icro,
.
on on
years
eds to
ng the
/ 2000
garding
to the
2/00

MONASH UNIVERSITY
THESIS ACCEPTED IN SATISFACTION OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
ON.....19 December 2001.....
.....[redacted]
for Sec. Research Graduate School Committee

A Smoothing Spline Approach to Nonlinear Inference for Time Series

A thesis submitted for the degree of

Doctor of Philosophy

by

Ivet Pitrun

Msc (Bus/Eng) Czechoslovakia

Department of Econometrics and Business Statistics

Monash University

Australia

October 2001.

Contents

Summary

There are many situations where the relationship of a response variable as a function of a quantitative variable (e.g., time) is of importance. The presence of other variables is also common. This thesis considers statistical models in which the additional variables are modeled by means of a parametric linear relationship and the time variable is modeled by an unknown non-parametric function that can be estimated by cubic smoothing splines.

Cubic smoothing splines have attracted much attention as an alternative to parametric regression. Perhaps their popularity is due to the fact that the criterion giving rise to cubic splines can be seen as a combination of a least squares goodness of fit criterion and a term that penalizes the degree of curvature of the regression function.

This thesis outlines a general regression methodology, which provides data-driven solutions to the following problems:

- the variance-covariance structure of the resultant non-parametric regression,

- diagnostic checking for departures from linearity,
- bandwidth selection,
- prediction.

The approach is data analytic in which the possible relationship is determined by data, instead of being limited to a certain functional form as in parametric analysis. Using non-parametric models in place of parametric ones introduces flexibility into the smoothing process. Our approach uses cubic smoothing splines in their state-space representation.

In this thesis we use the semiparametric model in which the time component is estimated by cubic smoothing splines. We express the semiparametric model in the state-space form and derive the variance-covariance matrix of the error disturbance vector by applying Yule-Walker equations.

A desirable capability in many economic and business situations is to be able to make reasonably accurate predictions of future values of the outcome variable. This thesis improves the potential use of the cubic smoothing splines in prediction by deriving the best unbiased predictor and the variance-covariance matrix of its prediction error. This allows us to construct prediction intervals and to test whether a recently realised observation on the dependent variable could have been generated from the same model that generated the past observations.

The derivation of the variance-covariance matrix of the disturbance vector of the general regression model leads naturally to the extension of the family of the likelihood ratio based tests. Two new tests for detecting nonlinearity are proposed in this thesis. One problem considered is that of testing for the inclusion of a possibly nonlinear component in the model and the other is testing for linearity of a possibly nonlinear component. Where alternative comparable test procedures are available, extensive simulations are used to benchmark the performance of the methodology and demonstrate that both new tests are competitive. A comparison study between mean and variance based tests is a part of this thesis.

The smoothing parameter, which controls the trade off between the smoothness of the solution and the fidelity to the data, is very important in non-parametric smoothing techniques. The marginal likelihood function is constructed for the new tests for nonlinearity. The smoothing parameter is a by-product of the maximization of the marginal likelihood. Simulations are used to compare the estimated value of the smoothing parameter along with the accuracy of the estimated nonlinear function for the time component in the model.

A wide variety of non-linear data sets that are difficult to model and estimate are used to illustrate the potential of the approach in applied data analysis.

Acknowledgments

I would like to thank Professor M.L. King and Associate Professor R.J. Hyndman of the Department of Econometrics and Business Statistics, Monash University, Australia, for their joint supervision of this thesis and their editorial help. They have done an excellent job and I am very grateful.

I am obliged to Professor A. Linczenyi of the Department of Management and Quality, Technical University, Slovak Republic, who showed me the way how to fall in love with standard deviation. He has introduced me to such a wonderful science, as statistics no doubt is.

Finally, I would like to thank my husband, Mike, for his support and understanding, and my two children, Oscar and Nicole for being so patient. I hope I have not missed out on too much of your growing up.

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgment is made in the text.

Ivet Pitrun,

Melbourne,

October 2001.

CHAPTER 1

Introduction

1.1 Background

Every day, economists make personal and professional decisions that are based upon predictions of future events. To be able to do this, they rely upon the relationship between what is already known and what is to be estimated. If decision makers can determine how the known is related to future outcomes, they can aid the decision-making process considerably. To be able to do this, they need to formulate and test a model based on what has been observed and utilize it for prediction.

Econometrics as a science involves estimation and testing of economic models. Maddala (1992) formulated three main aims of econometrics:

- formulation of economic models in empirically testable form,
- estimation and testing of these models with observed data,
- using these models for prediction and policy purposes.

Beginning with the initial set of assumptions, econometric theories provide the guidance to formulate an economic model, which is used to describe the phenomenon under study. A commonly used model in econometrics, which this thesis is mainly concerned with, is the regression model. Regression and correlation analyses are well known and widely used techniques to determine both the nature and the strength of a relationship between the variables. Francis Galton first used the term regression as a statistical concept in 1877. He designated the word regression as the name of the general process of predicting one variable from another. Later, statisticians coined the term multiple regression to describe the process by which several variables are used to predict another. This allows the use of more available information to predict the unknown values of the dependent variable.

Given a variable we want to explain and a group of potential explanatory variables, there may be several different regression equations we can look at, depending on which explanatory variables we include and how we include them. Each such regression equation is called a model. Modeling techniques are the various ways in which we can include the explanatory variables and check the appropriateness of the resultant models.

There are many modeling techniques. This thesis is concerned with one of the non-parametric modeling techniques called spline smoothing.

In the case of univariate modeling, a regression curve describes a general relationship between an explanatory variable and a response variable. The aim of regression analysis is to produce a reasonable approximation to the unknown response function. A look at a scatter plot of variables does not always suffice in order to establish an interpretable regression relationship. By describing the observational errors a regression model allows one to concentrate on important details of the mean dependence of the variables. The task of approximating the mean function can be done in two ways. The familiar parametric approach is to assume that the mean curve has some prespecified functional form. An example of a parametric model is a polynomial regression equation where the parameters are the coefficients of the independent variables. An important assumption of the parametric approach is that the approximation bias of the best parametric fit is a negligible quantity. An alternative approach is to estimate the mean curve non-parametrically without reference to a specific form. This overcomes the main disadvantage of the parametric approach which is that a preselected parametric model might be too restrictive or too rigid to fit unexpected features. The non-parametric approach offers a flexible tool in analyzing unknown regression relationships. The term "non-parametric" refers to the flexible functional form of the regression curve, therefore neither the functional form of the mean function nor the error distribution is prespecified.

Hardle (1990) listed the main advantages of the non-parametric approach to estimating a regression curve as follows:

- it provides a versatile method of exploring a general relationship between the variables,
- it gives predictions of the observations yet to be made without reference to a fixed parametric model,
- it provides a tool for finding spurious observations by studying the influence of isolated points,
- it constitutes a flexible method of substituting for missing values or interpolating between adjacent values.

The prediction of new observations is of particular interest in time series analysis. It has been found that in certain applications classical parametric models are too restrictive to give a reasonable explanation of observed phenomena. Robinson (1983) has investigated the non-parametric prediction of time series. In the nineties, non-parametric smoothing methods became popular techniques in applied research in many fields of economics such as electricity demand modeling, housing prices modeling, term structure of interest rate modeling, investment behavior modeling etc. The recent book by Schimek (2000) is a unique collection of a variety of points of view regarding non-parametric regression, smoothing and statistical modeling. This book provides an excellent reference for researchers in various fields wishing to obtain an overview of assorted techniques and practical issues related to implementing non-parametric techniques.

The question of which approach should be taken in data analysis is a compromise between the possibility of gross misspecification resulting from too high a model bias in the case of a pure parametric model and more variable estimates in the case of non-parametric models. The ideal approach seems to be a combination of the advantages of both methods in a semiparametric mixture.

1.2 Direction and motivation

The aim of this thesis is to extend a general approach to non-parametric smoothing by modeling the unknown, possibly nonlinear function of the time component by cubic smoothing splines in a semiparametric model. The optimal value of the smoothing parameter is determined by optimizing the marginal likelihood function. To be able to derive the marginal likelihood function, the variance-covariance matrix of the disturbance vector of the general regression model will be derived. Further applications of cubic smoothing splines in nonlinear inference for time series will be based on the identified error term structure.

