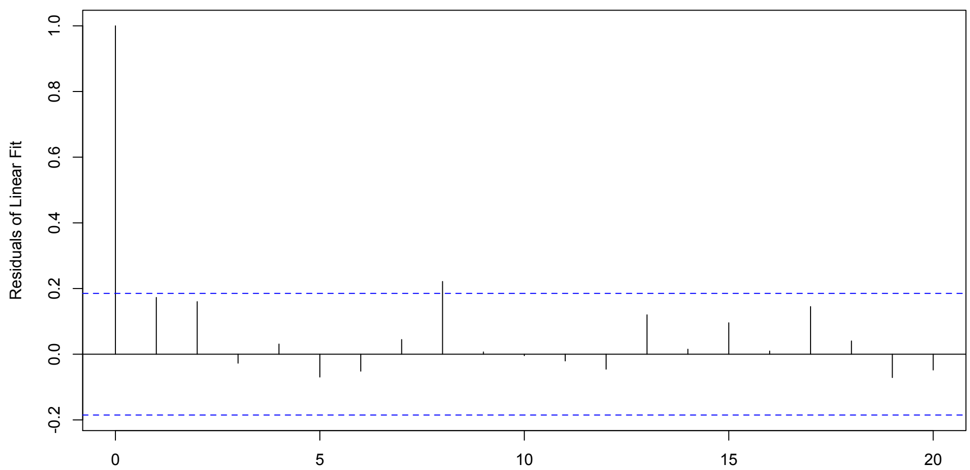

Figure 3.4: Sample autocorrelation function of residuals from linear fit of Argentina rainfall data.

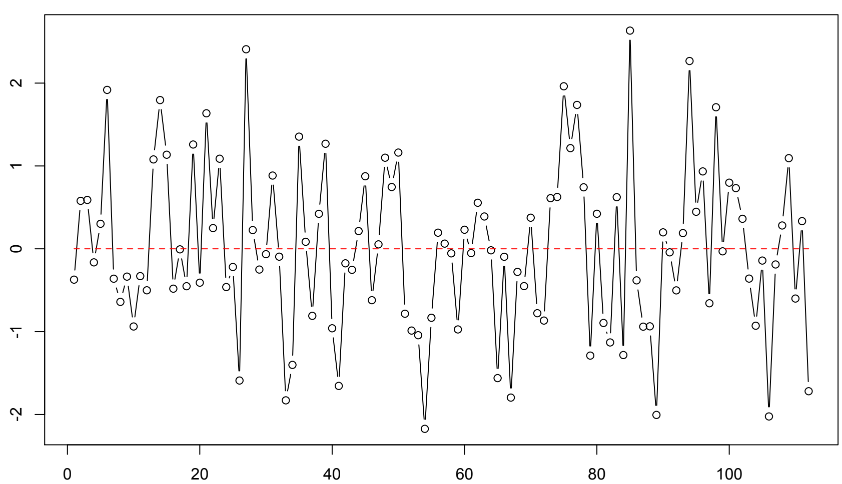

Figure 3.5: Plot of Standard residuals from constrained spline fit of Argentina rainfall data.

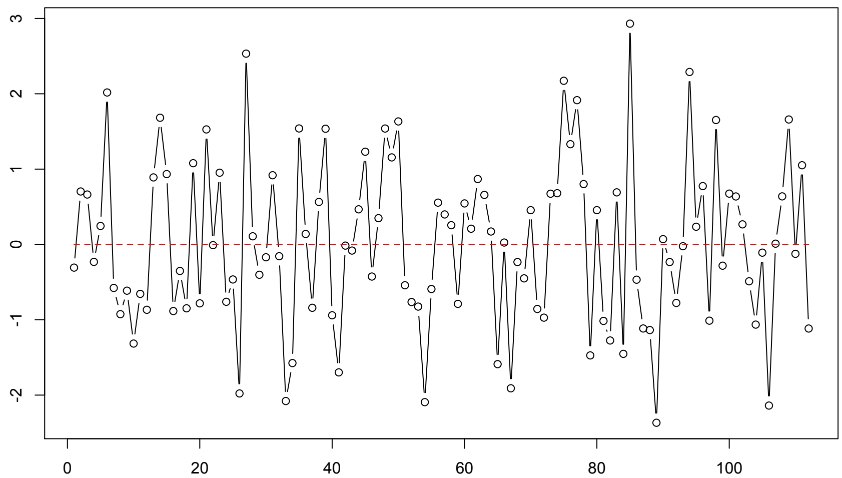

Figure 3.6: Plot of Standard residuals from unconstrained spline fit of Argentina rainfall data.

