

University of Montana

ScholarWorks at University of Montana

Graduate Student Theses, Dissertations, &
Professional Papers

Graduate School

2007

Comparison of Trend Detection Methods

Katharine Lynn Gray

The University of Montana

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

Let us know how access to this document benefits you.

Recommended Citation

Gray, Katharine Lynn, "Comparison of Trend Detection Methods" (2007). *Graduate Student Theses, Dissertations, & Professional Papers*. 228.
<https://scholarworks.umt.edu/etd/228>

This Dissertation is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

COMPARISON OF TREND DETECTION METHODS

By

Katharine Lynn Gray

B.S. University of Montana, Missoula, Montana 1997

M.S. University of Montana, Missoula, Montana 2000

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Mathematical Sciences

The University of Montana
Missoula, MT

Spring 2007

Approved by:

Dr. David A. Strobel, Dean
Graduate School

Dr. David Patterson, Committee Co-Chair
Department of Mathematical Sciences

Dr. Brian Steele, Committee Co-Chair
Department of Mathematical Sciences

Dr. Jonathan Graham
Department of Mathematical Sciences

Dr. George McRae
Department of Mathematical Sciences

Dr. Elizabeth Reinhardt
RMRS Fire Sciences Laboratory

Comparison of Trend Detection Methods

Committee Chair: Brian Steele, Ph.D and David Patterson, Ph.D.

Trend estimation is important in many fields, though arguably the most important applications appear in ecology. Trend is difficult to quantify; in fact, the term itself is not well-defined. Often, trend is quantified by estimating the slope coefficient in a regression model where the response variable is an index of population size, and time is the explanatory variable. Linear trend is often unrealistic for biological populations; in fact, many critical environmental changes occur abruptly as a result of very rapid changes in human activities. My PhD research has involved formulating methods with greater flexibility than those currently in use. Penalized spline regression provides a flexible technique for fitting a smooth curve. This method has proven useful in many areas including environmental monitoring; however, inference is more difficult than with ordinary linear regression because so many parameters are estimated. My research has focused on developing methods of trend detection and comparing these methods to other methods currently in use. Attention is given to comparing estimated Type I error rates and power across several trend detection methods. This was accomplished through an extensive simulation study. Monte Carlo simulations and randomization tests were employed to construct an empirical sampling distribution for the test statistic under the null hypothesis of no trend. These methods are superior over other smoothing methods of trend detection with respect to achieving the designated Type I error rate. The likelihood ratio test using a mixed effects model had the most power for detecting linear trend while a test involving the first derivative was the most powerful for detecting nonlinear trend for small sample sizes.

Acknowledgements

I would like to acknowledge the many people that have helped, supported, and guided me in writing this dissertation. I would especially like to thank Drs. Brian Steele and David Patterson for persevering with me as my advisors through out the time it took me to complete this research. In addition, I would like to thank the other committee members, Drs. Jon Graham, George McRae, and Elizabeth Reinhardt, for all their helpful comments and encouragement. I would also like to thank Guy Shepard for his help in the computer lab. Finally, I would like to thank my husband Tom for his unconditional support and love.

Contents

Abstract	ii
List of Tables	viii
List of Figures	x
1 Introduction	1
2 Spline Regression	5
2.1 Spline functions	6
2.1.1 Interpolation Splines	6
2.2 Splines in Statistics	7
2.3 Regression splines	7
2.3.1 Selecting the number and location of knots	10
2.4 Smoothing Splines	11
2.4.1 B-splines with difference penalty	12
2.4.2 Regression with a B-spline basis	16

2.4.3	Degrees of Freedom of a smoother	17
2.4.4	Smoothing Parameter	18
2.4.5	Residual variance	19
2.5	Penalized Regression splines	19
2.6	Penalized splines as mixed models	21
3	Trend Detection	25
3.0.1	Goals of trend analysis	26
3.1	Trend detection methods	27
3.1.1	Linear Regression	27
3.1.2	Nonparametric correlation coefficient	29
3.1.3	LOESS	31
3.1.4	Likelihood Ratio Test	33
3.1.5	Examining first derivatives	35
3.1.6	Test for trend using first derivatives	35
3.1.7	First derivatives for penalized spline regression using a linear mixed model	37
3.1.8	First derivatives for B-splines	38
3.2	Tests for trend with autocorrelated errors	41
3.2.1	Generalized linear regression with autoregressive errors	41
3.2.2	Sum of the first derivatives with autocorrelated errors	42
3.2.3	Likelihood ratio test with correlated errors	45

3.3	Example	46
3.3.1	Least squares regression	48
3.3.2	LOESS	48
3.3.3	Kendall's τ	49
3.3.4	Likelihood ratio test	49
3.3.5	Sum of the first derivatives	49
3.3.6	Discussion	50
4	Simulation Results	54
4.1	Simulation Procedure	56
4.1.1	Generated data	57
4.1.2	Comparison of Type I Error Rates	57
4.1.3	Power results	63
4.2	Autocorrelation methods	67
4.2.1	Simulation results for autocorrelation methods	69
4.3	Conclusions	71
5	Case Study	73
5.1	Data	74
5.2	Random site effect	75
5.3	Results	77

5.4	Most recent trend	78
5.5	Conclusion	79
6	Conclusions and Future Research	81
	Bibliography	84

List of Tables

3.1	Calculation of Kendall's tau coefficient for the Veery data.	31
3.2	P-values for trend tests over different time periods.	52
4.1	Description of generated data. Linear trend with the addition of influential observations were examined to explore the effect of influential observations on power.	59
4.2	Type I error rates for 1500 simulations. Standard errors for 1500 simulations are given in Figure 4.2. Standard errors ranged from .0055 to .0118. The methods compared were linear regression (LIN), Kendall's τ (Ken), the likelihood ratio test for trend (LRT), a randomization test for trend using the sum of the first derivatives (SUM.R), a test for trend using a locally weighted regression (LOESS), the trend detection test using the sum of the first derivatives using P-splines, and a test using the sum of the first derivatives using a linear mixed effects model and distributional assumptions about the test statistic (SUM). . .	60
4.3	Estimated power for 1500 simulations. Figure 4.2 gives the standard error for an estimated power. The methods compared were linear regression (LIN), Kendall's τ (Ken), the likelihood ratio test for trend (LRT), a randomization test for trend using the sum of the first derivatives (SUM.R), a test for trend using a locally weighted regression (LOESS), the trend detection test using the sum of the first derivatives using P-splines, and a test using the sum of the first derivatives using a linear mixed effects model (SUM).	66

4.4	Estimated Type I error rates for 1500 simulations for data simulated with no trend and linear trend. Data were simulated with ϕ equal to 0,.3 and .5. The methods that are compared are a generalized least squares method (LIN.AR), a randomization test to test the significance of the sum of the first derivatives (SUM.AR), and a likelihood ratio test (LRT.AR). All of these methods account for serial correlation.	69
4.5	Estimated power for 1500 simulations for data simulated with linear trend. Data were simulated with ϕ equal to 0,.3 and .5. The methods that are compared are a generalized least squares method (LIN.AR), a randomization test to test the significance of the sum of the first derivatives (SUM.AR), and a likelihood ratio test (LRT.AR). All of these methods account for serial correlation.	70
5.1	Results of linear mixed model fit for mourning dove data. The estimated parameter values and standard errors are given.	78

List of Figures

2.1	Carolina wren data collected on a BBS route. The graphs show linear spline regression fits with 1-4 knots. Graph b) has one knot located at 1977. Graph c) has 2 knots located at 1977 and 1981. Graph d) has knot locations at 1977, 1981 and 1991. Graph e) has knot locations at 1977, 1981, 1991 and 1996 . . .	9
2.2	Carolina wren data with spline regression fits of degree 1,2 and 3 with knot locations at 1977, 1981 and 1991	10
2.3	B-spline basis of degree 3 with knots at 1,2,3,4,5. A B-spline of degree 3 with 5 knots will have 7 B-splines. a) shows the first B-spline b) shows the second, c) the third, and d) shows all 7 B-splines.	15
3.1	Veery counts for 10 years for a route in Maryland.	30
3.2	Wolverine counts from 1752 to 1908. Five different regression fits are shown on individual graphs: b) Least squares regression, c) locally weighted regression (LOESS), d) cubic spline using a B-spline basis, e) a linear spline using a linear mixed effects model, f) a cubic spline using a linear mixed effects model	47
3.3	Fitted first derivative curve for wolverine data with 95% simultaneous confidence intervals. Regions for which the variability bands are positive correspond to time intervals in which the curve is significantly increasing while regions which the variability bands are negative correspond to time intervals in which the curve is significantly decreasing.	51
3.4	Wolverine data for certain time periods.	53

4.1	Examples of simulated data with line indicating the true function.	58
4.2	Standard error of power and Type I error rates estimates for a sample size of 1500.	61
4.3	Ninety-five percent confidence intervals for Type I error rates for n of 25. SUM is the method that used the sum of the first derivatives and SUM.R is the trend detection method when a randomization procedure was used to obtain the observed significance. The tests were all performed at an α level of .05. . .	61
4.4	Ninety-five percent confidence intervals for Type I error rates for n of 50. . . .	62
4.5	Ninety-five percent confidence intervals for Type I error rates for n of 200. . . .	62
4.6	Ninety-five percent confidence intervals for power to detect linear trend for n of 25. SUM is the method that used the sum of the first derivatives and SUM.R is the trend detection method when a randomization procedure was used to obtain the observed significance. The tests were all performed at an α level of .05.	64
4.7	Ninety-five percent confidence intervals for power to detect linear trend for n of 50.	65
4.8	Ninety-five percent confidence intervals for power to detect linear trend for n of 200.	65
5.1	Mourning dove counts at seven different BBS routes in Maryland from 1966 to 2005.	74
5.2	Main effects curve for the mourning dove analysis.	80

Chapter 1

Introduction

The purpose of my dissertation research is to develop a method to estimate and test for trend using ecological data. There are a number of methods already in use for detecting trend in ecological data; however, there is no consensus on the best way to quantify and test for trend (Thomas, [37]). Trend estimation, in particular the significance of trend, may differ depending on what estimation technique one uses. Smoothing methods have recently become popular to describe trend (see Fewster et al. [11], James et al. [17], and Anganuzzi [1]). Smoothing methods provide a flexible tool to describe trend because no prior information about the form of the trend is needed. For a smoothed curve, trend is described by a sequence of smoothed estimates rather than a single parameter such as the slope of a line (Anganuzzi, [1]). Trend is often quantified as the difference between the fitted values of the first few observations and the last few observations. Because this is a relatively new technique for trend detection very little work has been done comparing different methods of trend detection using smoothing techniques. There has been some research done comparing a smoothing method of trend detection to parametric methods such as linear or Poisson regression (see Anganuzzi, [1] and Peterjohn et al., [24]; however, to date there has not been a study comparing different trend detection methods that use smoothing to model trend.

One can think of a model for trend as consisting of the sum of a systematic component and a random component:

$$Z_t = Y_t + U_t \tag{1.1}$$

where Z_t is some population index at time t , U_t is a stochastic component, and Y_t may be a regression equation or smooth curve (Dagum and Dagum, [6]). If a population is stable over time, then Y_t is modeled as a constant term and the stochastic component, U_t , consists of random fluctuations around the mean population index. The random fluctuations in population indexes can be due to such things as unpredictable changes in the environment, autocorrelation or sampling error. If trend is not constant over time then trend may be modeled as a polynomial:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_p t^p. \tag{1.2}$$

A more flexible model for trend than (1.2) is

$$Y_t = f_t \tag{1.3}$$

where f_t is a smooth function of time. There are several options for estimating f_t such as smoothing splines, penalized spline regression, locally weighted regression and kernel smoothers (Fewster et al., [11]). Some of these methods will be discussed later. The advantage of using the trend model (1.3) is that no assumptions are made about the form of the trend. A smoothing parameter is often used to control the “roughness” of the fit. A very rough fit will result in too much of the random fluctuation being modeled while a very smooth fit will not adequately describe the underlying trend. Methods of selecting an appropriate smoothing parameter are discussed later.

Results of my research are applicable to surveys conducted annually at different sites. Many of the examples presented use data from the North American Breeding Bird Survey (BBS). The BBS is a continent-wide bird survey started by Chan Robbins in 1966 [33]. The BBS currently has over 3700 routes and covers all of North America with each route being approximately 24.5

miles long and consisting of stops every .5 miles (Sauer et al., [33]). An observer along a route stops every .5 miles and records all of the birds that he sees or hears. This survey is conducted annually at the height of breeding season and the average number of years of data for a route is 25 years (Sauer et al., [33]). For many of the bird species surveyed, managers do not have any other information about the status of the species and, therefore, many management decisions are based on these surveys (Sauer et al., [32]). However, these surveys, unlike methods such as mark/recapture, produce indexes instead of estimates of population size. In addition to this, other statistical challenges include possible autocorrelation, small sample sizes, variation in observer's abilities and measurement error due to weather.

Methods of data analysis for the BBS with regard to trend estimation and testing have been evolving continually since the beginning (Thomas and Martin, [36]). Although many methods have been compared there is no consensus as to which of the methods should be used in analyzing BBS survey data (Sauer et al., [32]). Analysis of the BBS survey data generally falls into one of three categories. Regression techniques such as least squares regression or Poisson regression is probably the most commonly used technique to analyze BBS data. Linear regression methods can be applied to estimate trend for each route and then a weighted average for routes is used as the estimate of trend. Linear regression has the benefit of producing an estimate of a parameter that serves as the estimate of trend, the slope. However, if the assumptions of linear regression are not satisfied then the slope parameter may be a poor choice to represent trend. A second class of methods are rank models which calculate a statistic based on ranks. A third class of methods are smoothing methods which have also been used to estimate and test for trend. Unlike regression techniques, there is not one or more parameters estimated from the model that serve as an estimate of trend; however, an estimate of trend can be calculated by taking the differences between the fitted values of the first few and the last few observations. These methods along with additional trend detection methods will be introduced in Chapter 3.

There have been some extensive studies done to compare the different methods of trend de-

tection used by the BBS. Thomas and Martin [36] compared two different regression methods and a nonparametric rank method. The regression methods differed with respect to how trend was estimated (Thomas and Martin, [36]). The authors concluded that “the method of trend estimation can affect the magnitude, direction, and statistical significance of population trends assigned to species”. Although the authors concluded that there are differences among the three methods that they compared, they were unable to determine which method gives the most accurate results. They suggested simulating data with known trends and then comparing the results.

The contribution of this work is to develop a method that identifies changes in trend, quantifies trend, and tests for the significance of trend. In addition to this, a comparison of different trend detection techniques will be made by estimating Type I error rates and power under different scenarios. Because autocorrelation is likely to exist in surveys such as the BBS, methods will also be introduced that accommodate serial correlation.

Chapter 2

Spline Regression

This chapter gives a brief introduction to splines and the use of splines in statistics. Spline theory has been used for many years in other areas such as computer graphics, engineering, and automotive science (Lancaster and Salkauskas, [18]). Historically, engineering draftsman used thin pieces of wood called splines to connect points with a smooth curve (Wegmen and Wright, [40]). The splines were “anchored in place by attaching lead weights called ducks at points along the spline” (Wegmen and Wright, [40]). The splines were made to pass through specific points as long as enough ducks were used (Wegmen and Wright, [40]). Suits et al. [35] define spline functions as “a device for approximating the shape of a curvilinear stochastic function without the necessity of pre-specifying the mathematical form of the function.” In statistics, spline regression presents a flexible tool for fitting nonlinear curves using standard regression techniques and has become a widely used method. Spline regression is nonparametric in the sense that there is no specified form to the curve but it is also parametric because linear regression techniques are used to fit the curve once a basis is created. Thus, splines in statistics are often referred to as a semiparametric technique. There are two different approaches to splines in statistics: regression splines and penalized spline regression. In regression splines the amount of smoothing is controlled by the number and placement of knots which separate the

spline functions. In penalized spline regression in which the amount of smoothing is controlled by a smoothing parameter thereby making knot selection less critical. A common criterion for deciding on the amount of smoothing is cross validation which will be discussed later. Penalized spline regression can also be set up as a mixed effects model and the smoothing parameter estimated by maximum likelihood methods.

2.1 Spline functions

Mathematically, splines can be defined as piecewise polynomials of a certain degree connected at join points called knots (Lancaster and Salkauskas[18]). Consider an interval $[a,b]$ and points t_1, t_2, \dots, t_n such that $a < t_1 < t_2 < \dots < t_n < b$. A function, f , is a cubic spline if within each interval $(a, t_1), (t_1, t_2), \dots, (t_n, b)$, f is a cubic polynomial and f, f' and f'' are continuous at t_1, t_2, \dots, t_n (Green and Silverman, [13]). In addition, a cubic spline is a natural cubic spline if the two intervals, (a, t_1) and (t_n, b) , are linear instead of cubic (Green and Silverman, [13]). Thus, a natural cubic spline has the restriction of being linear before the first knot and after the last knot.

2.1.1 Interpolation Splines

Generally in statistics one seeks to describe the underlying pattern of data without modeling too much random fluctuation due to noise. The use of splines in statistics stems from the theory behind interpolating splines. Suppose we have observation pairs $(t_1, z_1), \dots, (t_n, z_n)$ and the goal is to find a function, $f(t)$, that passes through all of these points. There are clearly many ways to do this, the simplest being to connect the points with straight lines. This may be undesirable because the derivatives at these points will be discontinuous. Suppose we restrict $f(t)$ to be a smooth curve with continuous first and second derivatives. There are still many possible curves that will pass through all points. One solution to finding

an optimal function that interpolates the points is to find the “smoothest” curve that passes through all of the points. A common measure of smoothness is the integrated squared second derivative of $f(t)$, $\int_a^b f''(t)^2 dt$ (Green and Silverman, [13]). Green and Silverman [13] state that the integrated squared second derivative of a curve is a “global measure of roughness that has considerable computational advantages”. It can be shown that the curve that minimizes $\int_a^b f''(t)^2 dt$ is a natural cubic spline, and this natural cubic spline is unique (Green and Silverman, [13]).

2.2 Splines in Statistics

The role of spline theory in statistics is to provide a class of functions from which to choose a curve to fit to paired data. The curve should display the primary features of the relationship without reflecting too much of the random fluctuation of the relationship. Smoothing techniques are considered semiparametric in the sense that there is some unspecified smooth function that is estimated from the data whereas a fully parametric technique assumes that the true relationship between two variables follows a specific form (Ruppert et al., [29]). One application of splines in statistics has been to fit smooth spline curves to data using standard regression techniques by creating design matrices whose columns are bases for a vector space.