This thesis is largely motivated by work of Shively, Kohn and Ansley (1994), who proposed a new exact test for nonlinearity in regression models with one possibly nonlinear component. Their test is based on the stochastic interpretation of the smoothing splines as given in Wahba (1978) and is a point optimal invariant test, as defined in King

(1987). Wecker and Ansley (1983) derived an exact maximum likelihood estimator for the parameters of Wahba's model. Shively et al. (1994) combined these findings to derive their exact test, which is computationally intensive and requires a numerical derivation of the critical values. Their approach is not easy to understand and loses some of its appeal because of the complicated computation required. We extend the Shively et al. (1994) work by deriving the exact covariance structure of the error term for cubic smoothing splines and equally spaced data. We apply our finding to prediction by deriving the best linear unbiased predictor and the variance-covariance matrix of its prediction error. This allows formulation of prediction intervals for future values.

We propose two new variance based tests for detecting by nonlinearity applying a likelihood ratio hypothesis testing approach. Our approach is computationally very friendly and easy to implement using conventional statistical software. We are interested in two testing problems. The first is to test for the inclusion of a possibly nonlinear component and the second one is to test for linearity against the alternative of nonlinear component. Thus in the first test we would like to know, whether an additional component, which we assume is nonlinear, should be included in a regression model. In the second test we are testing whether the additional component which should be included in the model, based on the first test result, is or is not linear indeed.

We also investigate the amount of smoothness of the estimated regression function, which is controlled by a smoothing parameter. This is commonly known as a bandwidth

selection in non-parametric regression techniques. We compare the existing methods of smoothing parameter selection with our approach.

1.3 Outline of the research

Chapter 2 provides an introduction to the theory of the univariate polynomial splines, and is largely based on the work of Silverman (1985), Mathews (1992), and Green and Silverman (1994). It briefly reviews the rapidly developing field of non-parametric smoothing with focus on spline smoothing techniques. We introduce the general spline classification. Attention is focused on spline functions in relation to estimating the unknown function as a data-driven solution. We briefly discuss the objective of regression splines and related problems of choosing the optimum number of knots and their optimum location. We describe the concept of smoothing splines, because smoothing spline applications are the focal point of this thesis.

Chapter 3 is motivated by the work of Wahba (1978), Wecker and Ansley (1983), and Judge et al. (1988). We introduce univariate cubic smoothing splines in a state-space form and derive the variance-covariance matrix Σ of its disturbance vector by Yule-Walker equations. This derivation leads to the application of Σ in prediction. We derive the best linear unbiased predictor and the variance-covariance matrix of its prediction error. Based on this derivation, we construct prediction intervals.

of the methodology presented here. We evaluate the accuracy of the estimation performance by the square root of the integrated squared error.

Chapter 7 summarizes the original contribution of this thesis. It also outlines associated problems with this research that require additional consideration.

CHAPTER 2

Spline functions

2.1 Introduction

Spline functions were introduced to the mathematical literature just over 70 years ago by Whittaker (1923) and extended later by Shoenberg (1964). The basic idea is to use simple functions, usually cubic polynomials, to construct a suitably smooth piecewise function. Since their introduction, splines have proved to be very popular in interpolation, smoothing and approximation and in computational mathematics generally. As well as their theoretical interest, splines are used as computational tools in science, engineering, computer-aided design, business, and the social sciences.

In this chapter we introduce the concept of spline functions. We will focus on spline functions in relation to estimating an unknown function based on a set of data.

In Chapters 4 and 5, we propose two variance based tests for detecting nonlinearity based on marginal likelihood functions. Both tests are likelihood ratio type tests. In the first test, which we denote MLR test, we formulate null hypotheses as that there is no need to include any additional component in a regression model, against the alternative one, that there is. We assume that this additional component is nonlinear. In the second test, which we denote MLR* test, we are in fact testing whether this additional component is or is not linear. Therefore we formulate the null hypotheses in this test as that the additional component which should be included in a regression model is linear against the alternative, that it is nonlinear. We compare the tests' performance with those of alternative comparable tests by computing their sizes and powers. We investigate their performance for ten nonlinear functions that are difficult to model by traditional parametric approaches. We perform a large Monte Carlo simulation subject to variation of the test functions and the variance of the error vector. We also compare the variance based tests with mean based tests for the inclusion and for the linearity of a possibly nonlinear component. These two chapters demonstrate the application of cubic smoothing splines in nonlinear inference, namely hypothesis testing

The work of Kohn, Ansley and Tharm (1991) motivates Chapter 6. This chapter extends the family of bandwidth selection methods of the smoothing parameter λ. A maximum marginal likelihood estimate of the smoothing parameter is a by-product of the likelihood function maximization from Chapters 4 and 5. A comparison between penalized regression and maximum likelihood based methods highlights the promising performance

A very common problem in practice, is to fit a function $s(t)$ of a known form to a set of data values (t_i, Y_i), for $i = 1,2,\ldots,n$, where $E(Y_i|t_i) = s(t_i)$. The most well-known technique for finding such a fit is *least squares curve-fitting*. The underlying model is of the form

$$Y_i = s(t_i) + \varepsilon(t_i). \tag{2.1}$$

The simplest form is the linear regression, which will fit a straight line to any data, regardless of the shape of the data. It is one of the most classical and widely used techniques. In other words, the data are regarded as realizations from the model

$$Y_i = \alpha + \beta t_i + \varepsilon(t_i). \tag{2.2}$$

The error $\varepsilon(t)$ is often assumed to be independent identically distributed noise with mean zero and constant variance σ^2. The main purposes of such regression analyses are to quantify the contribution of the covariate t to response Y per unit of value of t, to summarize the association between the two variables, to predict the mean response for a given value of t, and to extrapolate the results beyond the range of the observed covariate values. The linear regression technique is very useful if the mean response is linear

$$E(Y_i|t_i) \equiv s(t_i) = \alpha + \beta t_i. \tag{2.3}$$

If the scatter plot of the data appears to be non-linear, it is reasonable to use a non-linear relationship and linearise it by a suitable transformation. An important class of such models is specified by means of functions that are intrinsically linear. The non-linear yet intrinsically linear models involve functions of the independent variable that are either strictly increasing or strictly decreasing. In many situations, a scatter plot of data suggests that $s(t)$ has some local minimas and maximas. In this case a fit of a polynomial function, i.e. polynomial regression, is a very popular approach.

While the polynomial regression approach has been widely used, it suffers from a few drawbacks. One is that polynomials often suffer from excessive 'wiggle' as the order of the polynomial increases. The polynomial functions are not very flexible in modelling many problems because polynomial functions have all orders of derivatives everywhere. Another is that individual observations can have a large influence on remote parts of the curve. A third point is that the polynomial degree cannot be controlled continuously. There are several ways to repair the drawbacks of polynomial fitting. One is to allow possible discontinuity of derivative curves. This leads to the *spline approach.* Another possible direction is to expand the regression function into orthogonal series, then choose a few useful subsets of the basis functions, and use them to approximate the regression function. This approach is called the *orthogonal series method.* A third approach is that, instead of increasing the number of parameters, one can apply the linear model locally. This approach is termed the *local (linear) modelling approach.* In this thesis we will focus on the spline approach only.

2.2 Interpolating splines

In the least-squares method, we assume that there are errors and uncertainties in the data, so we do not demand that our fit passes through the data points (t_i, y_i). In the *interpolating-polynomial technique*, our fitted polynomial will pass through all data points (t_i, y_i). There are a number of techniques, such as Newton's, Lagrange interpolating polynomials and many more, coming from numerical mathematics that fit an interpolating polynomial to data points (t_i, y_i).

A more serious problem with polynomial interpolation is that it often results in a large amount of oscillation, especially for large polynomials of order n, therefore the interpolating fit is less accurate. Introducing *piecewise linear* or *piecewise quadratic interpolation* usually improves modelling. The resulting curve is not smooth because the derivative $s'(t)$ is discontinuous (i.e., the graph has "corners"). An alternative approach is still to apply lower-order polynomials but of larger degree than 2 to a subset of data points. Such connecting polynomials are called *spline functions*. The most frequently used spline functions are third-order polynomials known as *cubic splines*. These functions have the additional property that the connections between adjacent cubic equations are visually smooth. The interpolating cubic spline also connects all the points, but because it is limited to third-order changes, the oscillations are kept to a minimum. The spline usually provides a superior approximation to the behaviour of functions that have local, abrupt changes.

It is possible to construct cubic functions $s_k(t)$ on interval $[t_k, t_{k+1}]$, so that the resulting piecewise curve $y = s(t)$ and its first and second derivatives are all continuous on the larger interval $[a,b]$. The graph $Y = s(t)$ will not have sharp corners and the radius of curvature is defined at each point because $s_k'(t)$ and $s_k''(t)$ is continuous. Mathews (1992) presented the following definition of the *cubic spline interpolant*. Let $\{(t_k, t_{k+1})\}_{k=0}^{n}$ be $n+1$ points where $a = t_0 < t_1 < \ldots < t_n = b$. These points t_k are called *knots*. The function $s_k(t)$ is called a *cubic spline* if there exist n cubic polynomials $s_k(t)$ with the properties:

(i) $s(t) = s_k(t) = s_{k,0} + s_{k,1}(t - t_k) + s_{k,2}(t - t_k)^2 + s_{k,3}(t - t_k)^3$

for $t \in [t_k, t_{k+1}]$ for each $k = 0, 1, \ldots, n-1$.