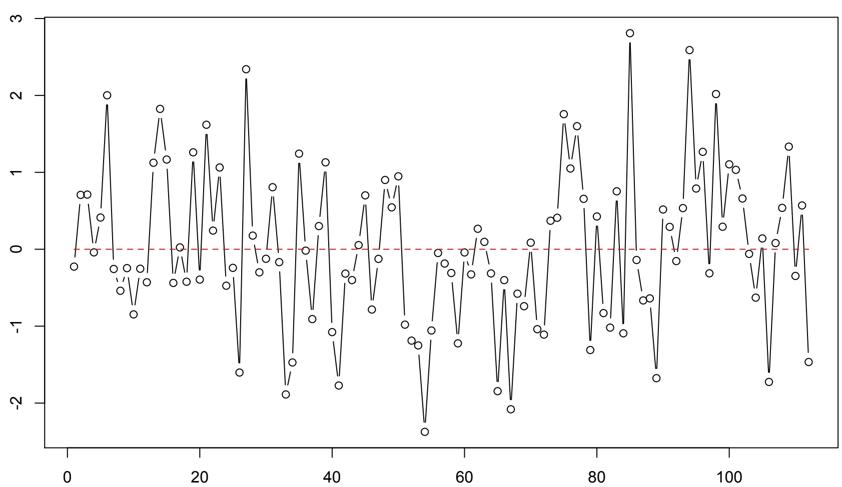

Figure 3.7: Plot of Standard residuals from linear fit of Argentina rainfall data.

Table 3.2: The power of the proposed test is compared with the F-test with unconstrained regression spline estimator as its alternative fit and one-sided t-test with simple regression estimator as its alternative. Data are generated by sigmoid, truncated cubic and linear trend with AR(1) errors, where $\eta = 0, 0.2, 0.4, 0.6$. For each kind of trend, three slopes are selected, by which the powers of the corresponding one-sided t-tests for independent normal errors are 0.25, 0.5 and 0.75. Sample size is 250.

		sigmoid			cubic			linear		
		0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
constrained	$\eta = 0$	0.225	0.428	0.688	0.316	0.637	0.869	0.243	0.474	0.696
	$\eta = 0.2$	0.225	0.357	0.548	0.253	0.500	0.731	0.198	0.374	0.559
	$\eta = 0.4$	0.173	0.277	0.414	0.200	0.357	0.541	0.168	0.298	0.397
	$\eta = 0.6$	0.134	0.177	0.287	0.164	0.248	0.357	0.142	0.212	0.283
unconstrained	$\eta = 0$	0.146	0.269	0.450	0.192	0.417	0.677	0.145	0.262	0.419
	$\eta = 0.2$	0.165	0.242	0.375	0.200	0.346	0.554	0.164	0.235	0.355
	$\eta = 0.4$	0.170	0.204	0.290	0.191	0.276	0.404	0.167	0.217	0.278
	$\eta = 0.6$	0.202	0.232	0.265	0.217	0.259	0.327	0.205	0.224	0.264
linear	$\eta = 0$	0.271	0.499	0.760	0.265	0.507	0.744	0.263	0.514	0.751
	$\eta = 0.2$	0.223	0.406	0.620	0.225	0.404	0.595	0.224	0.405	0.608
	$\eta = 0.4$	0.192	0.321	0.472	0.195	0.328	0.464	0.188	0.321	0.487
	$\eta = 0.6$	0.199	0.290	0.379	0.204	0.281	0.382	0.192	0.283	0.372