2.3 Regression splines

Regression splines are piecewise polynomials that connect at join points called knots (Green and Silverman, [13]). Choice of the number and location of knots is an important decision in spline regression. Once the number and location of the knots have been determined, then maximum likelihood techniques can be used to fit the curve. Given K knots, the regression

model of degree one can be represented as follows:

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^K b_k (x - \kappa_k)_+ \quad (2.1)$$

where $\kappa_1, \kappa_2, \dots, \kappa_K$ are the knot locations and $(x - \kappa_i)_+ = x - \kappa_i$ if $x > \kappa_i$ and zero otherwise.

The corresponding basis for (2.1) are the functions:

1 and x .

A design matrix may be constructed for fitting the regression:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & (x_1 - \kappa_1)_+ & \dots & (x_1 - \kappa_K)_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & (x_n - \kappa_1)_+ & \dots & (x_n - \kappa_K)_+ \end{bmatrix}$$

where n is the number of observations. Equation (2.1) is easily extendable to p degrees as follows:

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_p x^p + \sum_{k=1}^K b_k (x - \kappa_k)_+^p. \quad (2.2)$$

and the truncated power functions for (2.2) are written

$$1, x, \dots, x^p, (x - \kappa_1)_+^p, \dots, (x - \kappa_K)_+^p$$

and is known as the truncated power basis of degree p . A spline of degree three is commonly used in statistical applications. Often, it is desirable to have continuous first and second derivatives at the knot locations requiring polynomials of degree at least three and so parsimony leads to degree 3 splines. In addition to this, a measure of roughness is the integrated second derivative so the degree of the spline must be at least three (Green and Silverman, [13]).

Before fitting a regression spline model, the polynomial degree, the number of knots and knot

locations must be determined. It is apparent that these three choices leads to a huge number of candidate models. The graphs presented in Figure 2.1 illustrate the problem with selecting the number and location of knots for a piecewise linear spline. These data consists of counts of the Carolina Wren (*Thryothorus ludovicianus*) from 1966 to 2002 at a Breeding Bird Survey route in Maryland. The winter of 1976-1977 was severe and is thought to have reduced the population size of the Carolina Wren in this area (Link and Sauer, [19]). For each of the graphs, the number of knots is different. Visual inspection of the graphs does not reveal the most appropriate number of knots to choose. In addition to the number and location of knots,

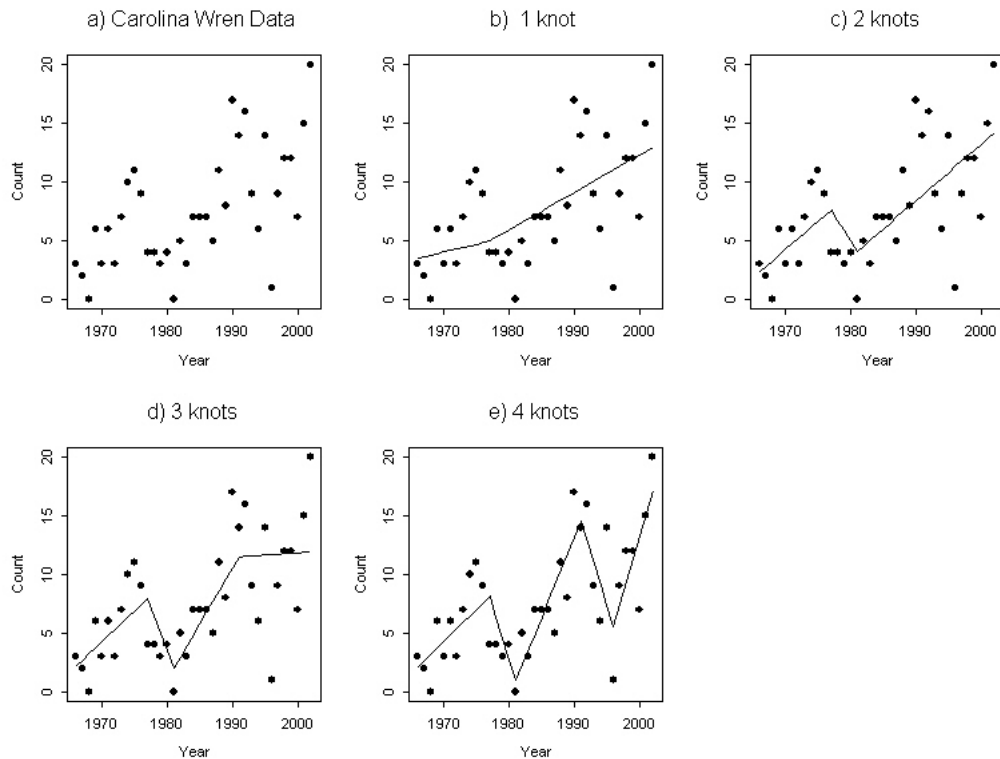


Figure 2.1: Carolina wren data collected on a BBS route. The graphs show linear spline regression fits with 1-4 knots. Graph b) has one knot located at 1977. Graph c) has 2 knots located at 1977 and 1981. Graph d) has knot locations at 1977, 1981 and 1991. Graph e) has knot locations at 1977, 1981, 1991 and 1996

the degree of the polynomial must be determined. Figure 2.2 shows the Carolina wren data with a linear, quadratic and cubic fit. The location of the three knots are 1977, 1981 and

1991. Again, it is not obvious which graph is “best”. One can argue that the quadratic and cubic fit are visually more pleasing than the linear fit, however; the linear fit may be adequate for describing the association.

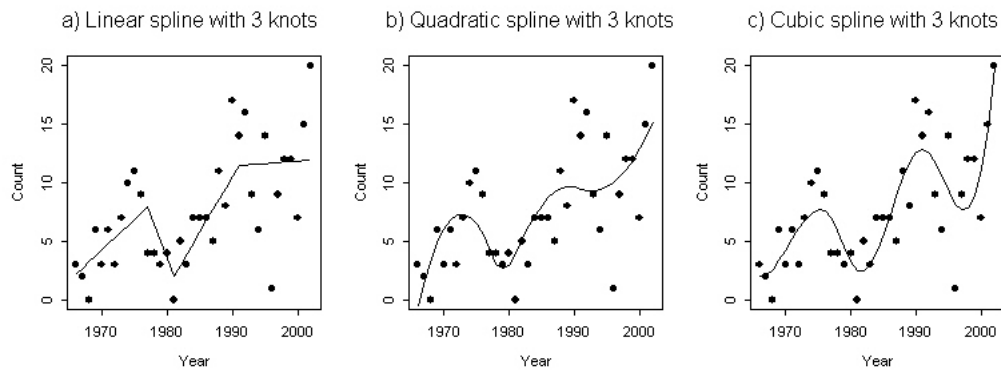


Figure 2.2: Carolina wren data with spline regression fits of degree 1,2 and 3 with knot locations at 1977, 1981 and 1991

2.3.1 Selecting the number and location of knots

Selection of number and location of knots is a well-studied subject. Sometimes there may be a reason for the placement of knots. For instance, we know that the Carolina wren suffered through a severe winter in 1976-1977. A knot could be placed at this year to adjust for the change in growth rate immediately following this winter. However, most times there are no obvious reasons for specific changes in growth rate.

Stepwise procedures are often used to select the number and location of knots. The idea is similar to stepwise regression; however, instead of finding the “best” subset of variables we find the “best” number and location of knots. Wand [38] gives a thorough explanation of the stepwise procedure as well as references to other methods of knot selection. The majority of knot selection methods are quite computer intensive due to the number of candidate models to choose from (Wand, [38]).

2.4 Smoothing Splines

Smoothing splines circumvent the knot selection problem by retaining a large number of knots (Wand, [38]). A smoothing penalty is imposed on the fit in order to avoid overfitting the model resulting in too much of the random fluctuation being modeled. Consider a twice differentiable function, $f(x)$, that minimizes

$$\sum_{i=1}^n \{y_i - f(x_i)\}^2 + \lambda \int_a^b \{f''(t)\}^2 dt \quad (2.3)$$

where λ is a fixed constant, and $a \leq x_1 \leq \dots \leq b$. The criterion (2.3) is called the penalized residual sum of squares (Hastie and Tibshirani, [15]). The first term in (2.3) is reduced when $f(x)$ falls close to the observed values, y_i . The second term in (2.3) is a measure of roughness of the fitted curve and is reduced when the curve is smooth. The integrated squared second derivative is a measure that quantifies the roughness of the curve (Green and Silverman, [13]). Large values of λ produce a smoother curve by forcing $f''(x)$ to be very small for all x while small values of λ result in a rougher curve. This roughness penalty is used often in smoothing splines although other penalties could be used (Eilers and Marx, [8]).

2.4.1 B-splines with difference penalty

The truncated power basis introduced earlier is a popular choice for the basis of a smoothing spline because of its simplicity. B-splines are another popular choice for a basis. De Boor [2] defines k th order B-splines as “appropriately scaled k th divided differences of the truncated power function”. Hence, B-splines can be computed as differences of truncated power bases (Eilers and Marx, [9]). Because the truncated power basis is not orthogonal, the B-spline basis is more numerically stable than the truncated power basis when fitting a curve (Ruppert et al, [29]).

A B-spline basis consists of polynomial pieces connected at join points called knots (Eilers and Marx, [8]). B-spline bases of any order are easy to create because they can be calculated recursively from a B-spline basis of degree zero (Eilers and Marx, [8]). Let $B_j(x; q)$ represent the value at x of the j th B-spline of degree q for a set of knots. A B-spline of degree zero can be created by setting $B_j(x; 0) = 1$ when $(j - 1)dx \leq x - x_{min} < jdx$ and 0 otherwise where $dx = (x_{max} - x_{min})/n'$ and n' is equal to the number of intervals between knots. Once the B-spline bases of degree 0 is created then a recursive formula given by deBorg [2] can be used to create bases of higher degree. The recursive formula for a basis that has equally spaced knots is:

$$B_j(x; q) = \frac{q + p - j + 1}{q} B_{j-1}(x; q - 1) + \frac{j - p}{q} B_j(x; q - 1) \quad (2.4)$$

where $p = x - x_{min}$ and $B_0(x; q) = 0$ for all values of x and q .

For example, suppose that we have observations at $x = 1, 2, \dots, 5$. The number of columns of the B-spline matrix is $n' + q$ where q is equal to the degree of the B-spline matrix, n' is equal to 4 and dx is equal to 1. With four intervals, the number of columns in the B-spline matrix of degree zero is four. When $j = 1$ we have that $B_1(x, 0) = 1$ when $0 \leq x - 1 < 1$ and zero otherwise. Therefore we have that $B_1(1, 0) = 1$ and $B_1(x, 0) = 0$ for $x = 2, 3, 4, 5$. In a similar

manner the remaining columns of the B-spline matrix are calculated:

$$B(x;0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Next, we use the recursive formula (2.4) to obtain the B-spline bases of higher degrees. For $q = 1$ and $j = 1$, formula (2.4) gives

$$B_1(x;1) = xB_0(x;0) + (2-x)B_1(x;0).$$

Recall that $B_0(x;0)$ is equal to zero which gives us

$$B_1(x;1) = (2-x)B_1(x;0).$$

Then

$$B_1(1,1) = 1B_1(x;0) = 1$$

and because $B_1(x,0) = 0$ for $x = 2, 3, 4, 5$ we have that $B_1(x,1) = 0$ for $x = 2, 3, 4, 5$. The other columns can be calculated in a similar manner to obtain:

$$B(x;1) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

For $q = 2$, formula (2.4) is

$$B_j(x; 2) = \frac{3+p-j}{2} B_{j-1}(x; 1) + \frac{j-p}{q} B_j(x; 1).$$

As an example, for $B_2(2, 2)$, $p = 1$ and $j = 2$, and is calculated as follows:

$$B_2(2, 2) = B_1(2; 1) + \frac{1}{2} B_1(2; 1) = .5,$$

and

$$B(x; 2) = \begin{bmatrix} .5 & .5 & 0 & 0 & 0 & 0 \\ 0 & .5 & .5 & 0 & 0 & 0 \\ 0 & 0 & .5 & .5 & 0 & 0 \\ 0 & 0 & 0 & .5 & .5 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 \end{bmatrix}.$$

Finally,

$$B(x; 3) = \begin{bmatrix} .167 & .667 & .167 & 0 & 0 & 0 & 0 \\ 0 & .167 & .667 & .167 & 0 & 0 & 0 \\ 0 & 0 & .167 & .667 & .167 & 0 & 0 \\ 0 & 0 & 0 & .167 & .667 & .167 & 0 \\ 0 & 0 & 0 & 0 & .167 & .667 & .167 \end{bmatrix}.$$

A B-spline of degree 3 with 5 knots will result in a basis matrix with 7 columns (the number of intervals plus the degree) because the B-spline polynomials overlap with neighbor polynomials. An illustration of the B-spline basis of degree three is shown in figure 2.3. Figure 2.3a represents $B_1(x, 3)$ for $1 \leq x \leq 5$. Notice that the value for y is .167 ($B_1(x, 3)$) at $x=1$ and is equal to zero when $x=2$. Figure 2.3b represents $B_2(x, 3)$ and Figure 2.3c represents $B_3(x, 3)$ for $1 \leq x \leq 5$. Finally, Figure 2.3d shows all seven B-splines calculated using formula (2.4) for $1 \leq x \leq 5$.

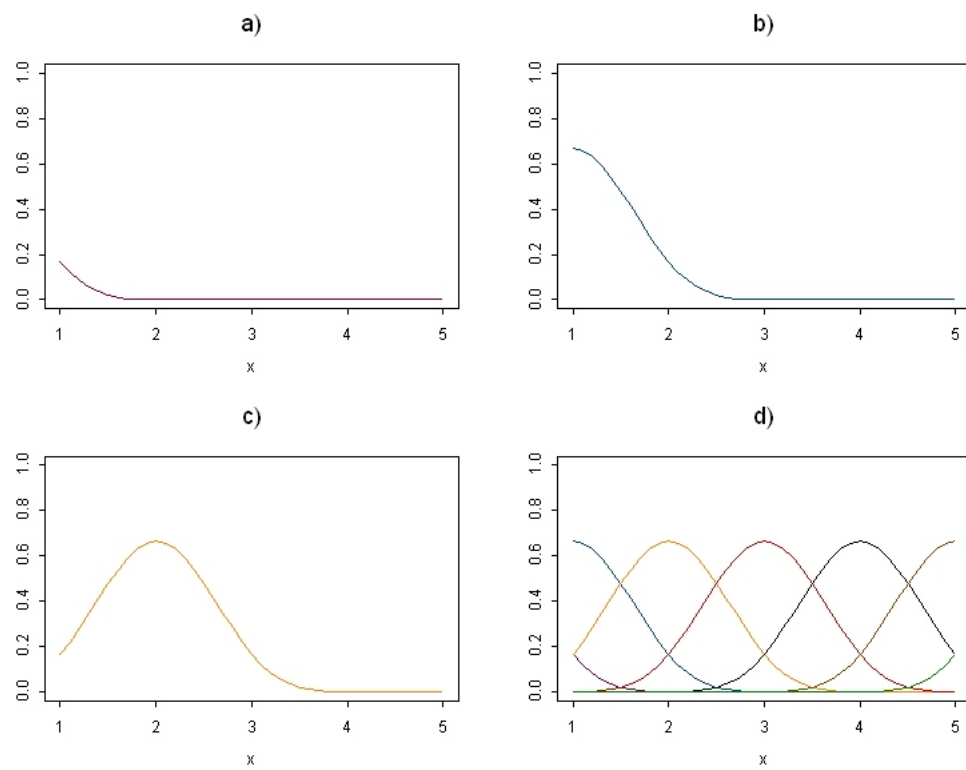


Figure 2.3: B-spline basis of degree 3 with knots at 1,2,3,4,5. A B-spline of degree 3 with 5 knots will have 7 B-splines. a) shows the first B-spline b) shows the second, c) the third, and d) shows all 7 B-splines.

2.4.2 Regression with a B-spline basis

Eilers and Marx [8] proposed using B-splines with a difference penalty instead of the integrated second derivative as a smoothing criteria. Eilers and Marx [8] termed the combination of B-splines with a difference roughness penalty “P-splines”. They proposed using the following objective function on which to base a smooth curve:

$$S = \sum_{i=1}^n \left(y_i - \sum_{j=1}^m a_j B_j(x_i) \right)^2 + \lambda \sum_{j=\kappa+1}^m \left(\Delta^k a_j \right)^2. \quad (2.5)$$

where $B_j(x_i)$ is a B-spline basis and the a_j 's are the m coefficients. The $\Delta^k a_j$'s are the k th order differences of the coefficient a_j . If $k = 1$, then $\Delta a_j = a_j - a_{j-1}$ and for $k = 2$, $\Delta^2 = \Delta \Delta a_j = a_j - 2a_{j-1} + a_{j-2}$ and so forth for larger values of k . Typical values of k are 2 or 3 (Eilers and Marx [8]).

Once the B-spline basis is established then maximum likelihood techniques are used to calculate the coefficients of the smoothed curve. The fitted curve is

$$\hat{\mathbf{y}} = \mathbf{B}\hat{\mathbf{a}} \quad (2.6)$$

where the elements of \mathbf{B} are $b_{ij} = B_j(x_i, q)$ and $\hat{\mathbf{a}}$ are the estimated coefficients. For P-spline regression (B-splines with a difference penalty), the following equation is minimized:

$$S = \|\mathbf{y} - \mathbf{B}\mathbf{a}\|^2 + \lambda \mathbf{D}_k^T \mathbf{D}_k \mathbf{a} \quad (2.7)$$

where \mathbf{D}_k is the matrix representation of the difference operator Δ^k . Taking the first derivative with respect to $\mathbf{a} = (a_1, a_2, \dots, a_m)^T$ we have

$$\frac{\partial S}{\partial \mathbf{a}} = -2\mathbf{B}^T \mathbf{y} + 2\mathbf{B}^T \mathbf{B}\mathbf{a} + 2\lambda \mathbf{D}_k^T \mathbf{D}_k \mathbf{a}.$$

Setting the above equation equal to zero and solving for \mathbf{a} :

$$\hat{\mathbf{a}} = (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{B}^T \mathbf{y}. \quad (2.8)$$

When λ is equal to zero, equation (2.8) reduces to linear regression least squares estimate with design matrix \mathbf{B} . When λ is greater than zero, equation (2.8) has the same form as the ridge regression estimate of the coefficients. Ridge regression was developed by Hoerl and Kennard [16] as an alternative to eliminating regressors when explanatory variables are highly correlated (Ryan, [31]). In least squares regression, the conventional estimator for the coefficients, $\boldsymbol{\beta}$, is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (2.9)$$

Then it can be shown (see Myers and Milton, [23]) that the estimator (2.9) is both unbiased and has minimum variance. However, when there are many correlated predictors, the variance of $\hat{\boldsymbol{\beta}}$ will be large (Marquardt and Snee, [22]). Hoerl and Kennard [16] proposed an estimate for $\boldsymbol{\beta}$ that focused on reducing the mean squared error as the relevant criterion rather than restricting the estimator to be unbiased. The resulting ordinary ridge regression estimator is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X} + kI)^{-1} \mathbf{X}^T \mathbf{y} \quad (2.10)$$

for $k \geq 0$. Values of k , the ridge parameter, between zero and one are explored to find the value that most reduces the mean square error (Marquardt and Snee, [22]).