This means that the spline is a cubic polynomial.

(ii) $s(t_k) = s_{k,0} = y_k$ for $k = 0, 1, \ldots, n$.

This indicates that the spline passes through each data point.

(iii) $s_k(t_{k+1}) = s_{k+1}(t_{k+1})$ for $k = 0, 1, \ldots, n-2$.

This condition means that the spline is a continuous function.

(iv) $s_k'(t_{k+1}) = s_{k+1}'(t_{k+1})$ for $k = 0, 1, \ldots, n-2$.

This means that the spline is a smooth function.

(v) $s_k''(t_{k+1}) = s_{k+1}''(t_{k+1})$ for $k = 0,1,\ldots,n-2$.

The second derivative is also continuous, which is related to the curvature of the cubic spline interpolant.

The data points give $(n+1)$ conditions and properties (iii), (iv) and (v) each give $(n-1)$ conditions. Therefore $n+1+3(n-1) = 4n-2$ conditions are specified. This leaves two additional degrees of freedom unaccounted for. These are called *endpoint constraints*; they involve either $s'(t)$ or $s''(t)$ at t_0 and t_n. The typical approach for the endpoint constraints can be summarized as follows:

(i) *Clamped cubic spline* : specify $s'(t_0), s'(t_n)$.

(ii) *Natural cubic spline* : a relaxed curve, $s''(t_0) = s''(t_n) = 0$, so that $s(t)$ is linear on the two extreme intervals.

(iii) *Extrapolated spline*: extrapolate $s''(t)$ to the endpoints.

(iv) *Parabolically terminated spline*: $s''(t)$ is constant near the endpoints.

(v) *Endpoint curvature-adjusted spline*: specify $s''(t)$ at each endpoint.

We refer to Mathews (1992) for more details about various interpolating cubic splines, their computation and their graphical presentation. The work of Schumaker (1981) provides a history of interpolating splines from a numerical analyst's point of view. Prenter (1975) describes their role in the numerical solution of differential equations and DeBoor (1978) is the standard reference on algorithms for generating univariate splines.

2.3 Univariate polynomial regression splines

Splines are of interest to scientists for their favourable properties in smoothing noisy data at least since Shoenberg's (1964) work. The two major types of non-interpolating splines as a function of one variable are regression splines and smoothing splines.

The concept of so-called *B-splines* (*B* stands for "basis") is used in regression splines. *B-splines* are not interpolating splines, because curves used for fitting data do not in general pass through the given data points. A special type of curve in this family are *Bezier curves*, see Geraald et al. (1989). We refer to DeBoor (1978) for more details about univariate regression splines.

In the previous section "knots"= $\{t_i\}$. Now assume that $\{k_1, k_2, \ldots, k_K\}$ are the K knots placed along the domain of the independent variable t, such that $min(t_i) < k_1 < \ldots < k_K < max(t_i)$. An important problem associated with fitting regression splines is the choice of both the number and location of the knots. A large number of knots leads to interpolation of the data. As the number of knots becomes much smaller than the sample size, the resulting regression spline will have an increasingly smooth appearance and the residuals will tend to increase. If the knots are badly located, details of the curve can be missed. The knots are often placed at locations where curvatures have a reasonably large change. The "eyeball" or trial and error method is frequently used, but it can be very cumbersome and time consuming.

We refer to work of Agarwal and Studden (1980) on the issue of determining the optimal number of knots. Friedman and Silverman (1989) proposed a

forward/backward stepwise knot-placement strategy. Their model selection and knot-placement strategy is known as the TURBO smooth model. Silverman (1985) proposed a *minimum span method* for knot placement. Fan et al. (1996) listed a *knot deletion* idea as an automatic procedure for selecting knots. Finding the optimum number of knots and their location in regression splines is similar to the finding optimum value of the smoothing parameter λ in smoothing splines.

For applications of regression splines, see Poirier (1973), Buse and Lim (1977), Niu (1996), Shively and Sager (1997), and Smith (1998a). Regression splines have attracted the attention of many scientists in economics and business, see for example Vasicek et al. (1982), Engle et al. (1986), Munson and Jernigan (1989), Kohsaka (1992), Smith and Huang (1993), Moyeed (1995), Smith (1998b), and Chen et al. (1999). Smith and Kohn (1996, 1997), and Smith et al. (1996, 1998) discuss applied non-parametric regression using regression splines in a Bayesian context.

2.4 Smoothing splines and the penalized least squares problem

Smoothing splines owe their origin to Whittaker (1923) whose work on graduating data stimulated Schoenberg (1964) to derive the smoothing spline estimator. Later, Reinsch (1967) gave an independent derivation for the case of a cubic spline. Smoothing splines were generally regarded as a numerical analysis tool until extensive research, pioneered by Grace Wahba, showed they had useful statistical properties and deserved consideration as a method for performing non-parametric

regression analysis. An introduction to her work, along with references to her more important papers, can be found in the Wahba (1990) monograph. It has since become evident that smoothing splines, and their variants, provide extremely flexible data analysis tools. As a result of this, they have became very popular in data analysis and have found applications in many areas, including economics and business. The Green and Silverman (1994) monograph has become a standard theory reference in this field of research.

Fitting the data well is an obvious aim of curve fitting. Another objective is to fit a reasonably smooth curve to the data. These two objectives are in conflict because a good fit can result in a widely fluctuating or rough curve. The solution to this conflict has been to introduce a roughness penalty term into the goodness-of-fit measure, so there is a reasonable compromise between a good fit to data and smoothness of the fitted curve.

The word spline comes from the name for a flexible piece of wood used to draw smooth curves in the days before computer graphics. If the $s(t)$ is the mathematical function of the curve to be drawn then the leading term in the mathematical expression for the strain energy in this piece of wood is a constant times $\int s''^2$. An obvious measure of goodness-of-fit is the sum of squared errors $\sum\left(y_i - s\left(t_i\right)\right)^2$ and if we take $\int s''^2$ as a reasonable approximation for the roughness of the curve (the less smooth the curve, the greater the strain on the spline) then we get

$$S_\lambda(s) = \sum_{i=1}^{n} \left(y_i - s(t_i)\right)^2 + \lambda \int_a^b \left(s''(x)\right)^2 dx \tag{2.4}$$

as a penalised goodness-of-fit measure for the fitted curve $s(t)$, which is required to be a twice-differentiable function on $[a,b]$. The smoothing parameter $\lambda > 0$ is an important parameter in (2.4). It regulates the 'rate of exchange' between goodness-of-fit and smoothness. Small values give rise to rougher fitted functions and larger values result in smoother curves with larger mean squared errors The penalized least squares estimator $\hat{s}(t)$ is the minimiser of the expression (2.4) from the range of all twice-differentiable functions $s(t)$ and the function $\hat{s}(t)$ is known as a cubic smoothing spline.

Consequently a *general smoothing spline* of order m that estimates $s(t)$ is obtained by minimizing the penalized sum of squares

$$S_\lambda(s) = \sum_{i=1}^{n} \left(y_i - s(t_i)\right)^2 + \lambda \int_a^b \left(s^{(m)}(t)\right)^2 dt \tag{2.5}$$

over all functions $s(t)$, having square integrable mth derivative, where $\int \left(s^{(m)}(t)\right)^2$ represents a roughness measure. The solution to minimizing (2.5) for given λ is a spline function $\hat{s}(t;\lambda)$ of degree $2m-1$, that is a piecewise polynomial of degree $2m-1$ between the design points t_i with $2m-2$ continuous derivatives across the points. *Cubic smoothing splines* are obtained when $m = 2$.

2.4.1 Smoothing parameter selection

The problem of choosing the smoothing parameter λ is very important in curve estimation. An incorrect value of λ tends to overestimate or underestimate the possible relationship. In the case of smoothing we use the terminology "over-smoothing" or "under-smoothing".

As noted above, the value of the smoothing parameter $\lambda(0 < \lambda < \infty)$ regulates the trade-off between goodness-of-fit and smoothness. The choice of λ is therefore vital for the shape of the fitted curve. Increasing λ increases the smoothness of the curve as the $\int \hat{s}''^2$ term becomes more important. This has the effect of forcing $\int \hat{s}''^2$ down. In the limit as λ approaches infinity, $\int \hat{s}''^2$ will be forced to zero and we end up with a straight line fitted by least squares.