The estimated order $\widehat{p} = 0$ for the constrained spline regression estimation. From 3.2, the plot of sample ACF of residuals from constrained spline fit also demonstrates that the residuals are independent. The p-value of kpss.test function in R is greater than 0.1 for the residuals. So we fail to reject that the residuals are stationary. The p-value of the proposed test is 0.001. The estimated order $\widehat{p} = 2$ and $\widehat{\boldsymbol{\eta}} = (0.150, 0.134)$. The p-value of one-sided t-test is 0.039. The estimated order $\widehat{p} = 12$ and $\boldsymbol{\eta} = (-0.021, 0.016, -0.193, -0.085, 0.207, -0.270, -0.112, 0.098, -0.191, -0.161, -0.113, -0.198)$. The p-value of F-test is almost 0. Without the constraint, spline regression quite follow the data too much. The curve of fit is wiggly and the estimation of the correlation among the data is hard to explain. For both t-test and F-test, the p-values of kpss.test function in R, referring to Kwiatkowski et al. (1992), are both greater than 0.1 for the residuals. So we fail to reject that the residuals are stationary.

3.5.2 Price of Liquefied U.S. Natural Gas Exports Data

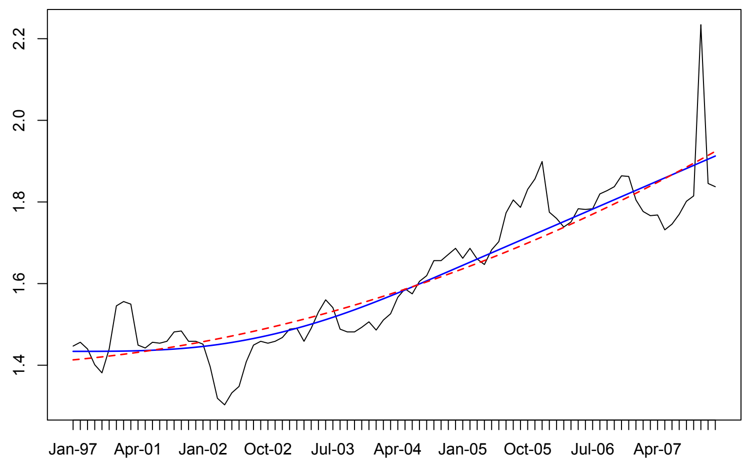

Figure 3.8: Price of Liquefied U.S. Natural Gas Exports (Dollars per Thousand Cubic Feet) Data: log transformed of monthly price from Jan, 1997 to Sep, 2007. The blue solid curve is the regression spline fit constrained to be increasing and convex, the alternative fit of the proposed test. The red dashed curve is the generalized linear model with a quadratic term. We use F-test to test whether the quadratic term is significant or not, for a comparison to our test.

Canada and Mexico are two main countries where U.S. exports its liquefied natural gas. The log transformed monthly data of price of the liquefied U.S. natural gas from January, 1997 to September, 2007 are in Figure 3.8. The original data are available on `http://tonto.eia.gov/dnav/ng/hist/n9133us3m.htm`. The price is increasing with the date significantly if we test whether the trend is constant or monotone increasing. Beside the monotone trend, we also want to known whether the price increase faster with the date. The proposed test of increasing convexity is used on this data to test whether the trend is linear or increasing and convex. The estimated order $\widehat{p} = 1$ and $\widehat{\eta} = 0.52$. The p-value is 0.005, significant as we expect. We also conduct a F-test to test whether the trend is linear or quadratic in date. Usually, if lack of information about the form of the data and incorrectly specifying a parametric form, hypothesis test with the parametric alternative fit could fail