2.4.3 Degrees of Freedom of a smoother

At first glance it appears that the fit for a smoother is over-parameterized because there are so many parameters. For spline regression with a B-spline basis there will be m estimated parameters, the number of columns of the B-spline matrix. However, a large value of λ has the effect of shrinking the coefficients toward zero (Harrell, [14]). Hastie and Tibshirani

[15] define what is called the equivalent degrees of freedom which “gives an indication of the effective number of parameters used in estimation of a fitted curve.” In linear regression the degrees of freedom is equal to the number of parameters being estimated (Ruppert et al., [29]). The hat or projection matrix in linear regression is defined as

$$\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

so that the vector of fitted values is $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ and \mathbf{X} is the design matrix. In least squares regression, the trace of the hat matrix is equivalent to the number of parameters being estimated. Hastie and Tibshirani [15] define the degrees of freedom of a smoother to be

$$df = tr(\mathbf{S}_\lambda) \tag{2.11}$$

where \mathbf{S}_λ is the smoother matrix for a given value of λ . For P-splines the smoother matrix is

$$\mathbf{S}_\lambda = \mathbf{B} (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{B}^T \tag{2.12}$$

which is akin to $\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ in ordinary least squares regression, and the equivalent degrees of freedom of the fit is equal to

$$tr(\mathbf{S}_\lambda) = tr \left[\mathbf{B} (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{B}^T \right]. \tag{2.13}$$

2.4.4 Smoothing Parameter

The choice of the smoothing parameter determines the amount of smoothing of the fitted curve. Of course, one can choose the amount of smoothing subjectively; however, a less subjective approach is usually desired. A common technique to determine the smoothing parameter is cross validation, a technique used frequently for model selection. The idea behind cross validation is to examine how well the model performs in predicting observations that are

not used in creating the model. Leave-one-out cross validation involves creating n models, $i = 1, \dots, n$, where for each model the i th observation is left out. Each model is then used to predict the observation which was left out in the model formulation. Let $\hat{f}_{-i}(x_i; \lambda)$ be equal to the fitted curve at observation x_i for smoothing parameter λ . Then the cross validation statistic is calculated as

$$CV(\lambda) = \sum_{i=1}^n \left(y_i - \hat{f}_{-i}(x_i; \lambda) \right)^2. \quad (2.14)$$

The value of λ for which equation (2.14) is smallest, $\hat{\lambda}$, is considered the optimal value of λ for the model.

2.4.5 Residual variance

Estimation of the residual variance, $\sigma_\epsilon^2 = \text{Var} = \left[y - \hat{f}(x; \hat{\lambda}) \right]$, can be obtained by calculating the variance of the sum of squares of the residuals around the fitted curve (Green and Silverman, [13]). For parametric regression, the residual sums of squares is divided by the degrees of freedom of the errors (Green and Silverman, [13]). For smoothing splines the residual sum of the errors is divided by the equivalent degrees of freedom as defined in a previous section

$$df_{\text{res}} = \text{tr} \left[\mathbf{I}_n - \mathbf{B} (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{B}^T \right] = n - \text{tr} \left[\mathbf{B} (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{B}^T \right].$$

As the value of λ increases, df_{res} will increase giving a smoother fit.

2.5 Penalized Regression splines

The difference between regression splines and smoothing splines is that for regression splines the number and location of knots must be selected while for smoothing splines a large number

of knots are retained and smoothing is controlled by a smoothing parameter. An advantage of regression splines is that there are far fewer parameters to estimate once the knot locations have been established. However, selecting the number and location of knots can be quite computer intensive (Ruppert, [30]). Penalized regression splines combine regression splines and smoothing splines. Penalized regression spline methodology invokes a smoothing parameter which has the effect of shrinking the coefficients toward zero; however, unlike smoothing splines not all of the knots are retained. Once the number of knots, K , is determined then the knots are placed at equal quantiles throughout the range of x (Ruppert et al., [29]). Via simulation studies, Ruppert [30] concluded that the number of knots is not a crucial decision as long as a sufficient number of knots is chosen. Ruppert [30] showed that once enough knots were used to model the major features of the data then additional knots did not improve substantially the fit of the model.

Consider again the truncated power basis of degree 3:

$$f(x) = \sum_{j=0}^3 \beta_j x^j + \sum_{k=1}^K \beta_{1k} (x - \kappa_k)_+^3 \quad (2.15)$$

where K is equal to the number of knots and κ_i , $i = 1, \dots, K$, are the knot locations. In fitting the above model via least squares we would minimize the sum of squares of the residuals. However, equation (2.15) will be overparameterized with even a moderate number of knots. Similar to smoothing splines, a roughness penalty is added to the usual residual sum of squares. Thus, we find the curve that minimizes

$$S(\beta) = \| \mathbf{y} - \mathbf{X}\beta \|^2 + \lambda \beta^T \mathbf{D}\beta \quad (2.16)$$

for some $\lambda > 0$ where \mathbf{D} is a $(K+p+1)$ diagonal matrix with zeros in the first $p+1$ diagonal positions and 1's in the remaining diagonal positions representing the knots (Ruppert et al., [29]). The effect of the penalty is that the coefficients shrink toward zero according to the value of λ , the smoothing parameter (Ruppert et al., [29]). The quantity in equation (2.16)

is minimized by finding the first derivative with respect to $\boldsymbol{\beta}$, setting the it equal to zero and solving for $\boldsymbol{\beta}$ as follows:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{y}^T \mathbf{X} + 2\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + 2\lambda \mathbf{D} \boldsymbol{\beta}.$$

Setting $\left. \frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0$ leads to

$$-2\mathbf{y}^T \mathbf{X} + 2\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + 2\lambda \mathbf{D} \boldsymbol{\beta} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0$$

$$\Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{D})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.17)$$

and the fitted values,

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X} (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{D})^{-1} \mathbf{X}^T \mathbf{y}.$$

The value for the smoothing parameter, λ , can be selected cross validation.

2.6 Penalized splines as mixed models

Although penalized spline regression does not eliminate the need to select the number and location of knots, it makes the decision less critical as long as there are enough knots to adequately cover the range of the data (Ruppert et al., [29]).

Consider the truncated power basis of degree 3:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=1}^K \beta_k (x - \kappa_k)_+^3 + \epsilon \quad (2.18)$$

where κ_k , $k = 1, \dots, K$, are knot locations. We can fit the above model using ordinary least squares regression. However, too many knots results in modeling too much of the random fluctuation of the data. The solution to reducing the number of knots is to use penalized

spline regression after choosing a smoothing parameter via cross validation. Wand [39] argued that equation (2.18) can be expressed as a mixed effects model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=1}^K u_k (x - \kappa_k)_+^3 + \epsilon \quad (2.19)$$

where the u_k , $k = 1, \dots, K$, are independent $N(0, \sigma_u^2)$, ϵ_i , $i = 1, \dots, n$ are independent $N(0, \sigma_\epsilon^2)$ and the $\{u_k\}$ and $\{\epsilon_i\}$ are mutually independent. Hence the random coefficients, u_k , are assumed to have constant variance, σ_u^2 , and a normal distribution assumption placed on the coefficients. Therefore, the variables u_1, u_2, \dots, u_K are constrained to be small, thereby leading to a smooth fit. We can rewrite equation (2.19) in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \quad (2.20)$$

where

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \mathbf{I}_K & 0 \\ 0 & \sigma_\epsilon^2 \mathbf{I}_n \end{bmatrix} \right),$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix}^T,$$

$$\mathbf{u} = \begin{bmatrix} u_1 & \dots & u_K \end{bmatrix}^T,$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix},$$

and

$$\mathbf{Z} = \begin{bmatrix} (x_1 - \kappa_1)_+^3 & \dots & (x_1 - \kappa_K)_+^3 \\ \vdots & \ddots & \vdots \\ (x_n - \kappa_1)_+^3 & \dots & (x_n - \kappa_K)_+^3 \end{bmatrix}.$$

The fitting for model (2.20) involves the estimation of $\boldsymbol{\beta}$, σ_ϵ^2 and σ_u^2 , and the prediction of

\mathbf{u} . Estimation of $\boldsymbol{\beta}$, σ_ϵ^2 , and σ_u^2 is achieved using maximum likelihood methods while the prediction of \mathbf{u} involves best linear unbiased prediction (BLUP).

The log-likelihood of $(\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_u^2)$ is

$$l(\boldsymbol{\beta}, \mathbf{V}) = -\frac{1}{2}\{n \log(2\pi) + \log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} \quad (2.21)$$

where $\mathbf{V} \equiv \text{Cov}(\mathbf{y}) = \sigma_u^2 \mathbf{Z}\mathbf{Z}^T + \sigma_\epsilon^2 \mathbf{I}$. Differentiating with respect to $\boldsymbol{\beta}$ we get

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{V})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} + 2\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \boldsymbol{\beta}. \quad (2.22)$$

Setting (2.22) equal to zero and solving for $\boldsymbol{\beta}$ gives us

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}) \quad (2.23)$$

and is the same form as the generalized least squares estimator. The profile log-likelihood for \mathbf{V} can be obtained by substituting (2.23) into (2.21) for $\boldsymbol{\beta}$:

$$l_P(\mathbf{V}) = -\frac{1}{2}\{n \log(2\pi) + \log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\}$$

and estimates of \mathbf{V} can be found by maximizing over σ_ϵ^2 and σ_u^2 (Ruppert et al., [29]). In most cases iterative methods are used to solve for the parameter estimates that maximize $l_P(\mathbf{V})$. The Newton-Raphson algorithm is used to maximize the likelihood function in S-plus (Demidenko, [7]).

The BLUP for the random coefficients, \mathbf{u} , can be shown (see Robinson [28]) to be

$$\tilde{\mathbf{u}} = \sigma_\epsilon^2 \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

It can be shown (see Robinson [28]) that the BLUP for (β, \mathbf{u}) can be expressed

$$\begin{bmatrix} \tilde{\beta} \\ \tilde{\mathbf{u}} \end{bmatrix} = (\mathbf{C}^T \mathbf{C} + \lambda \mathbf{D})^{-1} \mathbf{C}^T \mathbf{y}$$

where $\mathbf{C} = [\mathbf{X} \mathbf{Z}]$, $\lambda = \frac{\sigma_\epsilon^2}{\sigma_u^2}$ and $\mathbf{D} = \text{diag}(0, 0, 0, 0, 1, \dots, 1)$.

Hence λ serves as the smoothing parameter in the mixed model framework. The variance components, σ_ϵ^2 and σ_u^2 , are estimated iteratively by using an algorithm such as the Newton-Raphson algorithm.

The number and placement of knots must be decided for the mixed model approach to smoothing. Wand [39] gives the rule of thumb for the number of knots, K , as

$$K = \min \left(\frac{1}{4} \times \text{number of unique } x_i, 35 \right).$$

Knots are then located at the K_k , $k = 1, \dots, K$, sample quantile of the unique x_i where

$$K_k = \left(\frac{k+1}{K+2} \right).$$

One benefit to using the mixed effects representation of the smoothing spline is that maximum likelihood methods can be used to determine the smoothing parameter. Thus, there is no need for a smoothing parameter selection criterion such as cross validation. Furthermore, the mixed model methodology can be used to handle situations such as autocorrelation and measurement error [39].

Chapter 3

Trend Detection

This chapter discusses the importance of trend detection in ecology, examines some of the common methods used for detecting trend, and introduces some new methods.

Trend assessment is an important goal of many ecologists. Interest may lie in assessing the effect of a management decision on some environmental variable or in deciding whether an endangered species has recovered to the point where that species can be taken off the endangered species list. Some goals of an environmental monitoring program include: assessing long-term environmental change, observing responses to changes in management decisions, and recording the rate of change of a population (Goldsmith, [12]). My research focuses on trend estimation and testing using spline regression methods. In particular, I will be developing methods of trend detection that will be useful for studies similar to the North American Breeding Bird Survey (BBS).

Before discussing methods of estimating and detecting trend we must first define what is meant by “trend”. Esterby [10] defines trend as “general direction and tendency” and leaves it to the researcher to provide a more specific definition of trend. Link and Sauer [19] state that the Breeding Bird Survey analysis has been complicated by a failure to define exactly what is

being estimated. They give a definition of trend as the “average rate of change over a specific time interval.” Definitions of trend generally fall into two categories: to describe variation from a constant population over time and to describe the numeric change in population over a specific time period (Link and Sauer, [20]). Link and Sauer [20] suggest reserving the word “trajectory” for describing a pattern of population fluctuation and using the word “trend” to describe an average rate of change over a specific time interval. Trend will be defined for each of the tests that are introduced in this chapter.

One of the goals of my research is to compare different smoothing methods for trend detection. Although different tests for trend have been developed, very little work has been done to compare different trend detection methods that use smoothing techniques to fit the data. The methods currently used to test for trend focus on comparing confidence intervals from the beginning of the time period and end of the time period as a means to test for trend. Other methods exist that could be used to detect trend and may be an improvement over current methods. In addition to this, new methods will be introduced that may perform better than current methods of trend detection in certain situations.

3.0.1 Goals of trend analysis

Some possible goals when analyzing a time series would be the following:

1. Detection of trend
2. Estimation of the magnitude of trend
3. Identification of time periods when change is the greatest
4. Identification of time periods in which there was substantial trend and times in which there was negligible trend.
5. Prediction or forecasting of trend

Because my research focusses on methods that can be applied to the BBS, a trend detection method should also be able to accommodate serial correlation as well as small sample sizes.

My research focuses on finding optimal or near-optimal methods for trend detection when the observations across time are not necessarily linear. To accomplish this I compared different methods of trend detection via simulation. A successful method of trend detection should adhere to the correct Type I error rates and have high power to detect trend. In addition to this, the method for trend detection should have power similar to the power of an appropriate least squares regression approach when the assumptions for least squares regression are satisfied. Because both linear regression and nonparametric correlation coefficients for trend detection are commonly used, these methods will be used as a standard against which to judge other methods. A goal of this research is to find trend estimation methods based on smoothing splines and penalized spline regression that perform comparably to linear regression estimation methods when linear regression is appropriate and perform better than linear regression when linear regression assumptions are not satisfied. In addition to this, the methods should also be easily extended to more complicated situations such as autocorrelated time series.

3.1 Trend detection methods

The remainder of this chapter discusses the trend detection techniques that will be compared. Some of the methods have been used in the past for trend detection, while others are original applications.

3.1.1 Linear Regression

As previously mentioned, linear regression is the simplest and, perhaps, the most popular method to estimate trend when trend is defined as the growth rate of a population. Hence, a

short description of least squares linear regression is given. Let y_i denote the response variable observed at time, t_i . A conventional linear regression model for trend analysis is

$$y_i = \beta_0 + \beta_1 t_i + \epsilon_i \quad (3.1)$$

where $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ are independent for $i = 1, \dots, n$. The parameter β_1 represents the rate of change of y_i with respect to time. The least squares estimator is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

where

$$\mathbf{y} = [y_1, \dots, y_n]^T,$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix},$$

and

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}.$$

A t-test of the null hypothesis $H_0 : \beta_1 = 0$ versus the alternative hypothesis $H_a : \beta_1 \neq 0$ can be obtained from the ratio

$$\frac{\hat{\beta}_1}{\text{SE}(\hat{\beta}_1)}.$$

The standard error of $\hat{\beta}_1$ is given by

$$\text{SE}(\hat{\beta}_1) = \sqrt{\frac{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 t_i)^2}{(n-2) \sum (t_i - \bar{t})^2}}.$$

It can be shown that

$$t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \quad (3.2)$$

has a t -distribution with $n-2$ degrees of freedom when $\beta_1 = 0$.

3.1.2 Nonparametric correlation coefficient

Kendall's τ correlation coefficient is a nonparametric correlation coefficient that is often used when distributional assumptions of the residuals are violated or when there is a nonlinear association between two variables. For this test, trend is defined as a tendency for counts to increase or decrease over time (Thomas, [37]). The null hypothesis, $H_0 : \tau = 0$, versus the alternative hypothesis, $H_1 : \tau \neq 0$ can be tested by calculating the test statistic

$$z = \frac{\tau}{\sqrt{2(2n+5)/9n(n-1)}}.$$

When n is large, z is approximately normal in distribution and, for small n , statistical tables are used to interpolate a p-value [34].

A small example using BBS data illustrates the calculation of τ . The data consist of index counts (y) collected annually (t) for the bird species Veery (*Catharus fuscescens*) over a ten-year period. Figure 3.1 is a scatterplot of count versus year. Table 3.1 gives the data and demonstrates the calculation of Kendall's τ .

To calculate Kendall's tau coefficient, the paired data $\{(t_1, y_1), \dots, (t_n, y_n)\}$ are ordered with respect to one of the variables. For a time series, the data are already ordered by the variable time. Let P_i be the number of observations which exceed the value for, y_i , located to the right of y_i in Table 3.1. Let Q_i equal the number of observations to the right of y_i in Table 3.1

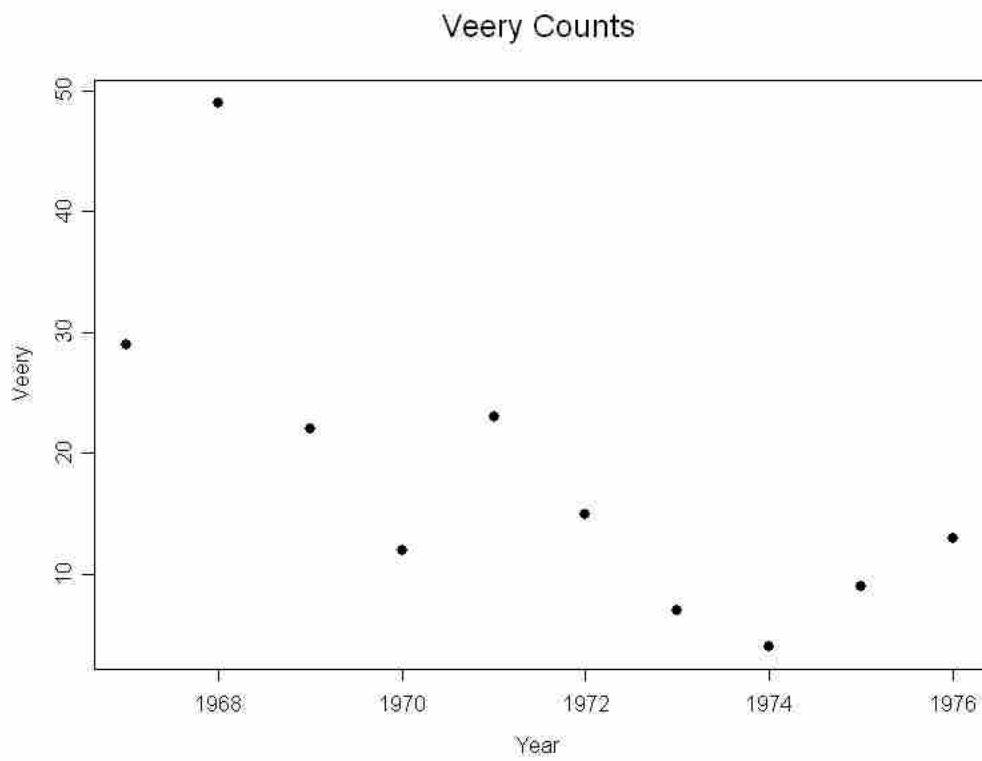


Figure 3.1: Veery counts for 10 years for a route in Maryland.