Figures 2.1 – 2.5 show cubic smoothing splines with different values of λ applied to the same artificial data set. The artificial set of data was created as follows: we generated a set of 100 independent design points, t_i, uniformly distributed over the interval $[0,1]$, and than generated a response variable $y_i = s(t_i) + \varepsilon_i$ via the function $s(t) = \sin(4t)^2 + 3t^8$ with the errors $\varepsilon_i, i = 1,\ldots,100$, being iid $\mathrm{N}(0, 0.5^2)$.

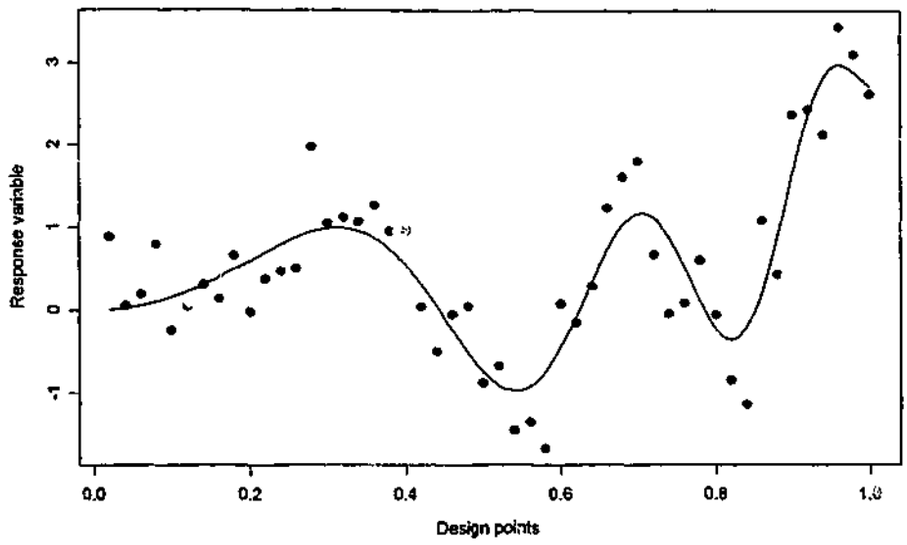

Figure 2.1: *Scatter plot of the artificial data with smooth function* $s(t)$ *superimposed*

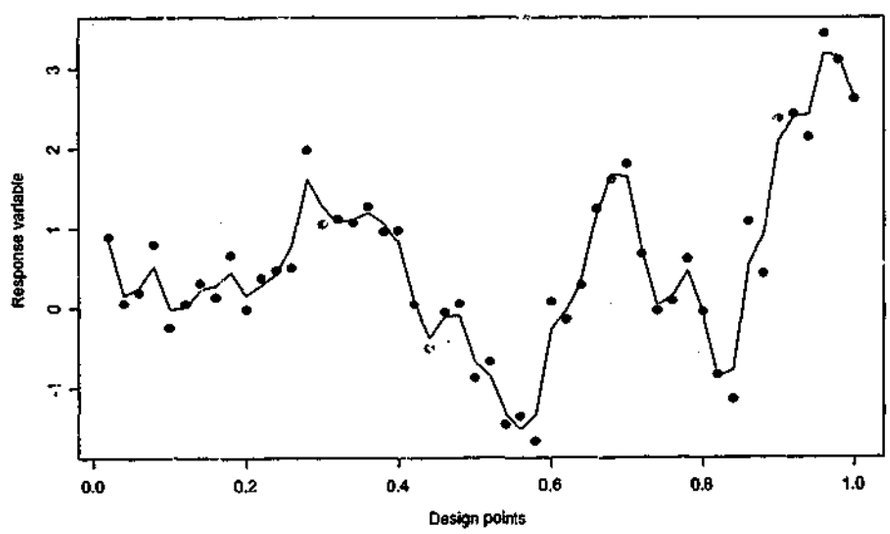

Figure 2.2: *Scatter plot of the artificial data and the estimated cubic spline with* λ *being too small*

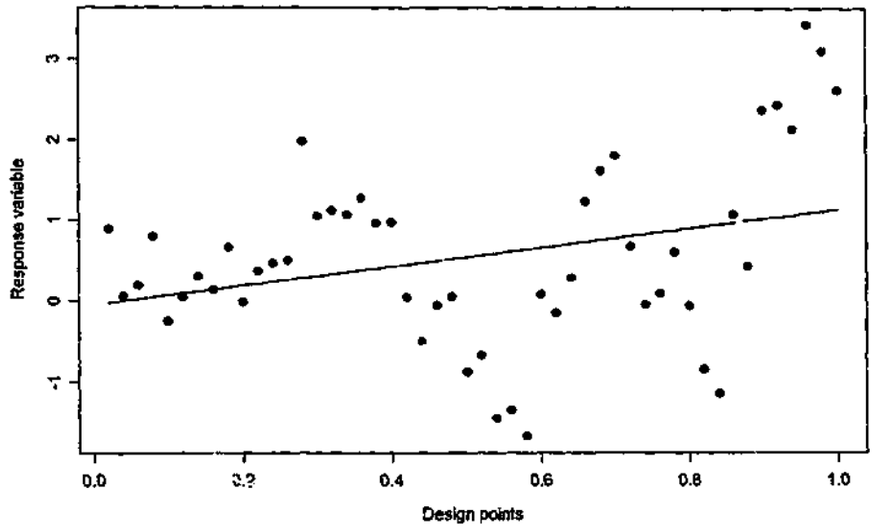

Figure 2.3: *Scatter plot of the artificial data and the estimated cubic spline with* $\lambda = \infty$

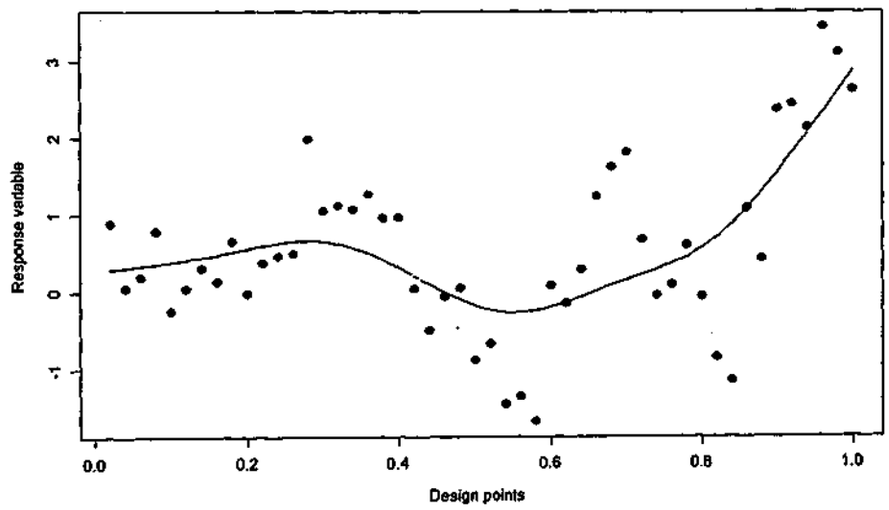

Figure 2.4: *Scatter plot of the artificial data and the estimated cubic spline with* λ *being too large*

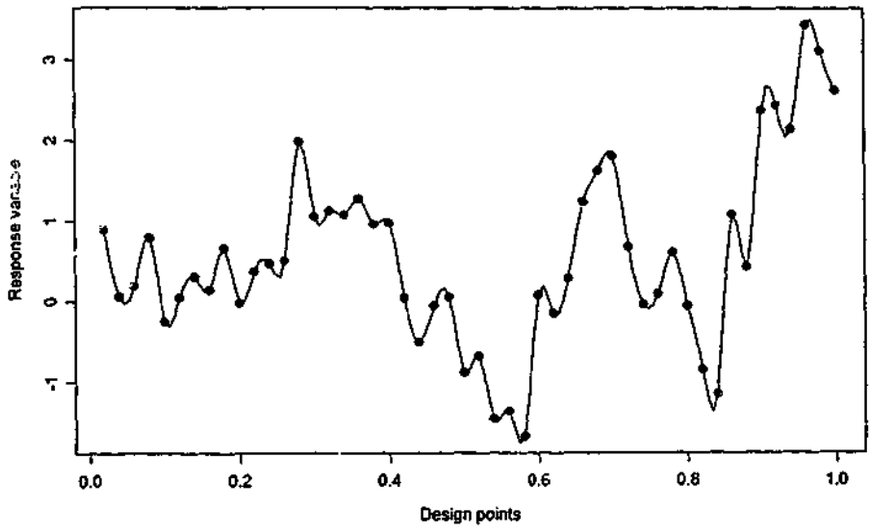

Figure 2.5: *Scatter plot of the artificial data and the estimated cubic spline with* $\lambda = 0$

In the two limiting cases we have interpolation for $\lambda = 0$ in *Figure 2.5,* and the linear fit for $\lambda = \infty$ in *Figure 2.3.* An incorrect value of the smoothing parameter tends to oversmooth *(Figure 2.4)* or undersmooth *(Figure 2.2)* the relationship between variables. If the aim is to smooth the data only, larger values of the smoothing parameter produce smoother curves while smaller values produce rougher curves. The values of smoothing parameter λ used in the above-mentioned exercise were chosen subjectively according to visual inspection of the artificial data set.