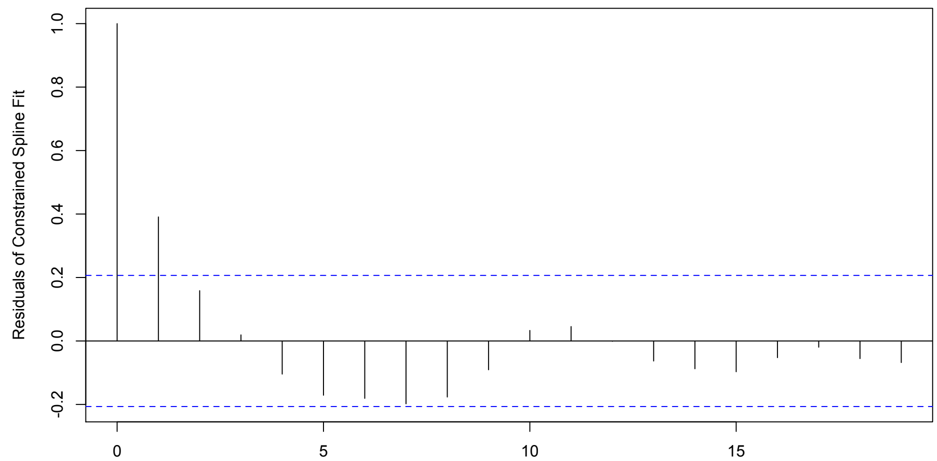

Figure 3.9: Sample autocorrelation function of residuals from constrained spline fit of price of liquefied U.S. natural gas exports data.

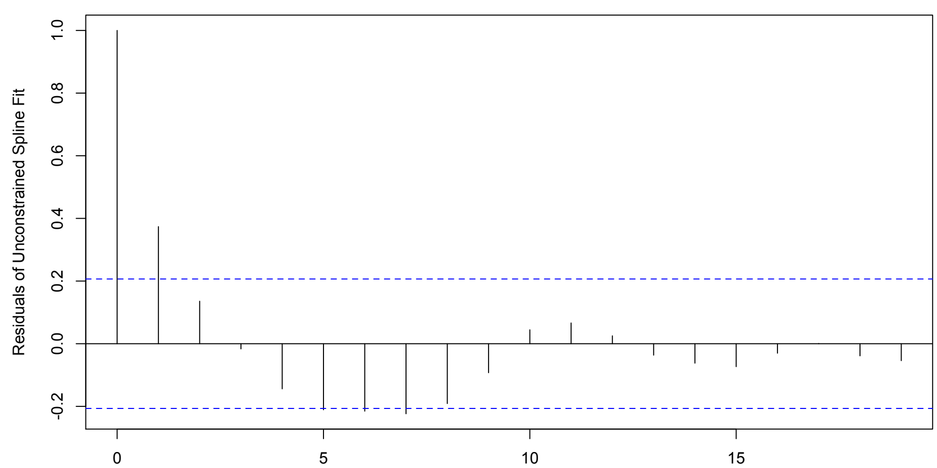

Figure 3.10: Sample autocorrelation function of residuals from unconstrained spline fit of price of liquefied U.S. natural gas exports data.

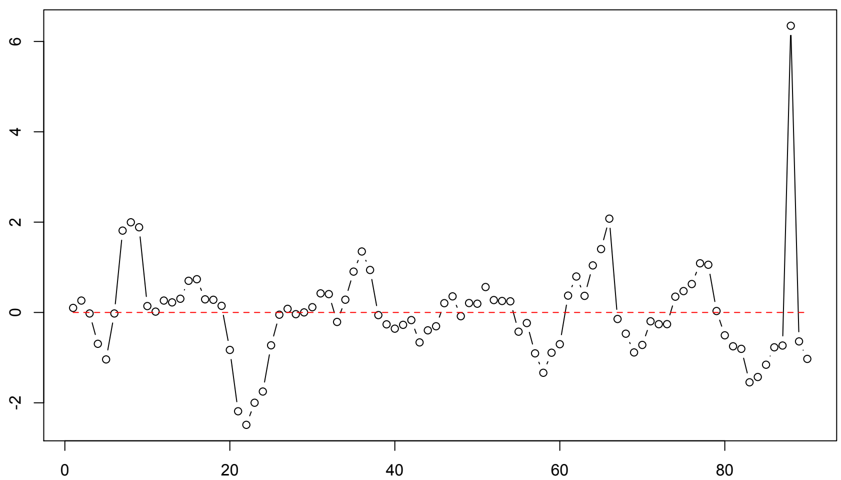

Figure 3.11: Plot of Standard residuals from constrained spline fit of price of liquefied U.S. natural gas exports data.