Year (t)	1	2	3	4	5	6	7	8	9	10
Count (y)	29	49	22	12	23	15	7	4	9	13
P	1	0	1	3	0	0	2	2	1	
Q	8	8	6	3	5	4	1	0	0	
P-Q	-7	-8	-5	0	-5	-4	1	2	1	

Table 3.1: Calculation of Kendall's tau coefficient for the Veery data.

which are less than y_i . S is then defined as

$$S = \sum_{i=1}^{n-1} (P_i - Q_i) = -25.$$

Kendall's τ is calculated as

$$\tau = \frac{2S}{n(n-1)} = \frac{-2 \times 25}{10 \times 9} = -.556$$

The test statistic, z is

$$z = \frac{\tau}{\sqrt{2(2n+5)/9n(n-1)}} = -2.24. \quad (3.3)$$

The p-value is .028 and can be obtained by using statistical tables if the sample size is small as in this case.

3.1.3 LOESS

A method that is currently used to analyze breeding bird survey data is a moving least squares regression smoother (LOESS) developed by Cleveland [4]. The LOESS method finds a fitted value for an observation y_i at t_i by using the values of t that are within an appropriate range of t_i (Ryan, [31]). A least squares regression line is fit within this range and the result is a smooth curve.

James et al. [17] developed a method of testing for trend that is currently used to analyze trend for BBS data. LOESS smoothers are fit to each route for a given species and the trend,

Δ , is estimated for each route as the difference between the average of the first three fitted observations and the average of the last three fitted observations (James et al, [17]). Given some region of interest with M routes, a test for trend is obtained by estimating the regional trend, computing

$$z = \frac{\bar{\Delta}}{\text{SE}(\bar{\Delta})},$$

whose $\text{SE}(\bar{\Delta})$ is an estimate of the standard deviation of the route estimates of Δ (James et al., [17]). James et al. [17] states that z can be approximated by a standard normal distribution. For estimates of trend at a single route, resampling is done using bootstrap techniques to calculate a 95% confidence interval, and the null hypothesis of no trend ($H_0 : \Delta = 0$) versus the two-sided alternative ($H_1 : \Delta \neq 0$) is rejected if the confidence interval does not include zero. The bootstrap procedure is conducted as follows:

1. Estimate Δ from the observed data.
2. Resample the observed paired data with replacement and estimate Δ using these data.
3. Repeat step 2 many times.
4. Calculate the bootstrap standard deviation as the standard deviation of the estimates obtained in step 3.

Assuming the bootstrap distribution for $\hat{\Delta}$ is approximately normal, a 95% confidence interval is obtained by computing

$$\hat{\Delta} \pm 1.96 (\text{Bootstrap standard deviation}). \quad (3.4)$$

3.1.4 Likelihood Ratio Test

Recall from Chapter 2 the linear mixed effects model used for smoothing

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$$

where

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \mathbf{I}_n & 0 \\ 0 & \sigma_\epsilon^2 \mathbf{I}_K \end{bmatrix} \right)$$

and

$$\mathbf{Z} = \begin{bmatrix} (t_1 - \kappa_1)_+ & \dots & (t_1 - \kappa_K)_+ \\ \vdots & \ddots & \vdots \\ (t_n - \kappa_1)_+ & \dots & (t_n - \kappa_K)_+ \end{bmatrix}.$$

Recall that $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ and $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I})$. The above model can be written

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}^*$$

where $\boldsymbol{\epsilon}^* = \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$ and $\text{cov}(\boldsymbol{\epsilon}^*) = \mathbf{V} = \mathbf{Z}^T \mathbf{Z} \sigma_u^2 + \sigma_\epsilon^2 \mathbf{I}$. Because $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, the log likelihood is as follows:

$$l(\boldsymbol{\beta}, \mathbf{V}) = -\frac{1}{2} \{ n \log(2\pi) + \log |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \}. \quad (3.5)$$

Consider the model

$$f(t) = \beta_0 + \beta_1 t + \sum_{k=1}^K u_k (t - \kappa_k)_+. \quad (3.6)$$

Ruppert et al. [29] suggest a test for an association between two variables using the above

model. The null and alternative hypotheses are

$$H_0 : \beta_1 = \sigma_u^2 = 0$$

versus

$$H_1 : \beta_1 \neq 0 \text{ or } \sigma_u^2 > 0.$$

The likelihood ratio test statistic is

$$-2 \log LR(\mathbf{y}) = -2 \left\{ l\left(\hat{\beta}_{0,0}, 0, 0, \hat{\sigma}_{\epsilon,0}^2; \mathbf{y}\right) - l\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2, \hat{\sigma}_\epsilon^2; \mathbf{y}\right) \right\} \quad (3.7)$$

where $(\hat{\beta}_{0,0}, \hat{\sigma}_{\epsilon,0}^2)$ minimizes $-2l(\hat{\beta}_0, 0, 0, \hat{\sigma}_{\epsilon,0}^2; \mathbf{y})$, and $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2, \hat{\sigma}_\epsilon^2)$ minimizes $-2l(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2, \hat{\sigma}_\epsilon^2; \mathbf{y})$.

Under the null hypothesis

$$\hat{\sigma}_{\epsilon,0}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$

and

$$\hat{\beta}_0 = \bar{y}.$$

If the test statistic is in the upper tail of the null distribution then we can conclude that $E(y)$ is not constant with respect to t . Ruppert et al. [29] report that standard asymptotic assumptions on which such likelihood tests are commonly based are not met and suggest using Monte Carlo simulations to approximate the null distribution of the likelihood ratio test statistic. The reason standard asymptotic assumptions fail is because for mixed effects models, \mathbf{y} cannot be partitioned into independent subvectors (Ruppert et al, [29]). In the Monte Carlo simulations, a large number of data sets are generated with the parameters equal to their estimates under the null hypothesis. For each simulation, the likelihood ratio statistic, formula (5.1), is computed to create an approximation of the null distribution. The observed trend is judged to be significant if the likelihood ratio test statistic is in the upper tail of the approximated null distribution. This test can be used to detect departures from a constant mean. Therefore, this test may yield significant results even if the fitted values are

similar between the beginning of the time period and the end of the time period and trend can be defined as variation from a constant mean.

3.1.5 Examining first derivatives

For simple linear regression an estimate for trend is the slope coefficient. When the rate of change is constant then the slope suffices as a measure of trend; however, more often the rate of change is not constant. For the case where the rate of change is not constant, examination of first derivatives of a fitted curve can yield useful information about trend. For instance, if trend is not constant, then it may be useful to identify periods in which the rate of change has increased or decreased.

Recall that trend can generally be defined two ways: to describe variation from a constant population over time and to describe the numeric change in population over a specific time period (Link and Sauer, [20]). The benefit of looking at first derivatives is that one can address both definitions. A test that quantifies the change in population size can be developed by looking at some numeric summary of the derivatives. In addition to this, if there are times in which the first derivative is significantly different from zero then one can conclude that there is variation from a constant population at those times.

3.1.6 Test for trend using first derivatives

Examining first derivatives can be used as a method to quantify the magnitude of the change in growth of a population as well as test the significance of this change. I propose a test statistic that uses the sum of the first derivatives over each time interval as an estimator of trend.

Consider the least squares estimate of the curve

$$\hat{f}(t) = \mathbf{X}\boldsymbol{\beta}.$$

Then an estimate of the first derivative of $f(t)$ with respect to t is

$$\hat{f}'(t) = \mathbf{X}'\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \hat{\boldsymbol{\beta}}.$$

Let

$$\mathbf{j}^T = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}.$$

Then the sum of the first derivatives with respect to t is

$$\sum_{i=1}^n \hat{f}'(t_i) = \mathbf{j}^T \mathbf{X}' \hat{\boldsymbol{\beta}} = n\hat{\beta}_1. \quad (3.8)$$

The variance is calculated as

$$\text{var}(\mathbf{j}^T \mathbf{X}' \hat{\boldsymbol{\beta}}) = \text{var}(\mathbf{j}^T \mathbf{X}' (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}) \quad (3.9)$$

$$= \sigma_\epsilon^2 \mathbf{j}^T \mathbf{X}' (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{j}^T \mathbf{X}')^T \quad (3.10)$$

Let

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{bmatrix}.$$

Then

$$\text{var}(\mathbf{j}^T \mathbf{X}' \boldsymbol{\beta}) = \sigma_\epsilon^2 n^2 c_{11}$$

and

$$t = \frac{n\hat{\beta}_1}{\sqrt{n^2 \hat{\sigma}_\epsilon^2 c_{11}}} = \frac{\hat{\beta}_1}{\hat{\sigma}_\epsilon \sqrt{c_{11}}} = \frac{\hat{\beta}_1}{\text{SE}(\hat{\beta}_1)}.$$

This t is the same as the test statistic when testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$.

3.1.7 First derivatives for penalized spline regression using a linear mixed model

Consider again the fitted mixed model representation of the penalized spline regression of degree 3:

$$\hat{f}(t) = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\beta}_3 t^3 + \sum_{k=1}^K \hat{u}_k (t - \kappa_k)_+^3. \quad (3.11)$$

The first derivative with respect to t is:

$$\hat{f}'(t) = \hat{\beta}_1 + 2\hat{\beta}_2 t + 3\hat{\beta}_3 t^2 + 3 \sum_{k=1}^K \hat{u}_k (t - \kappa_k)_+^2. \quad (3.12)$$

Let

$$\mathbf{X}' = \begin{bmatrix} 0 & 1 & 2t_1 & 3t_1^2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 2t_n & 3t_n^2 \end{bmatrix}$$

and

$$\mathbf{Z}' = \begin{bmatrix} 3(t_1 - \kappa_1)_+^2 & \dots & 3(t_1 - \kappa_K)_+^2 \\ \vdots & \ddots & \vdots \\ 3(t_n - \kappa_1)_+^2 & \dots & 3(t_n - \kappa_K)_+^2 \end{bmatrix}.$$

Then the fitted first derivatives are calculated as

$$\hat{\mathbf{f}}' = \mathbf{X}'\hat{\boldsymbol{\beta}} + \mathbf{Z}'\hat{\mathbf{u}}.$$

Let $\mathbf{C}' = [\mathbf{X}' \ \mathbf{Z}']$, then

$$\hat{\mathbf{f}}' = \mathbf{C}' \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \mathbf{C}' \left(\mathbf{C}'^T \mathbf{C}' + \frac{\hat{\sigma}_\epsilon^2}{\sigma_u^2} \mathbf{D} \right)^{-1} \mathbf{y}.$$

where $\mathbf{C} = [\mathbf{X} \ \mathbf{Z}]$ and \mathbf{D} is a $(K+4) \times (K+4)$ matrix, $\text{diag}(0, 0, 0, 0, 1, \dots, 1)$. Ruppert et al. [29] give the variance of estimated coefficients as

$$\widehat{\text{cov}} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \hat{\sigma}_\epsilon^2 \left(\mathbf{C}^T \mathbf{C} + \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_u^2} \mathbf{D} \right)^{-1}$$

The sum of the first derivatives with respect to t can be expressed as a linear combination of \mathbf{y} :

$$\sum_{i=1}^n \hat{f}'(t_i) = \mathbf{j}^T \mathbf{C}' \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \mathbf{j}^T \mathbf{C}' \left(\mathbf{C}^T \mathbf{C} + \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_u^2} \mathbf{D} \right)^{-1} \mathbf{C}^T \mathbf{y}. \quad (3.13)$$

If normally distributed errors are assumed then the test statistic

$$z = \frac{\mathbf{j}^T \mathbf{C}' \left(\mathbf{C}^T \mathbf{C} + \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_u^2} \mathbf{D} \right)^{-1} \mathbf{C}'^T \mathbf{y}}{\hat{\sigma}_\epsilon \sqrt{\mathbf{j}^T \mathbf{C}' \left(\mathbf{C}^T \mathbf{C} + \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_u^2} \mathbf{D} \right)^{-1} (\mathbf{j}^T \mathbf{C}')^T}} \quad (3.14)$$

is asymptotically normal in distribution with mean of 0 and variance of 1. For small sample sizes, the distribution of z can be estimated by a t -distribution with degrees of freedom equal to the residual degrees of freedom.

3.1.8 First derivatives for B-splines

De boor [2] gives the following formula for the first derivative of a B-spline curve, $b(t)$, for a given value of t :

$$b'(t) = \sum_j a_j B'_{j+1}(t; q) = \sum_j \Delta a_{j+1} B_j(t; q-1) \quad (3.15)$$

where the a_j 's are the B-spline coefficients, q is the degree of the B-spline, $\Delta a_{j+1} = a_{j+1} - a_j$ and $B_j(t; q)$ is the value at t of the j th B-spline of degree q .

Equation (3.15) gives us a simple way to calculate the first derivatives using first order differ-

ences of the estimated coefficients. Recall that \mathbf{B} is defined such that the elements of \mathbf{B} are $b_{ij} = B_j(t_i, q)$. Let \mathbf{B}_{-1} be defined as $B_j(t, q - 1)$ which is the B-spline matrix of degree $q-1$. Then using formula (3.15), the sum of the first derivatives with respect to t of a fitted curve of degree q is calculated as

$$\sum_{i=1}^n \hat{f}'(t_i) = \mathbf{j} \mathbf{B}_{-1} \Delta \hat{\mathbf{a}} \quad (3.16)$$

where $\Delta \hat{\mathbf{a}}$ is the vector of \hat{a}_j differences. Define

$$\mathbf{A}_1 = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & -1 & 1 \end{bmatrix}.$$

Then $\Delta \hat{\mathbf{a}} = \mathbf{A}_1 \hat{\mathbf{a}}$. Recall from Chapter 2 that

$$\hat{\mathbf{a}} = (\mathbf{B} \mathbf{B}^T + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{B}^T \mathbf{y}$$

where \mathbf{D}_k is the matrix representation of the difference operator Δ^k (Eilers and Marx, [8]). If it is assumed that the errors are independent and normally distributed with mean 0 and variance σ_ϵ^2 , then $\mathbf{y} \sim N(\mathbf{B} \mathbf{a}, \sigma_\epsilon^2 \mathbf{I})$. The sum of the first derivatives with respect to t can be expressed as

$$S(t) = \mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 \hat{\mathbf{a}} = \mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 (\mathbf{B} \mathbf{B}^T + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{B}^T \mathbf{y} \quad (3.17)$$

Because (3.17) is a linear function of \mathbf{y} , we have that

$\mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 \mathbf{a} \sim N(\mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 \mathbf{a}, \text{var}(\mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 \mathbf{a}))$. The estimated variance of $S(t)$ is calculated as

$$\begin{aligned} \widehat{\text{var}}(\mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 \hat{\mathbf{a}}) &= \mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 \widehat{\text{var}}(\hat{\mathbf{a}}) (\mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1)^T \\ &= \hat{\sigma}_\epsilon^2 \mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} (\mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1)^T. \end{aligned} \quad (3.18)$$

A test statistic for the null hypothesis that $H_0 : S(t) = 0$ versus $H_1 : S(t) \neq 0$ is calculated

as

$$z = \frac{\mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 \hat{\mathbf{a}}}{\hat{\sigma}_\epsilon \sqrt{\mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1 (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} (\mathbf{j}^T \mathbf{B}_{-1} \mathbf{A}_1)^T}}. \quad (3.19)$$

Because the sum of the derivatives is a linear combination of y which is asymptotically normal, z is asymptotically $N(0,1)$ under the null hypothesis. For small sample sizes, the distribution of z can be estimated by a t -distribution with degrees of freedom equal to the residual degrees of freedom.

Randomization tests

The assumptions that have been made with respect to the distribution of the test statistics obtained from either smoothing splines or spline regression fits have relied on the notion that the smoother matrix is a linear function of the data vector \mathbf{y} . However, the smoother matrix depends on \mathbf{y} through a smoothing parameter that must be estimated (Ruppert et al., [29]). Ruppert et al. [29] state that “it is common practice to pretend the data-based smoothing parameter is fixed, and, as an approximation, to treat the smoother as linear.” For this reason, a randomization test procedure is used for the test involving the sum of the first derivatives. A randomization test is a nonparametric method used to compute the significance of a test statistic when distributional assumptions about the test statistic are not known (Sokal and Rohlf, [34]). Randomization methods allow a researcher to provide an empirical sampling distribution of a statistic, and then determine an observed significance assuming the null hypothesis is true (Manly, [21]). This allows one to test whether the observed trend is consistent with the null hypothesis of no trend.

The randomization test procedure is conducted with the following steps:

1. Calculate the observed test statistic for the paired observations $\{(t_1, y_1), \dots, (t_n, y_n)\}$.
2. Randomly permute the order of y_i , $i = 1, \dots, n$ while retaining the original order of

t_1, \dots, t_n to obtain a new set of observation pairs.

3. Refit the model including the estimation of the smoothing parameter, λ . Calculate the test statistic for the paired observations obtained in step 2.
4. Repeat steps 2 and 3 a large number of times and construct an empirical sampling distribution for the test statistic under the null hypothesis.
5. If the observed value of the test statistic (from step 1) is above the .975 or below the .025 percentile, then it is concluded that there evidence of an association between t and y . The size of this test is .05.

3.2 Tests for trend with autocorrelated errors

Independence of errors is an important assumption for all of the methods that have been considered. It is often the case when analyzing time series data that errors close in time tend to be more similar than errors further away in time.