Figure 2.6 gives the spline estimate with λ "about right". Generalized cross validation (to be discussed) was used to choose λ.

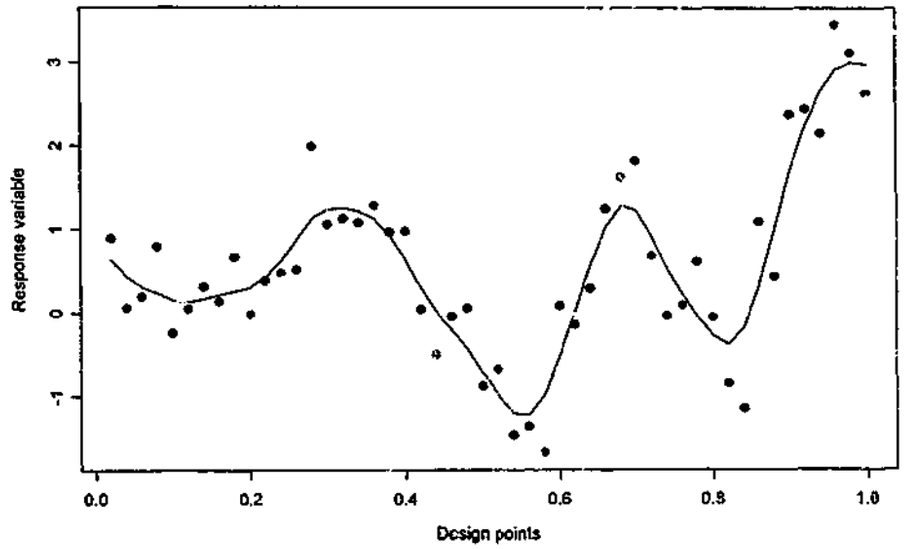

Figure 2.6: *Scatter plot of the artificial data with cubic spline estimate using GCV*

The correct value of the smoothing parameter λ is therefore crucial in fitting and smoothing the data by smoothing spline techniques. Theoretical and practical discussions concerning the selection of the smoothing parameter λ within the penalized regression framework are rich in the literature, see Eubank (1988) for a good review, also see Green and Silverman (1994).

Often the smoothing parameter, λ, is chosen by a *trial and error* approach. An arbitrary value of λ is selected until one is found which represents a visually satisfactory fit to the data. This can be time consuming and has the disadvantage that no optimality properties can be attributed to the value selected in this manner. It is preferable to use some data driven, objective or automatic methods for selecting λ.

The most well known automatic procedures are *cross-validation* (CV) and *generalized cross-validation* (GCV). Allen (1974) suggested CV in the context of regression splines and Wahba and Wold (1975) in the context of smoothing splines. Craven and Wahba (1979) introduced an improved version of CV known as GCV. It has become the standard automatic procedure for λ determination. Note that CV and GCV do not need σ^2 for estimating λ.

Robinson and Moyeed (1989) introduced a *robustification* of generalized cross-validation being motivated by the finding that CV can cause under-smoothing. For an appealing comparison between GCV and robustified GCV, see Neubauer and Schimek (1994).

Other procedures, which require estimating σ^2 for λ selection, are *maximum likelihood* (ML) or *restricted maximum likelihood* (REML). Anderssen and Bloomfield (1974) were the first to suggest the use of a maximum likelihood estimate for λ in a smoothing context. The Bayesian formulation of the smoothing spline estimator given in Wahba (1978) allows smoothing parameter estimation to be performed using the likelihood function. ML estimation of the smoothing parameter λ was incorporated in work by Ansley and Wecker (1981) and by Wecker and Ansley (1983). Their finding was later generalized to *marginal likelihood* (GML) by Wahba (1985), and Ansley and Kohn (1985). According to Speed (1991), REML estimation of λ coincides with the GML procedure discussed by Wahba (1985). Kohn, Ansley and Tharm (1991) used Monte Carlo simulations to compare CV, GCV and GML. They found that GCV and GML outperformed CV for unequally spaced data, while GCV and GML had similar properties for cubic smoothing splines.

In recent work, Lin and Zhang (1999) estimated a non-parametric function in their model by using smoothing splines. They jointly estimated the smoothing parameter λ and variance component σ^2 using a *marginal quasi-likelihood* method. Their approach is based on the generalized additive mixed models setting.

We devote Chapter 6 of this thesis to the problem of estimating the smoothing parameter λ.

2.4.2 Properties

This thesis focuses on natural smoothing splines following the approach of Green and Silverman (1994). They call this representation of smoothing splines the value-second derivative representation. The following discussion draws heavily on Green and Silverman (1994, pp. 11 – 13 and 17 - 20).

Assume s is the natural cubic spline with knots $t_1 < \ldots < t_n$. Let $s_i = s(t_i)$ and $\xi_i = s''(t_i)$ for $i = 1, \ldots, n$, $s = (s_1, \ldots, s_n)'$ and $\boldsymbol{\xi} = (\xi_2, \ldots, \xi_{n-1})'$. Note that $\boldsymbol{\xi}$ is an $(n-2)\times 1$ vector because for a natural cubic spline $s''(t_1) = s''(t_n) = 0$ so that ξ_i for $i = 1$ and n are zero. The curve $s(t)$ can be completely specified by s and $\boldsymbol{\xi}$; in fact it is possible to specify $s(t)$ and its derivatives as explicit functions of s and $\boldsymbol{\xi}$.

We now need to introduce the matrices Q, R and K. Assume that h_i is the difference between two subsequent knots such as $h_i = t_{i+1} - t_i$ for $i = 1, \ldots, n - 1$. Matrix Q is the $n \times (n-2)$ matrix with elements q_{ij}, for $i = 1, \ldots, n$ and $j = 2, \ldots, n-1$ given by $q_{j-1,j} = h_{j-1}^{-1}$, $q_{jj} = -h_{j-1}^{-1}$, and $q_{j+1,j} = h_j^{-1}$ for $j = 2, \ldots, n-1$ and $q_{ij} = 0$ for $|i-j| \geq 2$. Matrix R is of order $(n-2) \times (n-2)$ with entries r_{ij}, for i and j from 2 to $(n-1)$, specified by $r_{ii} = \frac{1}{3}(h_{i-1} + h_i)$ for $i = 2, \ldots, n-1$ and $r_{i,i+1} = r_{i+1,i} = \frac{1}{6}h_i$ for $i = 2, \ldots, n-2$ and $r_{ij} = 0$ for $|i - j| \geq 2$. The matrix K is defined as $K = QR^{-1}Q'$.

Theorem 2.1 [Green and Silverman (1994, p.13).] *The vectors s and ξ specify a natural cubic spline s if and only if the condition*

$$Q's = R\xi \tag{2.6}$$

is satisfied. If (2.6) is satisfied then the roughness penalty will satisfy

$$\int_a^b s''(t)^2 dt = \xi' R \xi = s' K s. \tag{2.7}$$

Now we wish to express the cubic smoothing spline given by (2.4) by the combination of matrices Q, R and K. The estimator of the unknown curve $\hat{s}$ is a minimiser of $S(s)$ over the set of functions that are differentiable on $a < t_1 < \ldots < t_n < b$ and have absolutely continuous first derivatives. Let s be a natural cubic spline with vectors $\boldsymbol{s}$ and $\boldsymbol{\xi}$, and matrices Q and R. Let $\boldsymbol{Y}$ be the vector $(y_1, \ldots, y_n)'$. The penalised sum of squares (2.4) can be expressed as the residual sum of squares about s plus the roughness penalty term. The residual sum of squares can be rewritten as

$$\sum_{i=1}^{n} \left\{ y_i - s\left(t_i\right) \right\}^2 = (\boldsymbol{Y} - \boldsymbol{s})' (\boldsymbol{Y} - \boldsymbol{s}),$$

where the vector $\boldsymbol{s}$ is the vector of $s\left(t_i\right)$. Substituting (2.7) for the roughness penalty term the original equation (2.4) changes to

$$\begin{aligned} S(s) &= (\boldsymbol{Y} - \boldsymbol{s})' (\boldsymbol{Y} - \boldsymbol{s}) + \lambda \boldsymbol{s}' K \boldsymbol{s} \\ &= \boldsymbol{s}' (I + \lambda K) \boldsymbol{s} - 2\boldsymbol{Y}' \boldsymbol{s} + \boldsymbol{Y}' \boldsymbol{Y}. \end{aligned} \tag{2.8}$$