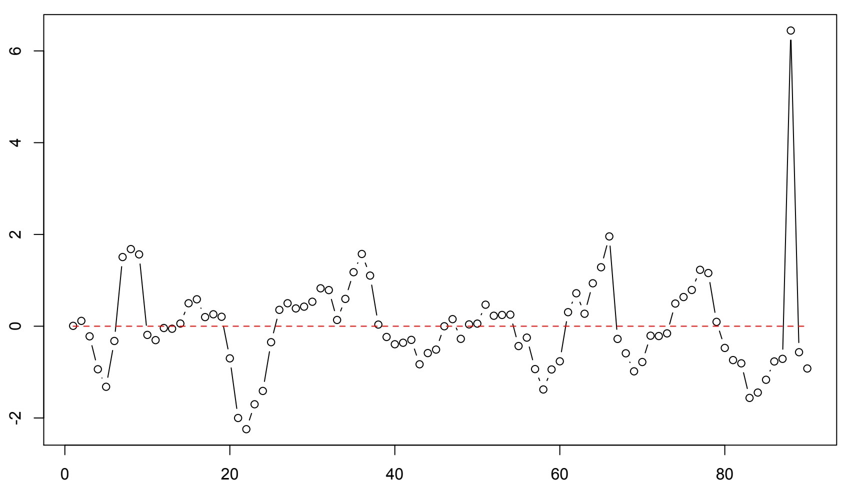

Figure 3.12: Plot of Standard residuals from unconstrained spline fit of price of liquefied U.S. natural gas exports data.

to detect the trend or give a larger p-value than the more powerful test with nonparametric alternative fit. But in this case, p-value for F-test is 0.0004, smaller than the p-value given by the proposed test. For both tests, we use kpss.test function in R to check the stationarity of the residuals. Both of p-values are greater than 0.1, so we fail to reject that the residuals are stationary. But whether the assumption that the log transformed price of liquefied U.S. natural gas export has quadratic form increasing trend on date from Jan, 1997 to Sep, 2007 is appropriate or not still need additional hypothesis test to determine.

CHAPTER 4

CONCLUSION AND FUTURE WORK

4.1 Conclusion

In this dissertation, we developed the constrained spline regression model and shape restricted hypothesis tests for data with stationary AR(p) errors. In Chapter 2, we proposed the estimation method and asymptotic properties of the estimator. The asymptotic rate for constrained spline estimators with estimation of correlation and ignoring the correlation have been proved to be the same. Even if we have an inconsistent estimator of correlation, as long as it satisfies appropriate conditions, the estimation of trend based on this estimator is still consistent and attains the optimal rate. However, estimation of the trend is substantially improved for moderate-sized samples under proposed iteration method. Further, the asymptotic variances of the three estimators are different. In Chapter 3, we studied the hypothesis tests of the trend, such as, constant vs. monotone, and linear vs. convex, in the presence of AR(p) errors. The asymptotic distributions of the test statistics depends on consistent estimation of the correlation. If correlation is unknown and the estimated correlation is consistent, the likelihood test statistics weighted by this consistent estimated correlation have their approximating distributions as a mixture of χ^2 distributions if variance is known or a mixture of *Beta* distributions if variance is unknown. If the correlation is existing and ignored, the test size will be greatly inflated.

4.2 Future Work

4.2.1 Tests of the Monotonicity and Convexity in the presence of correlation

Methods for testing the shape of a function, such as monotonicity and convexity, are useful in many applications, especially for time series data. We propose a set of hypothesis tests using regression splines and shape restricted inference in the presence of stationary AR(p) errors. The null hypothesis H_1 is that the function is constrained to be monotone or convex, and H_2 is that the function is unconstrained. The tests of H_1 versus H_2 use an estimate of the distribution of the minimum slope or second derivative of the spline estimator under the null hypothesis and proved to behave nicely both for small sample size and asymptotically. The test that H_1: the function is decreasing and convex, versus H_2: the function is unconstrained, is applied to intensity data from small angle X-ray scattering (SAXS) experiments. The proposed method serves as a useful pre-test in this context, because under H_1, a classical regression-based procedure for estimating a molecule's radius of gyration can be applied. Assume that the observed data $\{(x_i, y_i)\}$, for $1 \leq i \leq n$, are generated by the model

$$y_i = f(x_i) + \sigma\varepsilon_i,$$

where f is a smooth function. Suppose that $x_i \in [0, 1]$ and equally spaced.