3.2.1 Generalized linear regression with autoregressive errors

Reinsel et al. [27] developed a generalized linear regression method that assumes the error structure follows an autoregressive process. We will consider only the AR(1) process here. For an AR(1) process, the model for the residuals, ϵ_i , $i = 1, \dots, n$, is

$$\epsilon_i = \phi\epsilon_{i-1} + \xi_i \tag{3.20}$$

where $\phi \in [0, 1)$ is an autocorrelation parameter, and ξ_i are iid $N(0, \sigma_\epsilon^2)$. Maximum likelihood estimates of the regression coefficients are

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\mathbf{X}^T \hat{\mathbf{R}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{R}}^{-1} \mathbf{y} \quad (3.21)$$

where

$$\hat{\mathbf{R}} = \frac{\hat{\sigma}_\epsilon^2}{1 - \hat{\phi}^2} \begin{bmatrix} 1 & \hat{\phi} & \dots & \hat{\phi}^{n-2} & \hat{\phi}^{n-1} \\ \hat{\phi} & 1 & \hat{\phi} & \dots & \hat{\phi}^{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \hat{\phi}^{n-1} & \hat{\phi}^{n-2} & \dots & \hat{\phi} & 1 \end{bmatrix}. \quad (3.22)$$

The standard error of $\hat{\beta}_1$ is estimated by the square root of the (2,2) entry of the estimated covariance matrix:

$$\widehat{\text{cov}} \left(\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \right) = (\mathbf{X}^T \hat{\mathbf{R}}^{-1} \mathbf{X})^{-1}. \quad (3.23)$$

A 95% confidence interval for β_1 is

$$\hat{\beta}_1 \pm z_{.975} \sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}.$$

If 0 is not contained in the interval then the null hypothesis of $H_0 : \beta_1 = 0$ is rejected in favor of the alternative hypothesis $H_1 : \beta_1 \neq 0$.

3.2.2 Sum of the first derivatives with autocorrelated errors

An advantage of using the linear mixed effects model for smoothing is the ease with which serial correlation can be incorporated into the error structure. Recall the linear mixed effects model discussed in Chapter 2:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \quad (3.24)$$

where

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \mathbf{I}_K & 0 \\ 0 & \sigma_\epsilon^2 \mathbf{I}_n \end{bmatrix} \right)$$

and

$$\mathbf{Z} = \begin{bmatrix} (t_1 - \kappa_1)_+ & \dots & (t_1 - \kappa_K)_+ \\ \vdots & \ddots & \vdots \\ (t_n - \kappa_1)_+ & \dots & (t_n - \kappa_K)_+ \end{bmatrix}.$$

I propose modifying equation (3.24) to model AR(1) errors if evidence of serial correlation exists. If an AR(1) model is assumed, the residuals follow the model given by equation (3.20).

The linear mixed effects model with correlated errors is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \quad (3.25)$$

where

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \mathbf{I} & 0 \\ 0 & \mathbf{R} \end{bmatrix} \right)$$

where

$$\mathbf{R} = \frac{\sigma_\epsilon^2}{1 - \phi^2} \begin{bmatrix} 1 & \phi & \dots & \phi^{n-2} & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \dots & \phi & 1 \end{bmatrix}. \quad (3.26)$$

Variance components, σ_ϵ^2 and σ_u^2 , and the autocorrelation parameter, ϕ , are estimated using maximum likelihood methods (Coull et al., [5]). See Chi et al. [3] for details on estimating the linear mixed effects model with AR(1) errors. The test of the null hypothesis $H_0 : \sum_{i=1}^n f'(t_i) = 0$ versus the alternative hypothesis $H_1 : \sum_{i=1}^n f'(t_i) \neq 0$ is easily modified to accommodate serial correlation in the error structure in the linear mixed effects model. The

fitted first derivatives are calculated as

$$\hat{\mathbf{y}}' = \mathbf{X}'\hat{\boldsymbol{\beta}} + \mathbf{Z}'\hat{\mathbf{u}}.$$

Ruppert et al. [29] give the variance of estimated coefficients as

$$\widehat{\text{cov}} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \left(\mathbf{C}^T \hat{\mathbf{R}}^{-1} \mathbf{C} + \frac{1}{\hat{\sigma}_u^2} \mathbf{D} \right)^{-1}$$

The sum of the first derivatives is calculated

$$\sum_{i=1}^n \hat{f}(x_i) = \mathbf{j}^T \mathbf{C}' \left(\mathbf{C}^T \hat{\mathbf{R}}^{-1} \mathbf{C} + \frac{1}{\hat{\sigma}_u^2} \mathbf{D} \right)^{-1} \hat{\mathbf{R}}^{-1} \mathbf{y},$$

and

$$\text{var} \left(\sum_{i=1}^n \hat{f}(x_i) \right) = \mathbf{j}^T \mathbf{C}' \left(\mathbf{C}^T \hat{\mathbf{R}}^{-1} \mathbf{C} + \frac{1}{\hat{\sigma}_u^2} \mathbf{D} \right)^{-1} (\mathbf{j}^T \mathbf{C}')^T$$

The test statistic for the sum of the first derivatives with AR(1) errors is

$$z = \frac{\mathbf{j}^T \mathbf{C}' \left(\mathbf{C}^T \hat{\mathbf{R}}^{-1} \mathbf{C} + \frac{1}{\hat{\sigma}_u^2} \mathbf{D} \right)^{-1} \mathbf{C}'^T \hat{\mathbf{R}}^{-1} \mathbf{y}}{\sqrt{\mathbf{j} \mathbf{C}' \left(\mathbf{C}^T \hat{\mathbf{R}}^{-1} \mathbf{C} + \frac{1}{\hat{\sigma}_u^2} \mathbf{D} \right)^{-1} (\mathbf{j} \mathbf{C}')^T}} \quad (3.27)$$

A randomization test will be used to determine the significance of the test statistic (3.27).

The randomization test procedure is conducted with the following steps:

1. Calculate the observed test statistic.
2. Randomly permute the order of y_i , $i = 1, \dots, n$ while retaining the original order of t_1, \dots, t_n to obtain a new set of observation pairs.
3. Fit the model and estimate parameters using maximum likelihood methods.
4. Calculate the test statistic.

5. Repeat steps 2 through 4 a large number of times and construct an empirical sampling distribution for the test statistic under the null hypothesis.
6. If the observed value of the test statistic (from step 1) is not bracketed by the .975 and the .025 quantiles, then it is concluded that there is evidence of an association between t and y . The size of this test is .05.

3.2.3 Likelihood ratio test with correlated errors

The likelihood ratio test for the null hypothesis $H_0 : \beta_1 = \sigma_u^2 = 0$ versus the alternative $H_1 : \beta \neq 0$ or $\sigma_u^2 > 0$ can also be extended to accommodate serially correlated errors. The model for the likelihood ratio test is a linear mixed effects model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}.$$

The likelihood ratio test statistic for trend detection in the presence of autocorrelated errors is

$$-2 \log LR(\mathbf{y}) = -2 \left\{ l\left(\hat{\beta}_{0,0}, 0, 0, \hat{\sigma}_{\epsilon,0}^2, \hat{\phi}_{0,0}; \mathbf{y}\right) - l\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2, \hat{\sigma}_{\epsilon}^2, \hat{\phi}; \mathbf{y}\right) \right\} \quad (3.28)$$

where $\left(\hat{\beta}_{0,0}, \hat{\sigma}_{\epsilon,0}^2, \hat{\phi}_{0,0}\right)$ minimizes $-2l\left(\hat{\beta}_{0,0}, 0, 0, \hat{\sigma}_{\epsilon,0}^2, \hat{\phi}_{0,0}; \mathbf{y}\right)$ and $\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2, \hat{\sigma}_{\epsilon}^2, \hat{\phi}\right)$ minimizes $-2l\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2, \hat{\sigma}_{\epsilon}^2, \hat{\phi}; \mathbf{y}\right)$.

If we assume $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, then the log likelihood is:

$$l(\boldsymbol{\beta}, \mathbf{V}; \mathbf{y}) = -\frac{1}{2} \{ n \log(2\pi) + \log |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \}. \quad (3.29)$$

where

$$\mathbf{V} = \sigma_u^2 \mathbf{Z}\mathbf{Z}^T + \mathbf{R}.$$

The errors $\epsilon_i, i = 1, \dots, n$, are assumed to follow an AR(1) process and \mathbf{R} is defined as in

equation (3.26). Monte Carlo simulations will be used to calculate the observed significance of the test statistic given in (3.29). The Monte Carlo simulation procedure is conducted using the following steps:

1. Calculate the observed likelihood ratio test statistic using formula (3.29).
2. Data are then simulated by fixing the parameters equal to their specified values under the null hypothesis and estimating the model for these data. The parameters β_1 and σ_u^2 are set to zero and the other parameters are set to the estimated values.
3. The likelihood ratio test statistic is calculated for step 2.
4. Repeat steps 2 and 3 numerous times and construct an empirical sampling distribution for the test statistic under the null hypothesis.
5. The observed value of the test statistic is significant at the α -level of .05 if it falls in the upper 95 percentile of the distribution constructed in step 4.

3.3 Example

To illustrate the trend detection methods, an example using Wolverine (*Gulo gulo*) data is given. The data were obtained from the website, <http://cpbnts1.bio.ic.ac.uk> [26]. The wolverine count data were collected from 1752 to 1908 by the Hudson's Bay Company. Figure 3.2 shows a scatterplot of these data along with five different regression fits. A least squares regression fit is clearly not appropriate to describe the association between time, t , and count, y . The other smoothing techniques appear to fit the data adequately. The linear regression spline fit via a linear mixed effects model appears to be adequate in describing the overall tendency of the relationship.

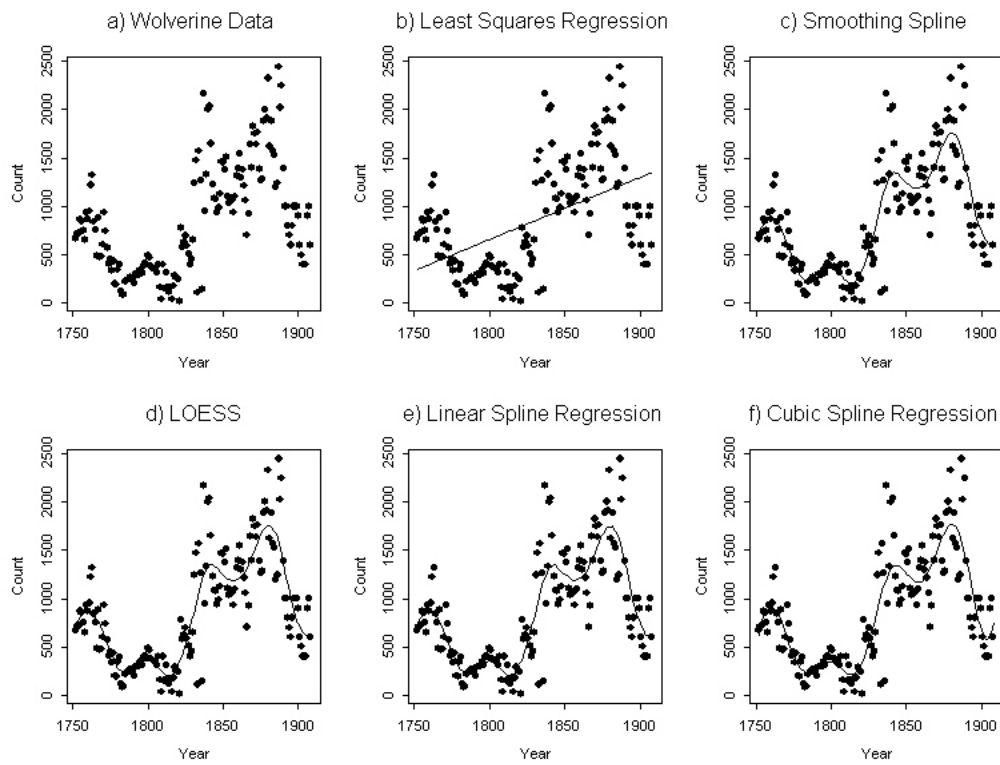


Figure 3.2: Wolverine counts from 1752 to 1908. Five different regression fits are shown on individual graphs: b) Least squares regression, c) locally weighted regression (LOESS), d) cubic spline using a B-spline basis, e) a linear spline using a linear mixed effects model, f) a cubic spline using a linear mixed effects model

3.3.1 Least squares regression

The null and alternative hypothesis for trend in least squares regression are

$$H_0 : \beta_1 = 0$$

versus

$$H_1 : \beta_1 \neq 0.$$

The estimated value for β_1 was 6.42 with a standard error of the estimate of .861. The p-value for this test was <.0001, giving strong evidence in favor of the alternative hypothesis and indicating positive trend. A 95% confidence interval for β_1 is (4.72,8.12).

3.3.2 LOESS

The method of trend detection that is currently in use for BBS data analysis is

$$H_0 : \Delta = 0$$

$$H_1 : \Delta \neq 0$$

where $\hat{\Delta}$ is calculated by taking the difference of the first three fitted values and the last three fitted values of the LOESS fit. The standard error for $\hat{\Delta}$ is calculated by bootstrapping the data as described in section 3.1.3. For the wolverine data, $\hat{\Delta}$ was 166.5 (SE=123). A 95% confidence interval for the difference between the last three fitted values and the first three fitted values is (-37.13,457.67). Because the 95% confidence interval includes zero, we have no evidence in favor of the alternative hypothesis.

3.3.3 Kendall's τ

The test involving the nonparametric correlation coefficient, Kendall's τ , is

$$H_0 : \tau = 0$$

$$H_1 : \tau \neq 0.$$

For the wolverine example, $\hat{\tau}$ was .310. The p-value was <.0001 which gives us strong evidence for the alternative hypothesis and indicates positive trend.

3.3.4 Likelihood ratio test

The null and alternative hypotheses for trend with the likelihood ratio test are

$$H_0 : \beta_1 = \sigma_u^2 = 0$$

$$H_1 : \beta_1 \neq 0 \text{ or } \sigma_u^2 > 0.$$

The likelihood ratio test statistic for the above test was 163.3 which gives us strong evidence (p-value<.0001) for the alternative hypothesis. The significant results reflect that the wolverine counts are not constant with respect to time but do not necessarily imply an upward or downward trend over the time period.

3.3.5 Sum of the first derivatives

The test involving the sum of the first derivatives is carried out using a smoothing spline with a B-spline basis (P-spline) and, also, the linear mixed model for spline regression. One difference between these two methods is the choice of the number of knots, K , and the placement of

knots. However, because n is fairly large for this example, a reduced number of knots is used for both fits. The other difference between the two methods is the estimation of the smoothing parameter, λ . P-splines use a difference penalty and cross-validation to determine the amount of smoothing of a fitted curve (Eilers and Marx, [9]). The spline regression fit using the linear mixed effects model estimates the smoothing parameter as $\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\epsilon^2}$, and maximum likelihood methods are used to estimate the variance components (Ruppert et al., [29]). Because no closed-form solution is available, an algorithm such as the EM algorithm must be used to estimate the variance components in the linear mixed effects model. For the wolverine data, the fit for the smoothing curve and the cubic spline regression curve appear similar. For both methods, the null hypothesis and alternative hypothesis are

$$H_0 : S(t) = 0 \text{ versus } H_1 : S(t) \neq 0$$

where $S(t) = \sum_{i=1}^n f'(t_i)$ and $f'(t_i)$ is the first derivative of the fitted curve at $t_i, i = 1, \dots, n$. For P-splines, $\hat{S}(t)$ was -474.96 (SE=649.52). The p-value for this test was .466 which gives no evidence for H_1 , that the sum of the first derivatives is not equal to zero. For the spline regression method using the linear mixed effects model, $\hat{S}(t)$ was 7.47 (SE=240.1). Again, this gives us no evidence in favor of the alternative hypothesis of trend.

3.3.6 Discussion

This example illustrates the differences between the trend detection methods. Both linear regression and Kendall's τ were highly significant and had positive trend estimates. Because the association is nonlinear and not monotone, this example is not appropriate for either linear regression or Kendall's τ . Due to the large number of species and routes in surveys such as the BBS, a trend detection method should be appropriate even when the trend is nonlinear and not monotone. The smoothing methods introduced in this chapter appear to provide a reasonable fit to these data. The LOESS method of trend detection and the

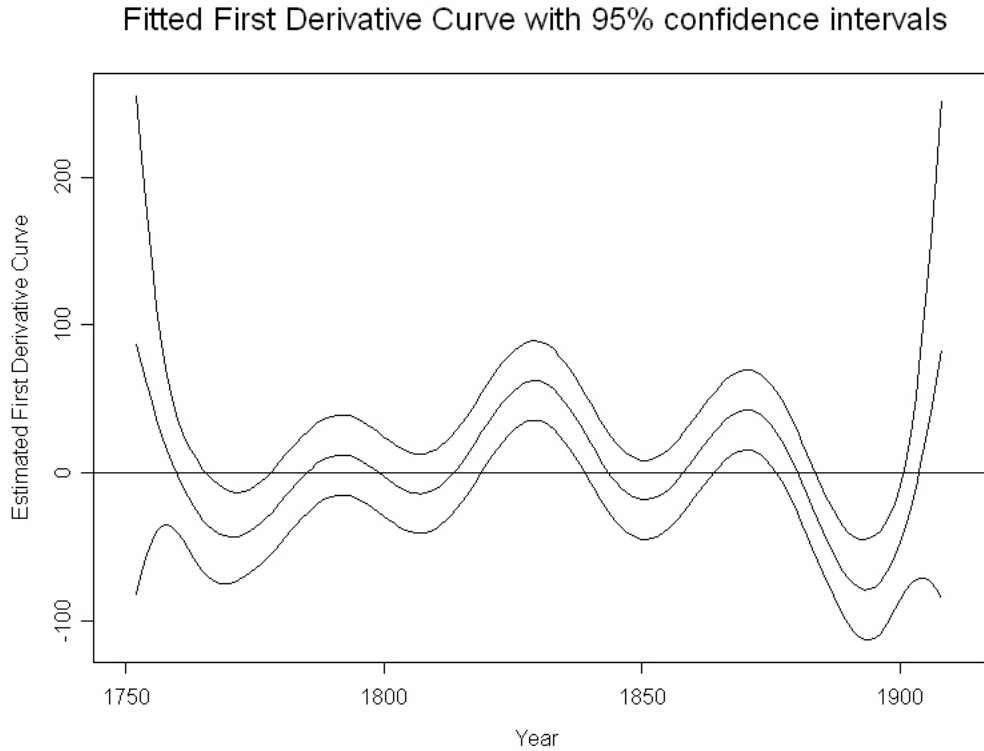


Figure 3.3: Fitted first derivative curve for wolverine data with 95% simultaneous confidence intervals. Regions for which the variability bands are positive correspond to time intervals in which the curve is significantly increasing while regions which the variability bands are negative correspond to time intervals in which the curve is significantly decreasing.

method that calculates the sum of the first derivatives provided similar results for trend detection. The insignificant results reflect that the starting and ending counts are similar for the wolverine data. The likelihood ratio test provided strong evidence for an association between t and y . The significant results for the likelihood ratio test reflect that the wolverine counts are not constant with respect to time but do not necessarily imply an upward or downward trend. Figure 3.3 shows the estimated first derivative curve with simultaneous 95% confidence intervals. Examining the first derivative curve with 95% confidence intervals allows a researcher to explore whether count, y , is constant over time, t . Time intervals for which the confidence intervals do not include zero represent periods that the growth rate is different than zero. This example illustrates the need to not only look at trend throughout the entire

test	1888-1908	1868-1908	1838-1908
Linear Regression	.0009	<.0005	.0155
Kendall's τ	.0019	<.0005	.0116
Sum of first derivatives	.0002	.0222	.0055
LOESS	<.0005	<.0005	<.0005
LRT	<.0001	<.0001	<.0001

Table 3.2: P-values for trend tests over different time periods.

time period but also interval specific trend. Figure 3.4 shows the wolverine data over specific time intervals. Table 3.2 gives the p-value of the test for trend over specific time periods. The time periods reflect the most recent counts. The p-values for the different tests of trend are all no larger than .0222. Excluding the likelihood ratio test for trend which does not indicate direction, all tests gave evidence of a negative trend over the time periods considered. The wolverine data show periods of positive trend, negative trend, and no trend. The first derivative curve is helpful to identify these periods.