Equation (2.8) has a unique minimum, because λK is non-negative definite and the matrix $(I + \lambda K)$ is positive-definite. This minimum is given by

$$\boldsymbol{s} = (I + \lambda K)^{-1} \boldsymbol{Y}, \tag{2.9}$$

therefore (2.8) can be expressed as

$$\{s - (I + \lambda K)^{-1}\boldsymbol{Y}\}'(I + \lambda K)\{s - (I + \lambda K)^{-1}\boldsymbol{Y}\} \tag{2.10}$$

plus a constant that depends only on $\boldsymbol{Y}$. Thus, $S(s)$ has a unique minimum given by (2.9) over the space of all natural cubic splines with knots at the points t_i. All this is summarised in the definition of the cubic smoothing spline given by Theorem 2.2, where $S_2[a,b]$ specifies the set of the set of functions that are differentiable on $a < t_1 < \ldots < t_n < b$ and have absolutely continuous first derivative.

Theorem 2.2 [Green and Silverman (1994, p.19).] *Suppose* $n \geq 3$ *and* $t_1,\ldots,t_n$ *are points satisfying* $a < t_1 < t_2 < \ldots < t_n < b$. *Given data points* $(y_1,\ldots,y_n)$ *and a strictly positive smoothing parameter* λ, *let* $\hat{s}$ *be the natural cubic spline with knots at the points* $t_1, \ldots, t_n$ *for which* $\boldsymbol{s} = (I + \lambda K)^{-1}\boldsymbol{Y}$. *Then for any* s *in* $S_2[a,b]$,

$$S(\hat{s}) \leq S(s)$$

with equality only if s *and* $\hat{s}$ *are identical.*

For the purpose of working in higher dimensions, an alternative representation of the natural cubic spline is given in Green et al. (1994 on p.140). We refer the reader to

Green et al. (1994), which has become a standard theory reference in this field, for more details on nonparametric regression with a roughness penalty.

2.4.3 Application and extension

Eubank and Spiegelman (1990) investigated the use of non-parametric regression methodology to test the adequacy of a parametric linear model. Their test is based on fitting cubic smoothing splines to residuals and has good power properties against several reasonable alternatives. Gu (1990) applied the concept of adaptive spline smoothing in non-Gaussian regression models. In the paper by Schlitten (1991), smoothing splines were applied to time series decomposition. His proposed method used structural time series models and adapted smoothing splines, which were suitably robustified. Coulson (1992) used smoothing splines to generate a non-parametric response of housing prices to floor space. In the study of Rodriguez (1992), the constrained smoothing spline estimator was introduced into time-of-day electricity demand modelling. Fisher et al. (1995) described a technique for fitting the term structure of interest rates using smoothing splines. They showed how their technique can be used to spline an arbitrary transformation of the discount function. Rodriguez (1999) applied smoothing splines to the estimation of a model of investment behaviour of the U.S. telephone industry.

We need to make a few remarks on the choice of *regression splines* versus *smoothing splines* for smoothing data from the model (2.1). Wahba (1975) and subsequently Agarwal and Studden (1980) showed that the asymptotic behaviour of the two methods is the same, provided the smoothing parameters, namely the number of knots

and the smoothing constant λ, are both chosen optimally. It was found that for very large data sets, the regression spline with the optimal number of B-splines does not have such a good resolution to follow the local features of data as smoothing splines. A hybrid approach for very large data sets has been suggested by Nychka et al. (1984).

A new concept of regression and smoothing splines is spatially adaptive smoothing, proposed by Luo and Wahba (1997). Spatially adaptive smoothing can handle a wide variety of shapes and spatial inhomogeneity. Traditionally, two techniques have been used to address this problem of spatial adaptivity. One technique uses local variable smoothing parameters (or bandwidth) in common smoothing methods, such as smoothing splines. The other technique is to place knots adaptively in a regression spline method.

Luo and Wahba (1997) combined some of the features of adaptive regression splines and traditional smoothing splines to obtain a hybrid smoothing procedure termed *hybrid adaptive splines* (HAS). This combines features from both regression and smoothing spline approaches. One of its advantages is its ability to vary the amount of smoothing in response to the inhomogeneous "curvature" of the true functions at different locations. This method can be applied to many estimation problems. The method's performance in Luo and Wahba's (1997) simulation study was found to be comparable to the wavelet shrinkage methods proposed by Donoho and Johnstone (1994, 1995), and Donoho, Johnstone, Kerkyacharin and Picard (1995). The HAS procedure is well suited to highly unequally spaced data.

2.5 Models with smoothing splines

The theory regarding univariate spline functions in Sections 2.2 – 2.4 focussed on the simple dependence of observations y on one independent variable t. Recall (2.1) and write the general form of the simple model again as $Y_i = s(t_i) + \varepsilon(t_i)$. If $s(t)$ is estimated non-parametrically by the use of a cubic smoothing spline, this model represents a simple non-parametric model. In real situations there are usually other variables, which are thought to provide information on the behaviour of the dependent variable. Multiple explanatory variables leads to the multivariate regression model. The way in which these additional variables are incorporated into the model results in different forms of models. If all explanatory variables are believed to be linear or if it is possible to achieve their linearity by a suitable transformation, the resulting multivariate linear regression model can be expressed as

$$Y_i = \boldsymbol{x}_i'\gamma + \varepsilon_i, \tag{2.11}$$

where $\boldsymbol{x}_i$ is a k - vector of independent explanatory variables for the ith observation, γ is a k-vector of regression coefficients to be estimated and ε_i are iid $N(0,\sigma^2)$. In this case, the variable t is included in the vector $\boldsymbol{x}_i$.

In many situations in practice, for example in time series modelling, the trend t is very often non-linear. As an example we refer to Figure 2.7 that represents a plot of

Australian production of crude steel in thousands of tons within a period of January 1969 and April 2001, based on the monthly observations from the OECD Main Economic Indicators official databases.

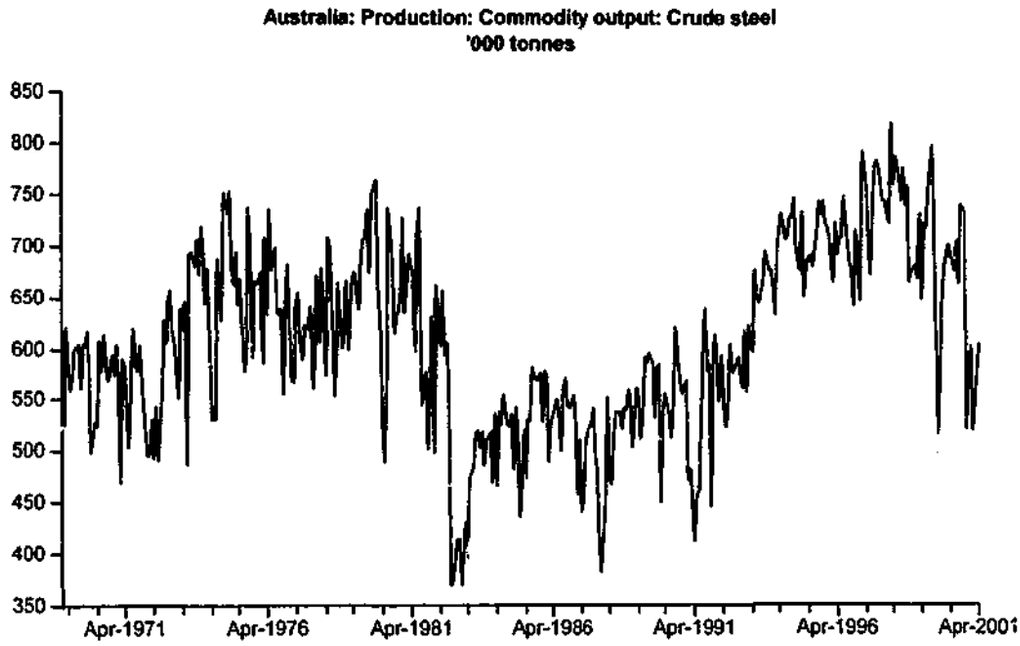

Figure 2.7: *Graph of Australian production of crude steel in '000* as an *example of a time series with a non-linear trend*

There is a possibility of being able to make this time series stationary in its mean, by for example, taking first or second order time differences. However there is no guarantee this will achieve stationarity in the mean, although this approach is often used in practice. Another way would be to model and extract the trend from the time series. In the case of a very complicated non-linear trend such as in Figure 2.7, the use

of higher order polynomial functions for modelling the trend component or its combination with any non-linear function might not necessarily lead to an appropriate form. A trend component like this seems to be best estimated non-parametrically.