$$\varepsilon_i = \sum_{j=1}^{p} \theta_j \varepsilon_{i-j} + z_i,$$

where z_i are standard normal variables. We want to test:

$$\mathbf{H}_1 : f \text{ is constrained} \quad vs. \quad \mathbf{H}_2 : f \text{ is unconstrained}.$$

This setup can be used to test: monotone v.s. unconstrained, convex/concave v.s. unconstrained, and any combinations, such as, monotone and convex v.s. unconstrained. We apply shape restricted inference with cone projection theories into the tests of constrained v.s. unconstrained of the trend. Nonparametric regression splines are used to estimate the null fit and alternative fit. The null fit is constrained spline regression estimator and the alternative fit is regular spline regression estimator. Both of them are estimated for data with stationary AR(p) errors. Following the setup in Chapter 3, it is equivalent to test: H_1: $\mathbf{Tb} \geq 0$ vs. H_2: $\mathbf{Tb} \neq 0$. The test procedures are:

1. Obtain the unconstrained estimator $\hat{\mathbf{b}}_u$ and determine the minimum of the slops at the knots, that is, calculate $\mathbf{t}_{\min} = \min(\mathbf{T}\hat{\mathbf{b}}_u)$;

2. If $\mathbf{t}_{\min}$ is non-negative, then do not reject H_1;

3. Otherwise, estimate distribution of $\mathbf{t}_{\min}$ under H_1, using $\hat{\mathbf{b}}_c$ for $\boldsymbol{\beta}$, and estimating the model variance using $\hat{\mathbf{b}}_u$;

4. If $\mathbf{t}_{\min}$ is smaller than estimated α-level percentile, reject H_1 in favor of H_2.

We conduct the simulations to study the test size and power of this procedures with data generated from different null distributions. Also the asymptotic properties will be investigated.

REFERENCES

Altman, N. (1992). An iterated cochrane-orcutt procedure for nonparametric regression. *Journal of Statistical Computation and Simulation 40*(1-2), 93–108.

Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association 85*(411), pp. 749–759.

Birke, M. and H. Dette (2006). Estimating a convex function in nonparametric regression. *Scandinavian Journal of Statistics 34*(2), 384–404.

Brillinger, D. R. (1989). Consistent detection of a monotonic trend superposed on a stationary time series. *Biometrika 76*(1), pp. 23–30.

Brockwell, P. and R. Davis (2009). *Time Series: Theory and Methods.* Springer Series in Statistics. Springer.

Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics 26*(4), pp. 607–616.

Brunk, H. D. (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Statist. 29*, 437–454.

Claeskens, G., T. Krivobokova, and J. D. Opsomer (2009). Asymptotic properties of penalized spline estimators. *Biometrika 96*(3), 529–544.

Cohen, A., J. Kemperman, and H. Sackrowitz (1993). Unbiased tests for normal order restricted hypotheses. *Journal of multivariate analysis 46*(1), 139–153.

Cohen, A., H. Sackrowitz, and E. Samuel-Cahn (1995). Constructing tests for normal order-restricted inference. *Bernoulli 1*(4), 321–333.

Dette, H., N. Neumeyer, and K. Pilz (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli 12*(3), 469–490.

Dette, H. and R. Scheder (2006). Strictly monotone and smooth nonparametric regression for two or more variables. *Canadian Journal of Statistics 34*(4), 535–561.

Diggle, P. J. and M. F. Hutchinson (1989). On spline smoothing with autocorrelated errors. *Australian Journal of Statistics 31*(1), 166–182.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical Science 11*(2), pp. 89–102.

Francisco-Fernandez, M. and J. Opsomer (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Canadian Journal of Statistics 33*(2), 279–295.

Fraser, D. and H. Massam (1989). A mixed primal-dual bases algorithm for regression under inequality constraints. application to concave regression. *Scandinavian Journal of Statistics*, 65–74.

Fuller, W. A. (2009). *Introduction to Statistical Time Series.* John Wiley & Sons, Inc.

Golub, G. H. and C. F. Van Loan (1996). *Matrix computations*, Volume 3. JHUP.