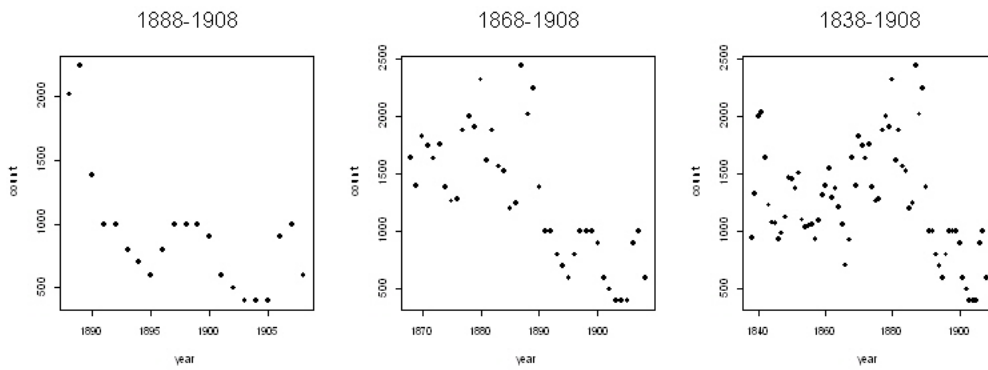


Figure 3.4: Wolverine data for certain time periods.

Chapter 4

Simulation Results

This chapter compares the trend detection methods introduced in Chapter 3. Data were generated with known trend to compare the Type I error rate and power among the different methods. A trend detection method should adhere to the designated type I error rate while having high power when trend is either linear or nonlinear. Ideally, the smoother should have power close to the power of methods based on least squares regression when the trend is linear. When the trend is nonlinear, smoothing methods should have larger power than methods based on linear regression.

Data were simulated for sample sizes ranging from 25 to 200. Data were simulated according to models with iid normally distributed observations, non-normal errors, and serially correlated errors.

A total of seven different trend detection methods will be compared. For detailed descriptions of the trend detection methods please refer to Chapter 3. Least squares regression and Kendall's tau coefficient are the trend detection methods that do not involve smoothing. These methods are used as a standard against which to judge other methods. A short description of the different smoothing methods for trend detection is now given.

Likelihood ratio test for trend: The null and alternative hypothesis for the likelihood ratio test are:

$$H_0 : \beta_1 = \sigma_u^2 = 0$$

$$H_1 : \beta_1 \neq 0 \text{ or } \sigma_u^2 > 0.$$

The linear mixed effects model for the above test is

$$f(t) = \beta_0 + \beta_1 t + \sum_{k=1}^K u_k (t - \kappa_k)_+ + \epsilon_t \quad (4.1)$$

where u_k are iid $N(0, \sigma_u^2)$ and κ_k , $k = 1, \dots, K$, are knot locations. Monte Carlo simulations are used to test the significance of the likelihood ratio statistic.

LOESS: A moving least squares regression (LOESS) is used to test the hypotheses

$$H_0 : \Delta = 0$$

$$H_1 : \Delta \neq 0$$

where Δ is equal to the difference of the mean of the first three fitted values and the mean of the last three fitted values. A bootstrap procedure is used to calculate the standard error of $\hat{\Delta}$.

P-splines: A B-spline matrix with a difference penalty is used to fit a smooth curve to the simulated data. The null and alternative hypothesis for the test using the P-spline fit are

$$H_0 : S(t) = 0$$

$$H_1 : S(t) \neq 0$$

where $S(t) = \sum_{i=1}^n f'(t_i)$ and $f'(t_i)$ is the first derivative of the fitted curve at t_i , $i = 1, \dots, n$.

In Chapter 3 it was argued that for small sample sizes, the test statistic for testing the above

hypotheses can be approximated by a t -distribution.

Penalized spline regression: The null and alternative hypotheses are:

$$H_0 : S(t) = 0$$

$$H_1 : S(t) \neq 0$$

where $S(t) = \sum_{i=1}^n f'(t_i)$ and $f'(t_i)$ is the first derivative of the fitted curve at t_i , $i = 1, \dots, n$.

The fitted model for testing the above hypotheses is

$$\hat{f}(t) = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\beta}_3 t^3 + \sum_{k=1}^K \hat{u}_k (t - \kappa_k)_+^3. \quad (4.2)$$

The fitted first derivative curve of (4.2) is calculated as:

$$\hat{f}'(t) = \hat{\beta}_1 + 2\hat{\beta}_2 t + 3\hat{\beta}_3 t^2 + 3 \sum_{k=1}^K \hat{u}_k (t - \kappa_k)_+^2. \quad (4.3)$$

In Chapter 3 it was argued that for small sample sizes, the test statistic for testing the above hypotheses can be approximated by a t -distribution because the smoother matrix is a linear function of y . However, because the smoothing parameter is estimated, the smoother matrix depends on y through a smoothing parameter (Ruppert et al., [29]). For this reason, a randomization test will also be used to judge the significance of the test statistic for this test.

4.1 Simulation Procedure

Data were generated according to a specified model and the performance of each method using these data was compared by estimating Type I error rates and power. A variety of sample sizes and trend forms were examined. The Type I error rate is defined as the probability that a particular test statistic is contained in the rejection region when, in fact, the null hypothesis

is true. Power is defined as the probability that a particular test statistic is contained in the rejection region when the null hypothesis of no trend is false. All the tests used an α -level of .05 as the boundary for the rejection region.

4.1.1 Generated data

Data were generated using sample sizes of 25, 50, and 200. For each scenario, 1500 data sets were generated and the proportion of times the null hypothesis was rejected at $\alpha = .05$ was observed.

For scenarios that included serially correlated errors, errors were generated that followed an autoregressive process of order 1, (AR(1)). Specifically, errors, ϵ_i for each time, t_i for $i = 1, \dots, n$, were generated such that

$$\epsilon_i = \phi\epsilon_{i-1} + \xi_i$$

where ϕ is the true serial correlation of the errors, and $\xi_i \sim N(0, 1)$. Serially correlated errors with ϕ equal to .3 and .5 were generated to examine the effects of serial correlation on Type I error rates. Table 4.1 summarizes the different scenarios in which simulations were performed and Figure 4.1 shows some examples of generated data for a n of 25. The true trend is shown on each graph.

4.1.2 Comparison of Type I Error Rates

Table 4.2 gives the results of the simulations when the null hypothesis of no trend is true. The proportion of instances the null hypothesis was rejected at $\alpha=.05$ is tabulated in Table 4.2. Standard errors for the estimated Type I error rates ranged from .0055 to .0118. Figure 4.2 gives standard error calculations for a sample size of 1500 for estimated Type I error rates

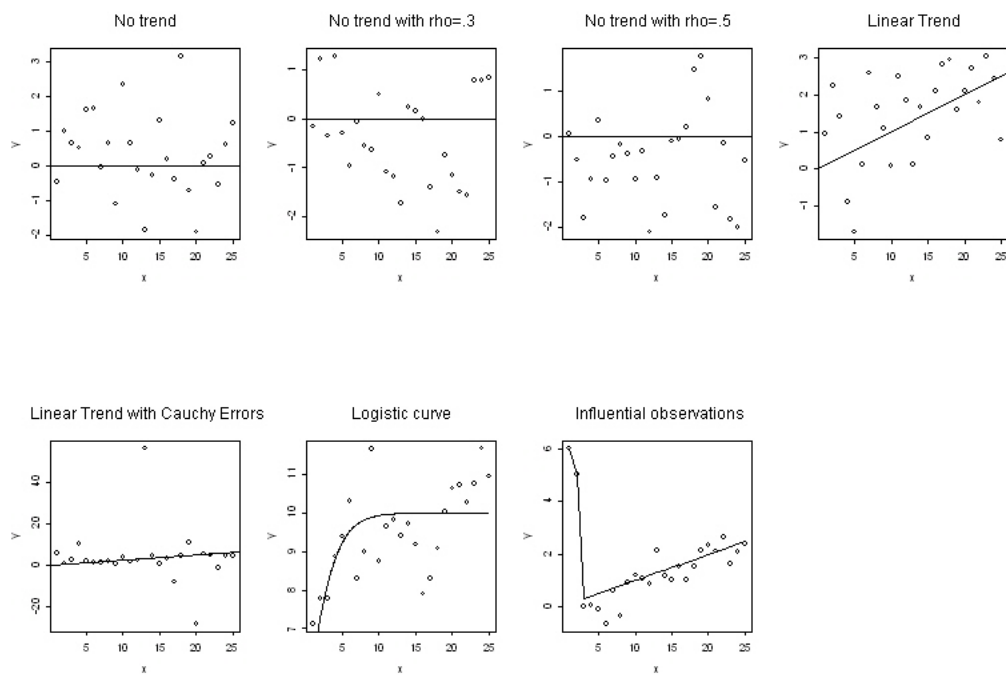


Figure 4.1: Examples of simulated data with line indicating the true function.

Sample size	Description	Equation	Generated errors
25,50,200	No trend	$y = \epsilon$	$\epsilon \sim N(0, 1)$
25,50,200	No trend	$y = \epsilon$ where $\epsilon_i = \phi\epsilon_{i-1} + \xi_i$	$\xi \sim N(0, 1)$
25	Linear	$y = .1t + \epsilon$	$\epsilon \sim N(0, 1)$
50	Linear	$y = .03t + \epsilon$	$\epsilon \sim N(0, 1)$
200	Linear	$y = .004t + \epsilon$	$\epsilon \sim N(0, 1)$
25	Linear	$y = .1t + \epsilon$	$\epsilon \sim CAU(0, 1)$
50	Linear	$y = .03t + \epsilon$	$\epsilon \sim CAU(0, 1)$
25	Linear	$y = .1t + \epsilon$	$\epsilon \sim POI(t)$
50	Linear	$y = .03t + \epsilon$	$\epsilon \sim POI(t)$
25	Logistic	$y = \frac{10}{1+e^{-.5t}} + \epsilon$	$\epsilon \sim N(0, 1)$
50	Logistic	$y = \frac{25}{1+e^{-.1t}} + \epsilon$	$\epsilon \sim N(0, 5)$
200	Logistic	$y = \frac{100}{1+e^{-.075t}} + \epsilon$	$\epsilon \sim N(0, 20)$
25	Exponential	$y = 2e^{.0275t} + \epsilon$	$\epsilon \sim N(0, 1)$
50	Exponential	$y = 10e^{.005t} + \epsilon$	$\epsilon \sim N(0, 1)$
25	Cyclic	$y = .1x + 10 \sin\left(5\pi(.02t)^{1.3}\right) + \epsilon$	$\epsilon \sim N(0, 2)$
50	Cyclic	$y = .2x + 10 \sin\left(5\pi(.02t)^{1.3}\right) + \epsilon$	$\epsilon \sim N(0, 5)$
25	Influential	$y_1 = 6, y_2 = 5,$ $y_i = .1t + \epsilon_i$ for $i = 3, \dots, 25$	$\epsilon \sim N(0, .5)$
50	Influential	$y_1 = 8, y_2 = 7, y_3 = 6, y_4 = 5$ $y_i = .1t + \epsilon_i$ for $i = 5, \dots, 50$	$\epsilon \sim N(0, 1)$

Table 4.1: Description of generated data. Linear trend with the addition of influential observations were examined to explore the effect of influential observations on power.

Sample size	ϕ	LIN	Ken	LRT	SUM.R	LOESS	P-spline	SUM
25	0	.049	.048	.050	.048	.069	.068	.066
25	.3	.155	.145	.178	.165	.160	.163	.166
25	.5	.251	.230	.270	.277	.279	.232	.228
50	0	.051	.052	.048	.049	.065	.062	.060
50	.3	.132	.138	.123	.164	.165	.161	.170
50	.5	.256	.245	.240	.235	.246	.248	.296
200	0	.053	.049	.051	.057	.059	.068	.067
200	.3	.118	.126	.172	.128	.159	.159	.128
200	.5	.265	.250	.270	.194	.268	.245	.194

Table 4.2: Type I error rates for 1500 simulations. Standard errors for 1500 simulations are given in Figure 4.2. Standard errors ranged from .0055 to .0118. The methods compared were linear regression (LIN), Kendall's τ (Ken), the likelihood ratio test for trend (LRT), a randomization test for trend using the sum of the first derivatives (SUM.R), a test for trend using a locally weighted regression (LOESS), the trend detection test using the sum of the first derivatives using P-splines, and a test using the sum of the first derivatives using a linear mixed effects model and distributional assumptions about the test statistic (SUM).

and power.

Linear regression, Kendall's τ , the likelihood ratio test and the randomization test for the sum of the first derivatives all achieved the stated Type I error rate (see Figure 4.3 - Figure 4.5). The LOESS, P-spline and first derivative sum with linear mixed model test failed to meet the Type I error rate of .05. A 95% confidence interval for the true Type I error rate for the LOESS smoothing method was (.0563,.0817). The P-spline smoothing technique and the technique using the sum of the first derivatives had 95% confidence intervals of (.055, .081) and (.053,.079), respectively.

All of the methods were sensitive to the effects of serially correlated errors. For instance, the type I error rate for the likelihood ratio test increased from a .050 (SE=.0056) to .178

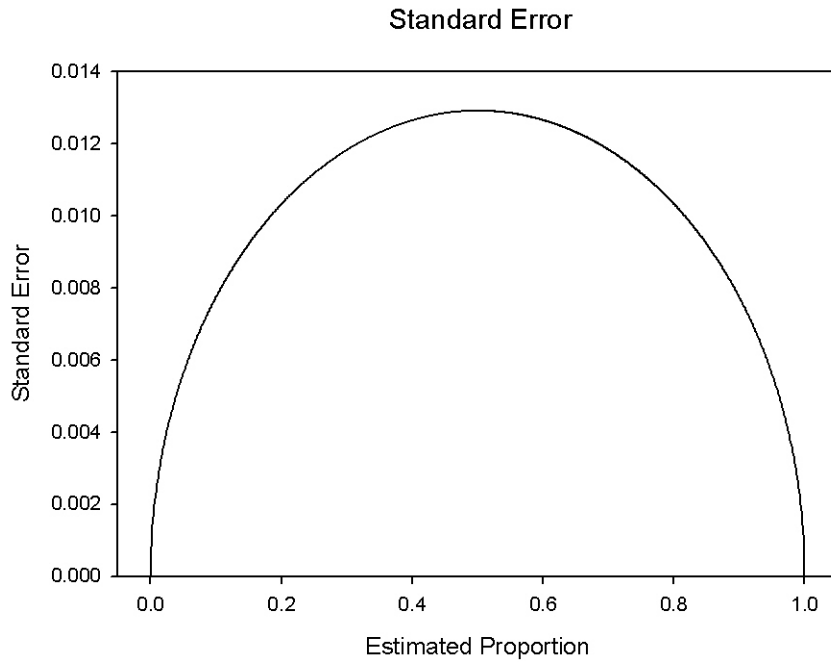


Figure 4.2: Standard error of power and Type I error rates estimates for a sample size of 1500.

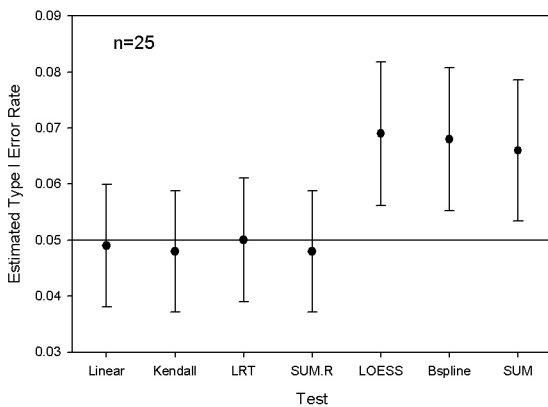


Figure 4.3: Ninety-five percent confidence intervals for Type I error rates for n of 25. SUM is the method that used the sum of the first derivatives and SUM.R is the trend detection method when a randomization procedure was used to obtain the observed significance. The tests were all performed at an α level of .05.

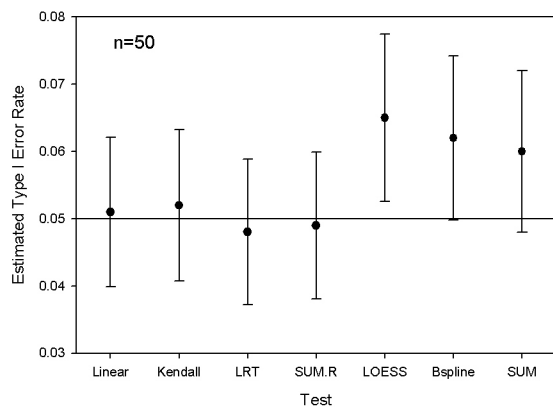


Figure 4.4: Ninety-five percent confidence intervals for Type I error rates for n of 50.

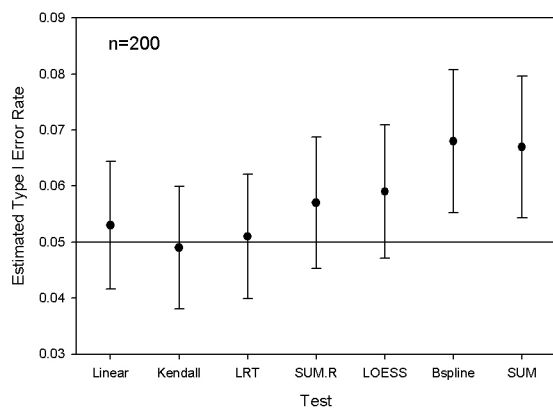


Figure 4.5: Ninety-five percent confidence intervals for Type I error rates for n of 200.