In a modelling situation when other variables are assumed to influence the dependant variable in a linear way but the dependant variable itself shows a complicated non-linear trend, it makes sense to include the variable t into the model separately from the other predictor variables. The trend variable t can be estimated non-parametrically and the resulting model is of the form

$$Y_i = x_i'\gamma + s(t_i) + \varepsilon_i, \tag{2.12}$$

where for each observation Y_i, we have k explanatory variables, x_i is now a $k-1$ vector and t_i is a scalar, γ is a $(k-1)$-vector of regression coefficients to be estimated along with the unknown smooth curve s. The model (2.12) is known as a semiparametric model. If we are interested in fitting a semiparametric model (2.12) to the data, we can estimate γ and s by minimizing the penalized sum of squares

$$S(\gamma, s) = \sum_{i=1}^{n} \left\{ Y_i - x_i'\gamma - s(t_i) \right\}^2 + \lambda \int s''(t)^2 \, dt. \tag{2.13}$$

As Green and Silverman (1994, pp.64 – 65) show, the penalized sum of squares $S(\gamma, s)$ can be written as

$$(\boldsymbol{Y} - X\gamma - \boldsymbol{s})'(\boldsymbol{Y} - X\gamma - \boldsymbol{s}) + \lambda \int s''(t)^2\, dt, \tag{2.14}$$

where $\boldsymbol{Y}$ is the n-vector with ith component Y_i and X the $n \times (k-1)$ design matrix whose ith row is $\boldsymbol{x}_i'$. The minimizing curve $\boldsymbol{s}$ is the natural cubic spline with knots replacing $t_1, t_2, \ldots, t_n$. Recall (2.7) expression (2.14) can be written as

$$(\boldsymbol{Y} - X\gamma - \boldsymbol{s})'(\boldsymbol{Y} - X\gamma - \boldsymbol{s}) + \lambda \boldsymbol{s}'\boldsymbol{K}\boldsymbol{s}. \tag{2.15}$$

It follows that (2.15) is minimized when γ and $\boldsymbol{s}$ satisfy the block matrix equation:

$$\begin{bmatrix} X'X & X' \\ X & \lambda K \end{bmatrix} \begin{pmatrix} \gamma \\ \boldsymbol{s} \end{pmatrix} = \begin{bmatrix} X' \\ I' \end{bmatrix} \boldsymbol{Y}. \tag{2.16}$$

We refer to Green et al. (1994) for more about how effectively to solve the system given by (2.16).

The semiparametric model (2.12) is a reasonable compromise between the fully linear regression (2.11) and fully non-parametric modelling. The main advantage of the semiparametric model is, that the mean response is additive in the independent variables $x_1, \ldots, x_{k-1}, t$, even though it is non-linear in the variable t. If all the explanatory variables are believed to influence the response variable in a non-linear fashion, it might be useful to estimate all piece functions s_j related to the variables

$x_1,\ldots,x_{k-1},t$ non-parametrically. The resulting model is known as an additive nonparametric model and can be written in form

$$Y_i = \sum_{j=1}^{k} s_j\left(x_{ij}\right) + \varepsilon_i, \tag{2.17}$$

where x_{ij} is the value of the jth variable for the ith observation and $j = 1,\ldots,k$ with $x_{ik} = t_i$. The estimated piece functions s_j could all be linear, in this case model (2.17) will be the linear model given by (2.11). If all but one of the piece functions s_j are linear this situation is equivalent to the semiparametric model (2.12).

When an additive model (2.17) is estimated using the penalized least squares approach, the unknown non-linear functions $\{s_i\}$ are obtained by minimizing

$$S\left(s_1,\ldots,s_k\right) = \sum_{i=1}^{n}\left\{Y_i - \sum_{j=1}^{k} s_j\left(x_{ij}\right)\right\}^2 + \sum_{j=1}^{k}\lambda_j \int s_j''(x)^2\,dx, \tag{2.18}$$

where $\lambda_1,\ldots,\lambda_k$ are separate smoothing parameters for each of the piece smooth functions s_j.

Additive models have recently become very popular in applied work for their flexibility to model each variable separately. The application of spline functions to additive models was mentioned very briefly. The interested reader is refereed to Hastie and Tibshirani (1990) for more details.

2.6 Concluding remarks

This chapter reviewed the literature related to univariate polynomial splines. One observation is that the smoothing techniques are a driving force in the development of non-parametric regression techniques. Furthermore semiparametric regression models are attracting more and more attention. What all non-parametric techniques have in common is that they are computationally demanding. Therefore before the advent of high-speed computers, their theoretical development appeared to be of limited practical value. Since then smoothing techniques in regression have been increasingly applied in many economics and marketing fields.

In the remaining chapters of this thesis we will utilize the semiparametric model and describe the non-linear component of the model non-parametrically by smoothing splines. We will consider the smoothing spline model as a stochastic process. We will rewrite the semi-parametric model in a state space form and apply the model in non-linear inference.

CHAPTER 3

Stochastic process formulation

3.1 Introduction

This chapter is devoted to the stochastic process and state space form of smoothing splines. Suppose we observe a variable Y that depends on multiple explanatory variables. In a multiple linear regression, it would be assumed that dependence was linear and the theory of the general linear model would be used to estimate the model. We consider a semiparametric model where we relax the assumption of linearity on one of the explanatory variables which we shall call t, but retain the linear dependence on the remaining k variables. We write the model as

$$Y_i = x_i'\gamma + s(t_i) + e_i, \tag{3.1}$$

where x_i is a k-vector of explanatory variables, γ is a vector of regression coefficients, s is a smooth function of t, and the errors e_i are iid $N(0,\sigma^2)$. Model (3.1) is called a semiparametric model, because the response Y is assumed to depend in a parametric (linear) fashion on some, but not all, of the explanatory variables. In this thesis, we shall use cubic smoothing splines to estimate s.The original idea proposed by Wahba (1978) and developed in subsequent work of Wecker and Ansley (1983) links polynomial smoothing splines with stochastic processes. Wahba showed that polynomial smoothing splines could be obtained using a signal plus noise model where the signal is distributed as an integrated Wiener process of the same order. She showed that the polynomial smoothing spline is the conditional expectation of the signal given the variable that is observed. Wecker and Ansley (1983) used Wahba's model to develop a state space model for smoothing splines of all orders and suggested the use of Kalman filtering and related tools for estimation and prediction.

In Sections 3.2 – 3.4, we describe a smoothing spline as a stochastic process in a state space form. We apply the results of Wahba (1978) and Wecker and Ansley (1983) to cubic polynomial splines and include a linear term in the general statistical model. This model is similar to that discussed by Shively et al. (1994). We extend their model to allow covariates in Section 3.5. We express this semi-parametric model in state space form. We derive the error variance-covariance matrix for this model in the case of equally spaced data using Yule – Walker equations in Section 3.6, which is then used in further analysis in Chapters 4, 5 and 6 of this thesis. Further generalization of the covariance matrix derivation is also discussed. Section 3.7 is devoted to the application of the derived variance-covariance matrix to prediction. Section 3.8

highlights the significance of the theoretical work. Appendix 3A.1 gives a proof of the variance-covariance matrix derivation.

3.2 Integrated Wiener process

The stochastic process $\{Y(t): t \in [0,1]\}$ is said to be *Gaussian* if $\{Y(t_1),\ldots,Y(t_n)\}$ has a multivariate normal distribution for any finite collection of points $t_1,\ldots,t_n$ from $[0,1]$. To specify a Gaussian process we need to specify its mean value function $\mu(t) = \mathrm{E}[Y(t)]$ and its covariance function $K(l,t) = \mathrm{Cov}(Y(l), Y(t))$. A *white noise* process is a Gaussian process with mean zero and covariance function

$$K(l,t) = \begin{cases} v^2 & l = t \\ 0 & l \neq t. \end{cases}$$

A *Wiener process* $W = \{W(t): t \in [0,1]\}$ is a Gaussian process, which satisfies the following conditions

- $W(0) = 0$,
- W has independent increments, that is, if $u \le v \le l \le t$, then $W(t) - W(l)$ and $W(v) - W(u)$ are independent;
- $W(t+l) - W(t)$ is $N(0, v^2 l)$ for all $l, t \ge 0$ where v^2 is a positive constant.