Hall, P. and L.-S. Huang (2001). Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics 29*(3), 624–647.

Hall, P. and I. Keilegom (2003). Using difference-based methods for inference in nonparametric regression with time series errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 443–456.

Hall, P. and J. D. Opsomer (2005). Theory for penalised spline regression. *Biometrika 92*(1), 105–118.

Hansen, J., M. Sato, R. Ruedy, K. Lo, D. W. Lea, and M. Medina-Elizade (2006). Global temperature change. *Proceedings of the National Academy of Sciences 103*(39), 14288–14293.

Hart, J. D. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society. Series B (Methodological) 53*(1), pp. 173–187.

Hart, J. D. (1994). Automated kernel smoothing of dependent data by using time series cross-validation. *Journal of the Royal Statistical Society. Series B (Methodological) 56*(3), pp. 529–542.

Horn, R. A. and C. R. Johnson (1990). *Matrix analysis.* Cambridge university press.

Huang, J. Z. (1998). Projection estimation in multiple regression with application to functional anova models. *The Annals of Statistics 26*(1), pp. 242–272.

Kauermann, G., T. Krivobokova, and L. Fahrmeir (2009). Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 71*(2), 487–503.

Kim, T., B. Park, M. Moon, and C. Kim (2009). Using bimodal kernel for inference in nonparametric regression with correlated errors. *Journal of Multivariate Analysis 100*(7), 1487–1497.

Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics 54*(1), 159–178.

Lawson, C. and R. Hanson (1995). *Solving least squares problems*, Volume 15. SIAM.

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

Mammen, E. and C. Thomas-Agnan (1999). Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics 26*, 239–252.

Meyer, M. (2013). A simple new algorithm for quadratic programming with applications in statistics.

Meyer, M. C. (2003). A test for linear versus convex regression function using shape-restricted regression. *Biometrika 90*(1), 223–232.

Meyer, M. C. (2008). Inference using shape-restricted regression splines. *The Annals of Applied Statistics 2*(3), pp. 1013–1033.

Meyer, M. C. (2012). Constrained penalized splines. *Canadian Journal of Statistics 40*(1), pp. 190–206.

Mukerjee, H. (1988). Monotone nonparametric regression. *The Annals of Statistics 16*(2), 741–750.

Opsomer, J., Y. Wang, and Y. Yang (2001). Nonparametric regression with correlated errors. *Statistical Science 16*(2), pp. 134–153.

Ramsay, J. O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 60*, 365–375.

Raubertas, R., C. Lee, and E. Nordheim (1986). Hypothesis tests for normal means constrained by linear inequalities. *Communications in Statistics-Theory and Methods 15*(9), 2809–2833.

Robertson, T., F. T. Wright, and R. Dykstra (1988). *Order Restricted Statistical Inference*. John Wiley & Sons, Inc.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Cambridge University Press.

Siddiqui, M. M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *The Annals of Mathematical Statistics 29*(2), pp. 585–588.

Silvapulle, M. J. and P. K. Sen (2004a, November). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions* (1 ed.). Wiley-Interscience.

Silvapulle, M. J. and P. K. Sen (2004b, November). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions* (1 ed.). Wiley-Interscience.

Slutsky, E. (1925). *Über stochastische Asymptoten und Grenzwerte*. Amministrazione del "Metron".

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics 10*(4), pp. 1040–1053.

Tantiyaswasdikul, C. and M. B. Woodroofe (1994). Isotonic smoothing splines under sequential designs. *Journal of Statistical Planning and Inference 38*, 75–87.

Wang, H., M. Meyer, and J. Opsomer (2013). Constrained spline regression in the presence of correlated errors. Technical report, Department of Statistics, Colorado State University.

Wang, X. and J. Shen (2010). A class of grouped brunk estimators and penalized spline estimators for monotone regression. *Biometrika 97*(3), 585–601.

Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association 93*(441), pp. 341–348.

Woodward, W. and H. Gray (1993). Global warming and the problem of testing for trend in time series data. *Journal of Climate 6*(5), 953–962.

Wu, W., M. Woodroofe, and G. Mentz (2001). Isotonic regression: Another look at the changepoint problem. *Biometrika 88*(3), 793–804.