(SE=.0099) with a ϕ of .3. When serial correlation was present in the error structure, none of the methods maintain the Type I error rate of .05.

In summary, it was discovered that the smoothing methods which rely on distributional assumptions about the test statistic all failed to meet the designated Type I error rate for a sample size of 25. For n of 200, the P-spline test for trend and the penalized spline regression test for trend still failed to achieve the Type I error rate of .05. The violation of Type I error rate is most likely due to variability in estimation of the smoothing parameter. Ruppert et al. [29] state that it is common practice to ignore this source of variability. The consequence of doing so is obtaining slightly inflated Type I error rates. Both the likelihood ratio test and the randomization test for trend achieved the designated Type I error rate for all sample sizes when no serial correlation was present in the errors.

4.1.3 Power results

Table 4.3 compares the estimated power of each method when data were simulated with a specified form. Standard error estimates for a given estimated power can be extracted from Figure 4.2. The likelihood ratio test had the greatest power for detecting linear trend compared to the other smoothing methods. The estimated power of the likelihood ratio test was approximately equal to the power of Kendall's τ for the sample sizes considered. This method was superior for detecting linear trend compared to the other smoothing methods (see Figure 4.6- Figure 4.8). Ninety-five percent confidence intervals for the difference between the power of linear trend and the likelihood ratio test method for a n of 25, 50 and 200 are: (.013,.053), (.016,.068) and (.008,.056). Ninety-five percent confidence intervals for the difference between the estimated power of linear trend and the estimated power of the randomization test for n of 25, 50, and 200 are: (.099, .145), (.109,.165), and (.057, .109).

The randomization test using the sum of the first derivatives had the lowest estimated power

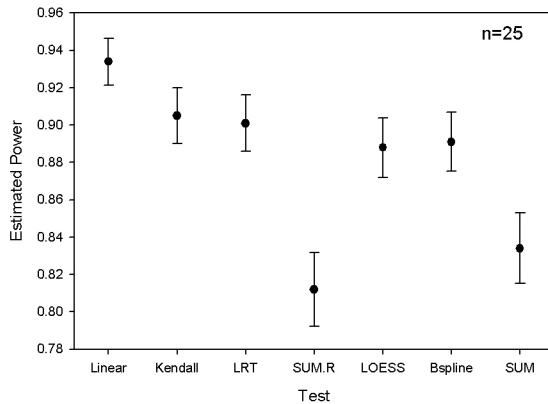


Figure 4.6: Ninety-five percent confidence intervals for power to detect linear trend for n of 25. SUM is the method that used the sum of the first derivatives and SUM.R is the trend detection method when a randomization procedure was used to obtain the observed significance. The tests were all performed at an α level of .05.

for detecting linear trend for n of 25 and 50. Ninety-five percent confidence intervals for the difference between the estimated power of linear trend and the estimated power of the randomization test for n of 25, 50, and 200 are: (.099, .145), (.109,.165), and (.057, .109). Ninety-five percent confidence intervals for the difference between the estimated power of the likelihood ratio test and the randomization test for the sum of the derivatives for n of 25, 50, and 200 are: (.064,.114), (.066,.124), and (.024, .078).

The LOESS method had an estimated power similar to the likelihood ratio test for a sample size of 25 for detecting linear trend, however, had significantly lower estimated power than the likelihood ratio test for a sample size of 50. When the sample size was 200, the LOESS method had the lowest estimated power for detecting linear trend of all the methods.

The smoothing methods were superior to linear regression and Kendall's τ test for detecting logistic trend for n of 25. The tests that involve the sum of the first derivative had higher estimated power than either the likelihood ratio test or the LOESS test for trend for this logistic case. The P-spline method had the highest estimated power of .917 (SE=.0071) for detecting logistic trend followed by the randomization test with a power of .906 (SE=.0075).

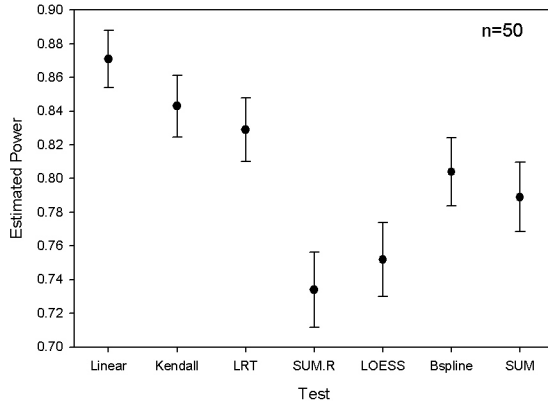


Figure 4.7: Ninety-five percent confidence intervals for power to detect linear trend for n of 50.

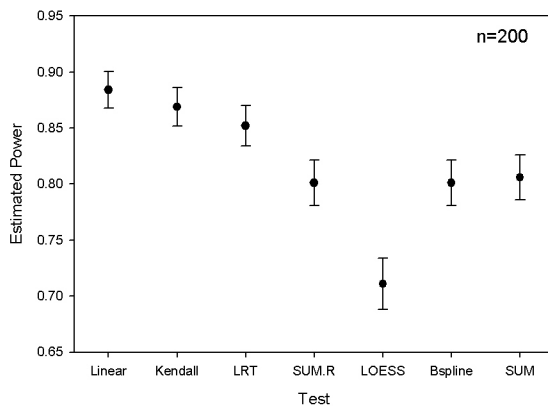


Figure 4.8: Ninety-five percent confidence intervals for power to detect linear trend for n of 200.

Trend	n	LIN	Ken	LRT	SUM.R	LOESS	P-spline	SUM
Linear	25	.934	.905	.901	.812	.888	.891	.834
Linear	50	.871	.843	.829	.734	.752	.804	.789
Linear	200	.884	.869	.852	.801	.711	.801	.806
Linear (Cauchy)	25	.428	.909	.415	.330	.246	.435	.300
Linear (Cauchy)	50	.527	1.0	.505	.449	.678	.522	.503
Linear(Poisson)	25	.819	.858	.800	.720	.724	.756	.736
Logistic	25	.761	.614	.831	.906	.713	.917	.900
Logistic	50	.989	.95	.969	.979	.961	.91	.946
Logistic	200	.992	.977	.997	.999	.997	.98	.999
Influential	25	0	.937	.974	.999	0	1.0	.999
Influential	50	.844	1.0	1.0	1.0	.05	.931	1.0
Exponential	25	.802	.774	.764	.700	.766	.744	.705
Exponential	50	.758	.722	.677	.678	.721	.677	.685
Cyclic	25	.0127	.0687	.0987	.097	.028	.937	.099
Cyclic	50	.416	.352	.998	.531	.608	.796	.550
Catastrophic	25	.007	.236	.994	.603	.435	.952	.450
Catastrophic	50	.204	.596	.996	.84	.996	.856	.852

Table 4.3: Estimated power for 1500 simulations. Figure 4.2 gives the standard error for an estimated power. The methods compared were linear regression (LIN), Kendall's τ (Ken), the likelihood ratio test for trend (LRT), a randomization test for trend using the sum of the first derivatives (SUM.R), a test for trend using a locally weighted regression (LOESS), the trend detection test using the sum of the first derivatives using P-splines, and a test using the sum of the first derivatives using a linear mixed effects model (SUM).

The LOESS method of trend detection had the lowest estimated power of .713 (SE=.0117) of the smoothing methods for this sample size. For larger sample sizes of 50 and 200 there were no significant differences between the power of the trend detection methods for logistic trend.

Least squares regression is sensitive to the effects of influential observations. Influential observations were added to simulated data with linear trend (see Table 4.1). As expected, least squares regression performed poorly in detecting trend. The LOESS test for trend also had low power for detecting trend in the presence of influential observations. All other methods had a large estimated power in the presence of influential observations.

For the simulation that resembled a catastrophic event, the likelihood ratio test and the test using P-splines had significantly higher power than the other methods for a sample size of 25. The estimated power of the likelihood ratio test was .994 (SE=.002) compared to the estimated power for linear regression of .007 (SE=.002) or the estimated power of Kendall's τ of .236 (SE=.011).

In summary, the likelihood ratio test performed the best of the smoothing methods for detecting linear trend for all sample sizes considered. For nonlinear trend such as logistic or exponential, the randomization test using the sum of the first derivative had a larger estimated power than the likelihood ratio test in some cases. The LOESS method of trend detection had significantly less power for detecting logistic trend for n equal to 25. In addition, for large n , the LOESS method had significantly less power for detecting linear trend.

4.2 Autocorrelation methods

Independence of errors is an important assumption for all of the methods that have been considered. It is often the case when analyzing time series data that errors close in time tend to be more similar than errors further away in time. The type I error rates when the errors

were serially correlated were well above the designated α -level of .05 for all of the methods.

In Chapter 3, three methods were introduced that incorporate serial correlation into the error structure. Generalized linear regression was developed by Reinsel et. al. [27]. This method assumes the error structure is modeled by an autoregressive process of order 1. Generalized linear regression is then used to estimate the regression coefficients.

An advantage of using the linear mixed effects model for smoothing is the ease with which serial correlation can be incorporated into the error structure. The linear mixed effects model with correlated errors is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} \quad (4.4)$$

where

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \mathbf{I} & 0 \\ 0 & \mathbf{R} \end{bmatrix} \right)$$

where

$$\mathbf{R} = \frac{\sigma_\epsilon^2}{1 - \phi^2} \begin{bmatrix} 1 & \phi & \dots & \phi^{n-2} & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \dots & \phi & 1 \end{bmatrix}. \quad (4.5)$$

Both the likelihood ratio test and the test involving the sum of the first derivatives can be modified to account for serial correlation in the error structure (see Chapter 3 for details). For the likelihood ratio test, Monte Carlo simulation procedures are used to calculate the observed significance of the test statistic. For the sum of the first derivatives test, a randomization test is used to calculate the observed significance of the test statistic.

Trend	Sample size	ϕ	slope	LIN.AR	SUM.AR	LRT.AR
No trend	25	0	0	.053	.049	.042
No trend	50	0	0	.052	.049	.048
No trend	200	0	0	.052	.062	.051
No trend	25	.3	0	.072	.061	.051
No trend	50	.3	0	.072	.051	.049
No trend	200	.3	0	.053	.047	.052
No trend	25	.5	0	.09	.07	.056
No trend	50	.5	0	.074	.059	.056
No trend	200	.5	0	.048	.040	.044

Table 4.4: Estimated Type I error rates for 1500 simulations for data simulated with no trend and linear trend. Data were simulated with ϕ equal to 0,.3 and .5. The methods that are compared are a generalized least squares method (LIN.AR), a randomization test to test the significance of the sum of the first derivatives (SUM.AR), and a likelihood ratio test (LRT.AR). All of these methods account for serial correlation.

4.2.1 Simulation results for autocorrelation methods

Table 4.4 shows the results of the test methods that incorporate serial correlation into the error structure. Errors, ϵ_i for each time, t_i for $i = 1, \dots, n$, were generated such that

$$\epsilon_i = \phi\epsilon_{i-1} + \xi_i$$

where ϕ is the true serial correlation of the errors and $\xi_i \sim N(0, 1)$.

The generalized linear regression method with autocorrelated errors failed to meet the designated Type I error rate for small sample sizes. A 95% confidence interval for the true Type I error rate of this method with n equal to 25 is (.058, .085) when ϕ is .3 and (.076,.104) when ϕ is .5. For an n of 50, the Type I error rates again failed to achieve the specified value of .05. A 95% confidence interval for the true Type I error rate when n is 50 is (.058, .085)

Trend	Sample size	ϕ	slope	LIN.AR	SUM.AR	LRT.AR
Linear	25	0	.1	.909	.723	.810
Linear	50	0	.03	.851	.697	.822
Linear	200	0	.004	.886	.779	.805
Linear	25	.3	.1	.68	.466	.524
Linear	50	.3	.03	.573	.400	.496
Linear	200	.3	.004	.645	.506	.644
Linear	25	.5	.1	.486	.308	.364
Linear	200	.5	.004	.384	.224	.340

Table 4.5: Estimated power for 1500 simulations for data simulated with linear trend. Data were simulated with ϕ equal to 0,.3 and .5. The methods that are compared are a generalized least squares method (LIN.AR), a randomization test to test the significance of the sum of the first derivatives (SUM.AR), and a likelihood ratio test (LRT.AR). All of these methods account for serial correlation.

when ϕ is .3. For n of 200, the estimated Type I error rate when ϕ is equal to .3 is .053 (SE=.006). The generalized linear regression method had the greatest estimated power for detecting linear trend. For an n of 200, this method is comparable to the likelihood ratio test for autocorrelated errors. A 95% confidence interval for the difference in power between the two methods is (-.033,.035) for n equal to 200.

The likelihood ratio test with autocorrelated errors achieved the stated α level of .05 for all sample sizes considered. All 95% confidence intervals for the true Type I error rate capture .05 for this method.

Table 4.5 reports the estimated power for detecting linear trend for the autocorrelation methods. The estimated power for detecting linear trend is higher than for the method that uses the sum of the first derivatives. With 95% confidence, we can say that the likelihood ratio test has between .022 and .094 higher power than the method using the sum of the first derivatives when n is 50 and ϕ is .3. For small sample sizes, the power for detecting linear trend was

significantly reduced compared to the methods without serial correlation in the error structure (see Table 4.3). This is expected due to dependence of errors. The results given in Table 4.2 indicated that estimated Type I error rates are inflated due to the presence of autocorrelation, therefore, methods that account for serial correlation in the response variable should be used when evidence exists of serial correlation. The method with the largest estimated power for the autocorrelation methods was the generalized least squares regression method. The estimated power for detecting linear trend when n is 50 and ϕ is .3 was .573 (SE=.013). The estimated power for the likelihood ratio test when n is 50 and ϕ is .3 was .496 (SE=.013).

4.3 Conclusions

In this chapter we examined which trend detection methods held their type I error rates and also had high power. A trend detection method should perform well in a variety of situations. For instance, the association between time, t and count, y , may be linear or nonlinear. Therefore, a trend detection method that performs well when the association is either linear or nonlinear is desired. Other issues which may occur would be instances where a population has suffered a drastic decline due to some catastrophic event.

It was discovered that the smoothing methods did not meet the designated type I error rate (see Figure 4.6). Although not serious violations of the designated type I error rate of .05, methods that achieve the designated type I error rate would be preferred. The randomization test using the sum of the first derivatives achieved the desired type I error rate; however, the power of this method was lower than for all other smoothing methods for detecting linear trend (see Figure 4.6). The likelihood ratio test not only achieved the designated type I error rate but also had high power for detecting linear trend. Additionally, the likelihood ratio test had high power for detecting trend in the presence of influential observations and when a catastrophic event was simulated (see Table 4.3).

Serial correlation in the error structure had the effect of inflating the type I error rates for all of the methods. An advantage of the linear mixed effects model is the ease with which serial correlation can be incorporated into the error structure. The methods that were developed to accommodate serial correlation had lower type I error rates in the presence of serial correlation, however, also lower power for detecting linear trend. The presence of serial correlation should be examined, and if evidence of such correlation exists, one should use one of the methods developed that incorporates the serial correlation into the model. For the autocorrelation methods, the likelihood ratio test appears to be superior. For small sample sizes, the likelihood ratio test achieved the designated Type I error rate. None of the estimated Type I error rates for the likelihood ratio test were significantly different than .05 for all sample sizes examined. In addition to this, the likelihood ratio test had significantly larger power than the test for the sum of the first derivatives.

In conclusion, this chapter focused on comparing type I error rates and the power of detecting trend. It was discovered that the distributional assumptions for the smoothing methods may not be appropriate based on the fact that type I error rates are too large for these smoothing methods. The smoothing methods that adhered to the type I error rates were the randomization test for the sum of the first derivatives and the likelihood ratio test. Of these two methods, the likelihood ratio test had higher power for detecting linear trend than the randomization test using the sum of the first derivatives. Both the randomization test for the sum of the first derivatives and the likelihood ratio test were easily extendable to accommodate serial correlation, however, the likelihood ratio test was superior over the randomization test in terms of both power and achieving Type I error rates.

Chapter 5

Case Study

This chapter concludes the research with some case studies and discusses some extensions to the trend detection methods that were introduced in previous chapters. The likelihood ratio test was concluded to be a powerful method of trend detection for both linear and nonlinear associations. Recall the model used for this test

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}.$$

The likelihood ratio test statistic when no evidence of serial correlation exists is

$$-2 \log LR(\mathbf{y}) = -2 \left\{ l\left(\hat{\beta}_{0,0}, 0, 0, \hat{\sigma}_{\epsilon,0}^2; \mathbf{y}\right) - l\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2, \hat{\sigma}_{\epsilon}^2; \mathbf{y}\right) \right\} \quad (5.1)$$

where $(\hat{\beta}_{0,0}, \hat{\sigma}_{\epsilon,0}^2)$ minimizes $-2l(\hat{\beta}_0, 0, 0, \hat{\sigma}_{\epsilon,0}^2; \mathbf{y})$ and $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2, \hat{\sigma}_{\epsilon}^2)$ minimizes $-2l(\hat{\beta}_{0,1}, \hat{\beta}_1, \hat{\sigma}_u^2, \hat{\sigma}_{\epsilon}^2; \mathbf{y})$.

The null and alternative hypotheses for trend are

$$H_0 : \beta_1 = \sigma_u^2 = 0$$

$$H_1 : \beta_1 \neq 0 \text{ or } \sigma_u^2 > 0.$$

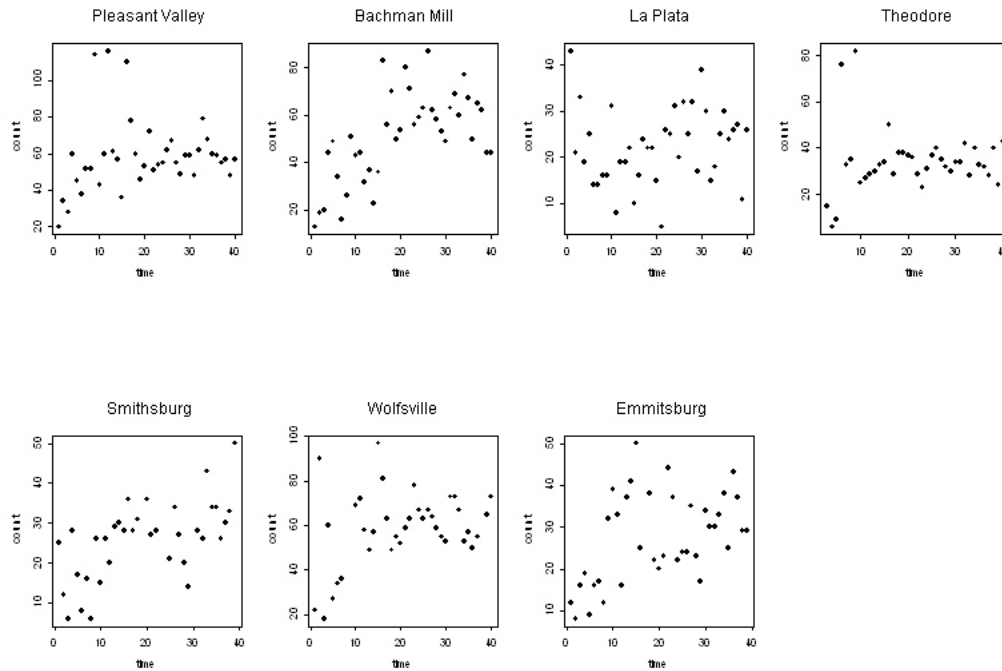


Figure 5.1: Mourning dove counts at seven different BBS routes in Maryland from 1966 to 2005.

Monte Carlo simulations are used to test the significance of the observed test statistic.

5.1 Data

Data for the BBS survey can be obtained at the website, www.pwrc.usgs.gov/bbs. For the case study presented in this chapter, Mourning dove (*Zenaida Macroura*) data are presented. Data were obtained for seven different routes in Maryland from the years 1966 to 2005. Figure 5.1 shows scatterplots at each route. There appears to be a positive association between time, t , and count y , however, the form appears nonlinear. With the exceptions of some unusual values, all seven routes appear to have approximately the same overall tendency of positive trend for the first ten years of the survey and no trend for the remainder of the years.

5.2 Random site effect

The focus of my research has been to find an optimal method of trend detection for observations at a single site. If a survey covers a random sample of sites, then site should be considered a random effect. The BBS, as well as other surveys, covers a large number of sites (routes) across a region and interest may lie in drawing conclusions about trend at multiple sites. The routes for the Mourning dove data can be treated as a random effect for the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & t_{11} \\ \vdots & \vdots \\ 1 & t_{1n_1} \\ \vdots & \vdots \\ 1 & t_{m1} \\ \vdots & \vdots \\ 1 & t_{mn_m} \end{bmatrix},$$

$$\mathbf{Z} = \begin{bmatrix} 1 & \dots & 0 & (t_{11} - \kappa_1)_+ & \dots & (t_{11} - \kappa_K)_+ \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \dots & 0 & (t_{1n_1} - \kappa_1)_+ & \dots & (t_{1n_1} - \kappa_K)_+ \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & (t_{m1} - \kappa_1)_+ & \dots & (t_{11} - \kappa_K)_+ \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & (t_{mn_m} - \kappa_1)_+ & \dots & (t_{1m_m} - \kappa_K)_+ \end{bmatrix}$$

and

$$\mathbf{u} = \begin{bmatrix} U_1 \\ \vdots \\ U_m \\ u_1 \\ \vdots \\ u_K \end{bmatrix}$$

where

$$\text{cov} = \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \sigma_U^2 I_m & 0 & 0 \\ 0 & \sigma_u^2 I_K & 0 \\ 0 & 0 & \sigma_\epsilon^2 I_n \end{bmatrix}.$$

Hence, U_1, \dots, U_m are random intercepts for each route, and u_1, \dots, u_K controls the amount of smoothing of the curve. Random intercepts for the fitted curve appear reasonable for the Mourning Dove data. The log-likelihood is

$$l(\beta_0, \beta_1, \sigma_\epsilon^2, \sigma_u^2, \sigma_U^2; \mathbf{y}) = \left(n \log(2\pi) + \log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right)$$

where $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \sigma^2 \mathbf{I}$ and

$$\mathbf{G} = \begin{bmatrix} \sigma_U^2 I & 0 \\ 0 & \sigma_u^2 I \end{bmatrix}.$$

The parameters $\boldsymbol{\beta}$, σ_U^2 , σ_ϵ^2 and σ_u^2 are estimated by the maximum likelihood methods. There is no closed-form solution to compute the maximum likelihood estimates so estimation is accomplished through an algorithm such as the Newton-Raphson algorithm[25]. The test for trend is

$$H_0 : \beta_1 = \sigma_u^2 = 0$$

$$H_1 : \beta_1 \neq 0 \text{ or } \sigma_u^2 > 0.$$

The likelihood ratio test statistic for the above hypotheses is

$$-2\log LR(\mathbf{y}) = -2 \left\{ l\left(\hat{\beta}_{0,0}, 0, \hat{\sigma}_{U,0}^2, 0, \hat{\sigma}_{\epsilon,0}^2; \mathbf{y}\right) - l\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_U^2, \hat{\sigma}_u^2, \hat{\sigma}_{\epsilon}^2; \mathbf{y}\right) \right\} \quad (5.2)$$

where $(\hat{\beta}_{0,0}, \hat{\sigma}_{U,0}^2, \hat{\sigma}_{\epsilon,0}^2)$ minimizes $-2l(\hat{\beta}_{0,0}, 0, 0, \hat{\sigma}_{U,0}^2, \hat{\sigma}_{\epsilon,0}^2; \mathbf{y})$ and $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_U^2, \hat{\sigma}_u^2, \hat{\sigma}_{\epsilon}^2)$ minimizes $-2l(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_U^2, \hat{\sigma}_u^2, \hat{\sigma}_{\epsilon}^2; \mathbf{y})$.

Monte Carlo procedures can be employed to test the significance of formula 5.2. The Monte Carlo procedure follows the steps:

1. Calculate the likelihood ratio statistic for the observed data.
2. Data is then simulated by fixing the parameters equal to their specified values under the null hypothesis.
3. The likelihood ratio statistic is calculated for step 2.
4. Repeat steps 2 and 3 numerous times and construct an empirical sampling distribution for the test statistic under the null hypothesis.
5. The observed value of the test statistic is significant at the α -level of .05 if it falls in the upper 95 percentile of the distribution constructed in step 4.

5.3 Results

Figure 5.2 shows the Mourning Dove data with the main effects curve shown on the graph. There appears to be a positive association overall, however, a tendency for the counts to stay fairly constant after the 18th year. Table 5.1 shows the estimated coefficients and variance components. The observed likelihood ratio test statistic was 30.62. The p-value for this based on 10,000 simulations was less than .0001 giving us strong evidence that there is an association between t and y .

Parameter	Estimate	standard error
β_0	21.89	6.807
β_1	1.636	.573
σ_ϵ	13.44	1.19
σ_U	15.77	9.80
σ_u	.544	.664

Table 5.1: Results of linear mixed model fit for mourning dove data. The estimated parameter values and standard errors are given.

5.4 Most recent trend

The likelihood ratio test is a powerful method for detecting both linear and nonlinear associations. However, of interest to researchers is not only detecting trend, but also quantifying trend. It is common to estimate trend for a smoothed curve by taking the difference between the first few values and the last few fitted values. Although we can estimate trend in this manner for any smooth curve, if the rate of change of the population of interest has not remained constant then this estimate may not be useful as a measurement of trend. The model for the likelihood ratio test provides a piecewise linear fit between knot locations. Therefore, I propose to estimate the most recent trend as the slope of the linear spline between the last knot, κ_K and the last observation, t_n . If we redefine t to be time since the most recent observation, then this slope can be estimated by the parameter, $-\hat{\beta}_1$. For the mourning dove data, the estimate for $-\beta_1$ is -.154 with a standard error of .577. Thus, we can conclude that the most recent trend is not significantly different than zero. For this example, knots were placed approximately every 4 observations. For the purpose of getting more precise estimates of the most recent trend, we can remove the first knot. If we delete a knot, then the estimate of the most recent trend (β_1) is -.111 with a standard error of .399.

5.5 Conclusion

This chapter concludes the research with a case study. The two issues that were addressed were extending the linear mixed effects model for smoothing and quantifying the most recent trend. Mourning dove count data collected at seven routes in Maryland were analyzed. Although the likelihood ratio test for trend gave significant results, the observed trend over the last 8 years was not significantly different than zero.

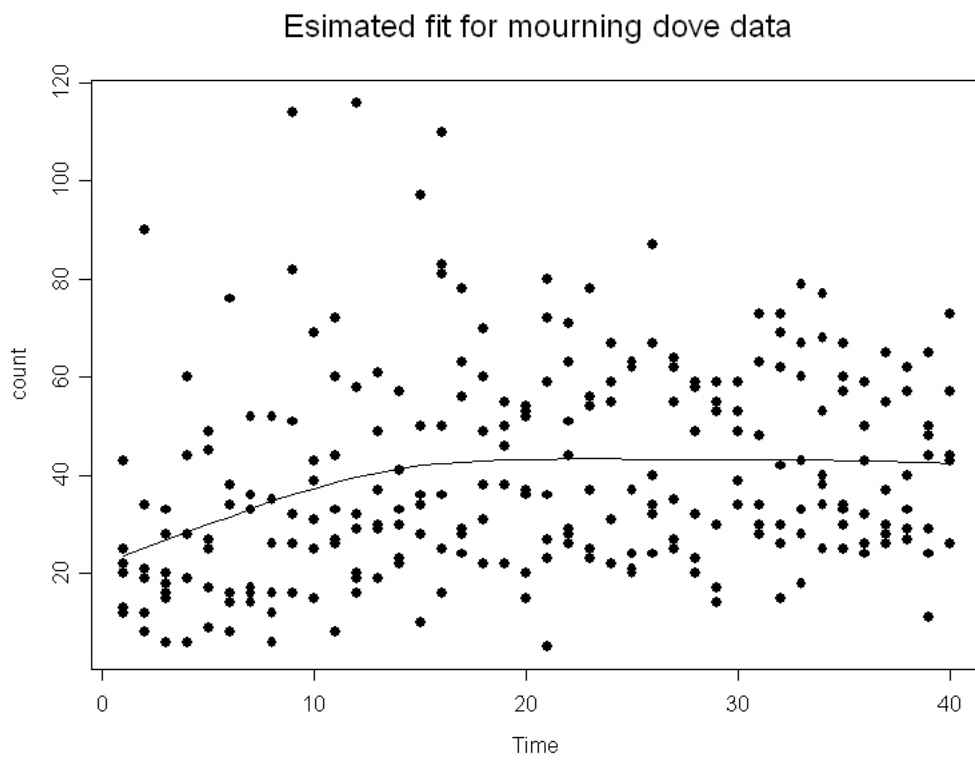


Figure 5.2: Main effects curve for the mourning dove analysis.

Chapter 6

Conclusions and Future Research

The main results of this research will be summarized in this chapter and potential future research will be suggested.

Of the tests for trend that were compared, the tests that held the stated type I error rate were the randomization test involving the sum of the first derivatives and the likelihood ratio test. The tests for trend that relied on asymptotic distribution assumptions for the distribution of the test statistic all failed to meet the stated Type I error rate of .05. This is due to the fact that the distributional assumptions were based on a fixed smoothing parameter when in fact the smoothing parameter is estimated from the data. The likelihood ratio test for trend had the largest estimated power of the smoothing methods for the detection of linear trend. This test also performed well for nonlinear trend and had large estimated power when influential observations were added into simulations for linear trend. In addition to this, the model was easily extendable to account for autocorrelation in the errors. Even for small sample sizes, the likelihood ratio test achieved the stated Type I error rate when errors followed an AR(1) process. The disadvantage of this test is that a significant result does not indicate whether the population is increasing or decreasing only that population is not constant across

time. Usually researchers are interested in quantifying trend and this test does not lend itself to quantifying trend. However, the difference between the first and last value (or any time interval) can be used to estimate trend if trend is quantified as the population change over a time interval.

Because using smoothing techniques as a method for trend detection is still relatively new, there have been few methods that use the first derivatives to quantify and test for trend (see Fewster et al. [11]). The first derivative curve is useful for displaying interesting features of a curve. For instance, a researcher may be interested in addressing the following questions:

1. During what time interval is trend increasing, decreasing or staying constant?
2. When does growth rate start to plateau?
3. Is the most recent trend positive or negative?

The first derivative curve with confidence intervals can address these questions. The first derivative curve can also be used to quantify trend. My research examined using the sum of the first derivatives as a means of quantifying trend; however, other methods such as the median first derivative or a weighted average of first derivatives giving the most recent first derivatives the most weight could also be used to quantify trend. The advantage of examining the derivatives as a means of trend detection is that this method is useful in not only quantifying trend but also addressing questions about the population trajectory across time. In addition, serial correlation can be accommodated in the error structure using the mixed effects model. The disadvantage of this method is that the power for detecting linear trend for small sample sizes is low compared to other methods and confidence intervals for first derivatives at the end and beginning of time periods tend to be very wide.

Examining derivatives has a bright future with regard to describing, quantifying, and testing for trend. The tools introduced in this research are widely applicable to all types of monitoring

data and will be even more useful for annual surveys such as the BBS when sample sizes become larger. Because a mixed model is used to model the smooth curve, extensions such as autocorrelation are easy to implement using standard software such as S-PLUS.

The penalized spline regression methods assume normally distributed errors. Often for monitoring data, including the BBS, the population index is a count. Poisson regression is often used to model count data. If a reasonable assumption is that the errors follow a Poisson distribution, then the linear mixed effects model can be extended to accommodate Poisson errors via a generalized linear mixed effects model (see Ruppert et al. [29]). A potential area of research could be to examine the properties of a test for trend that assumes Poisson distributed errors and employs a mixed effects model to model the fitted curve.

Bibliography

- [1] A. Anganuzzi, *A comparison of tests for detecting trends in abundance indices of dolphins*, Bird Populations **3** (1997), 48–66.
- [2] C. de Boor, *A practical guide to splines*, revised, Springer, New York, 2001.
- [3] E. Chi and G. Reinsel, *Models for longitudinal data with random effects and ar(1) errors*, Journal of the American Statistical Association, 452.
- [4] W. Cleveland, *Robust locally weighted regression and smoothing scatterplots*, Journal of the American Statistical Association **74** (1979), 829–836.
- [5] B. Coull, J. Schwartz, and M. Wand, *Respiratory health and air pollution: additive mixed model analyses*, Biostatistics **2** (2001), 337–349.
- [6] C. Dagum and E. Dagum, *Encyclopedia of statistical sciences*, 1st ed., John Wiley and Sons, New York, New York, 2001.
- [7] E. Demidenko, *Mixed models: Theory and applications*, 1st ed., Wiley, New York, New York, 2004.
- [8] P. Eilers and B. Marx, *Flexible smoothing with b-splines and penalties*, Statistical science **11** (1996), 89–101.
- [9] ———, *Splines, knots, and penalties*, 2004. Unpublished paper.
- [10] S. Esterby, *Trend analysis methods for environmental data*, Environmetrics **4** (1993), 459–481.
- [11] R. Fewster, S. Buckland, G. Siriwardena, S. Bailie, and J. Wilson, *Analysis of population trends for farmland birds using generalized additive models*, Ecology **81** (2000), 1970–1984.
- [12] F. Goldsmith, *Monitoring for conservation and ecology*, 1st ed., Chapman and Hall.
- [13] P. Green and B. Silverman, *Nonparametric regression and generalized linear models: a roughness penalty approach*, 1st ed., Chapman and Hall, USA, 1994.
- [14] F. Harrell, *Regression modeling strategies*, 1st ed., Springer-Verlag New York, Inc.
- [15] T. Hastie and R. Tibshirani, *Generalized additive models*, 1st ed., Chapman and Hall, USA, 1990.

- [16] A. Hoerl and R. Kennard, *Ridge regression: Applications to nonorthogonal problems*, *Technometrics* **12** (1970), 69–82.
- [17] F. James, C. McCulloch, and D. Wiedenfeld, *New approaches to the analysis of population trends in land birds*, *Ecology* **77** (1996), 13–27.
- [18] P. Lancaster and K. Salkauskas, *Curve and surface fitting an introduction*, 1st ed., Academic Press LTD, 1986.
- [19] W. Link and J. Sauer, *Estimating population change from count data: Application to the north american breeding bird survey*, *Ecological Applications* **8** (1998), 258–268.
- [20] W. Link and J. Sauer, *New approaches to the analysis of population trends in land birds: Comment*, *Ecology* **78** (1997), 2632–2634.
- [21] B. Manly, *Randomization, bootstrap and monte carlo methods in biology*, 2nd ed., Chapman and Hall, England, 2005.
- [22] D. Marquardt and R. Snee, *Ridge regression in practice*, *The American Statistician* **1** (1975), 3–20.
- [23] R. Myers and J. Milton, *A first course in the theory of linear statistical models*, 2nd ed., The McGraw-Hill Companies, Inc., New York, New York, 1997.
- [24] B. Peterjohn and W. Link J. Sauer, *The 1994 and 1995 summary of the north american breeding bird survey*, *Bird Populations* **3** (1997), 48–66.
- [25] J. Pinheiro and D. Bates, *Mixed-effects models in s and s-plus*, 1st ed., Springer, New York, New York, 2000.
- [26] Imperial College NERC Centre for Population Biology, *The global population dynamics database*, 1999.
- [27] G. Reinsel, *Statistical analysis of stratospheric ozone data for trend detection*, *Environmetrics* **81** (1981), 215–235.
- [28] G. Robinson, *That BLUP is a good thing: The estimation of random effects*, *Statistical Science* **6** (1991), 15–51.
- [29] D. Ruppert, M. Wand, and R. Carroll, *Semiparametric regression*, 1st ed., Cambridge University Press.
- [30] D. Ruppert, *Selecting the number of knots for penalized splines*, *Journal of Computational and Graphical Statistics* **11** (2002), 735–757.
- [31] T. Ryan, *Modern regression methods*, 1st ed., John Wiley and Sons, Inc., New York, New York, 1997.
- [32] J. Sauer, J. Fallon, and R. Johnson, *Use of north american breeding bird survey data to estimate population change for bird conservation regions*, *Journal of Wildlife Management* **67** (2003), 372–389.
- [33] J. Sauer, J. Hines, G. Gough, I. Thomas, and B. Peterjohn, *The north american breeding bird survey results and analysis*, 96th ed.

- [34] R. Sokal and J. Rohlf, *Biometry*, 3rd ed., W. H. Freeman and Company, New York, New York, 1995.
- [35] D. Suits, A. Mason, and L. Chan, *Spline functions fitted by standard regression methods*, *The Review of Economics and Statistics* **5** (1977), 132–139.
- [36] L. Thomas and K. Martin, *The importance of analysis method for breeding bird survey population trend estimates*, *Conservation Biology* **10** (1996), 479–490.
- [37] L. Thomas, *Monitoring long-term population change: Why are there so many analysis methods?*, *Ecology* **77** (1996), 49–58.
- [38] M. Wand, *A comparison of regression spline smoothing procedures*, *Computational Statistics* **15** (2000), 443–462.
- [39] ———, *Smoothing and mixed models*, *Computational Statistics* **18** (2003), 223–249.
- [40] E. Wegman and I. Wright, *Splines in statistics*, *Journal of the American Statistical Association* **78** (1983), 351–363.