The covariance function of W is given by $K(l,t)=v^2\min\{l,t\}$. W is called a *standard* Wiener process if $v^2=1$.

Let W denote a standard Wiener process and let

$$W_m(t)=\int_0^t \frac{(t-u)^m}{m!}dW(u), \quad 0\le t\le 1, \quad m=0,1,2,... \tag{3.2}$$

denote the *m-fold integrated* Wiener process, see Shepp (1966). Then,

$$W_0(t)=W(t),\, W_m(t)=\int_0^t W_{m-1}(u)\,du,\, m=1,2,...$$

The process W_m has a covariance function given by

$$K_m(l,t)=\left(\frac{1}{m!}\right)^2\int_0^1 (l-u)_+^m (t-u)_+^m\,du, \quad 0\le l,t\le 1, \tag{3.3}$$

where $(u)_+=u$ if $u\ge 0$ and 0 otherwise.

An m-fold Wiener processes can be written in a dynamic linear representation. Let

$$w_m(x,y)=\int_x^y \frac{(y-u)^m}{m!}dW(u).$$

Then for $x, y, z \in [0,1]$

$$w_m(x,y) = w_m(z,y) + \sum_{j=0}^{m-1} \frac{(y-z)^j}{j!} w_{m-j}(x,z) \tag{3.4}$$

and $W_m(t) = w_m(0,t)$. See Wecker and Ansley (1983) for a proof of this result.

3.3 Wahba's approach

Suppose that observations $(t_1, y_1), \ldots, (t_n, y_n)$ are obtained by sampling from a stochastic process $\{Y_c(t) : t \in [0,1]\}$ at predetermined points $t_1, \ldots, t_n$ where c is an arbitrary large number, theoretically $c \to \infty$. The process $Y_c(t)$ is assumed to be the sum of a polynomial of degree $m-1$ plus $(m-1)$-fold integrated Wiener process. Specifically, let

$$Y_c(t) = \mu(t) + e(t), \quad t \in [0,1], \tag{3.5}$$

in which $\{e(t)\}$ is a white noise process with zero mean and variance σ^2, and

$$\mu(t) = \sum_{j=0}^{m-1} \beta_j \frac{t^j}{j!} + \tau W_{m-1}(t), \tag{3.6}$$

where $\beta = (\beta_0, \ldots, \beta_{m-1})'$ is normally distributed with zero mean and covariance matrix cI and τ is a scaling parameter. The random vector β and the integrated Wiener process $W_{m-1}(t)$ are assumed to be uncorrelated.

Let $\mu_{n,\lambda}$ denote the smoothing spline of degree $2m-1$ and smoothing parameter λ, defined as the function, which minimizes

$$\sum_{j=1}^{n}\left(y_j - s\left(t_j\right)\right)^2 + \lambda \int_0^1 \left[s^{(m)}(l)\right]^2 dl$$

over all functions with $2m-2$ continuous derivatives on $[0,1]$. Wahba (1978) showed that the smoothing spline $\mu_{n,\lambda}(\cdot)$ is the mean of the stochastic process $\mu(t)$ if the distribution of β is diffuse. That is,

$$\mu_{n,\lambda}(t) = \lim_{c\to\infty} E\left\{\mu(t) | \mathbf{Y} = \mathbf{y}\right\}, \tag{3.7}$$

where $\mathbf{Y} = \left(Y_1, \ldots, Y_n\right)'$, $\mathbf{y} = \left(y_1, \ldots, y_n\right)'$, $\lambda = \sigma^2 / \tau^2$. The smoothing parameter λ is controlling the "rate of exchange" between the residuals error described by the residuals sum of squares and local variation represented by the squared integral of the mth derivative of s. In practice, λ is not generally known and it should be estimated. Chapter 6 of this thesis is devoted to the problem of choosing the smoothing parameter λ.

3.4 State space form

Wecker and Ansley (1983) proposed a state space version of Wahba's stochastic process model (3.6), which enables the estimation of the parameters λ and σ^2 as well as an estimate for $\mu(t)$ together with confidence bounds.

If in (3.4), we let $t = t_1$, $z = t_{i-1}$, $y = t_i$, then we have for $i = 2,\ldots,n$,

$$w_m\left(t_1,t_i\right) = w_m\left(t_{i-1},t_i\right) + \sum_{j=0}^{m} \frac{\left(t_i - t_{i-1}\right)^j}{j!} w_{m-j}\left(t_1,t_{i-1}\right). \tag{3.8}$$

Now define

$$\boldsymbol{\alpha}_i = \tau\left(w_{m-1}\left(t_1,t_i\right), w_{m-2}\left(t_1,t_i\right),\ldots,w_0\left(t_1,t_i\right)\right)'$$

$$\boldsymbol{u}_i = \tau\left(w_{m-1}\left(t_{i-1},t_i\right), w_{m-2}\left(t_{i-1},t_i\right),\ldots,w_0\left(t_{i-1},t_i\right)\right)'$$

and let $T(u)$ be the $m \times m$ matrix

$$T(u) = \begin{bmatrix} 1 & u & \cdots & \dfrac{u^{m-1}}{(m-1)!} \\ 0 & 1 & \cdots & \dfrac{u^{m-2}}{(m-2)!} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Then (3.8) implies that

$$\alpha_i = T_i \alpha_{i-1} + u_i, \tag{3.9}$$

where $T_i = T(h_i)$, $h_1 = t_1$ and $h_i = t_i - t_{i-1}$.

Note that Wecker and Ansley (1983) derived the covariance matrix of u_i incorrectly. As pointed out by Mazzi and de Jong (1998), the covariance matrix of u_i, which we denote by V_i, has (j,k)th cntry

$$\begin{aligned}[V_i]_{jk} &= \tau^2 \int_{t_{i-1}}^{t_i} \frac{(t_i - u)^{m-j}(t_i - u)^{m-k}}{(m-j)!(m-k)!} du \\ &= \tau^2 \frac{h_i^{2m-(j+k)+1}}{(2m-(j+k)+1)(m-j)!(m-k)!}.\end{aligned} \tag{3.10}$$

Letting $r = (1,0,\ldots,0)'$, $\beta = (\beta_0,\ldots,\beta_{m-1})'$, Wahba's model (3.5) and (3.6) can be written in state space form as

$$Y_i = z_i'\beta + r'\alpha_i + e_i \tag{3.11}$$

and

$$\alpha_i = T_i \alpha_{i-1} + u_i, i = 1,\ldots,n, \tag{3.12}$$

where e_i are iid $N(0,\sigma^2)$ and $z_i = \left(1, t_i, \frac{t_i^2}{2!}, \ldots, \frac{t_i^{m-1}}{(m-1)!}\right)'$ with starting condition $\alpha_0 = (0,\ldots,0)'$. It follows from the definitions that α_{i-1} is independent of u_i. Note the difference in the same notation z in (3.4) and (3.11).

3.5 State space form with covariates

Let us assume *cubic* smoothing splines $(m=2)$ and modify Wahba's model by adding some covariates as follows

$$Y_i = x_i'\gamma + s(t_i) + e_i, \tag{3.13}$$

where e_i are iid $N(0,\sigma^2)$, $s(t_i) = z_i\beta + g(t_i)$, and

$$g(t_i) = \tau W_1(t_i) = \tau \int_0^{t_i} (t-u)\,dW(u).$$

The additional covariates are given by $x_i = (x_{i,1},\ldots,x_{i,k})'$. Let $e = \{e_1,\ldots,e_n\}'$, $s = \{s(t_1),\ldots,s(t_n)\}'$ and $g = \{g(t_1),\ldots,g(t_n)\}'$. Then, as observed by Ansley et al. (1994), the model can be written in matrix form as

$$Y = X\gamma + Z\beta + g + e \tag{3.14}$$

where the ith row of X is x_i', the ith row of Z is $z_i' = (1, t_i)$, $E(g) = 0$, $Var(g) = \lambda^{-1}\Sigma$, $E(s) = 0$, $Var(s) = cZZ' + \lambda^{-1}\Sigma$, $E(\beta) = \mathbf{0}$, and $\mathrm{Var}(\beta) = cI$. We use the notation Σ for an unknown variance-covariance matrix of the nonlinear component $g(t_i)$ in $s(t_i)$, recall (3.13). The structure of Σ will be derived in Section 3.6.

If $\tau = 0$, then $g(t) \equiv 0$ and $s(t)$ is a linear function of t. In this case, solving the Kalman recursion is equivalent to letting $\lambda \to \infty$ and fitting an ordinary least squares regression line.

Following the approach of Wecker and Ansley (1983), we model (3.14) in state space form. Let $\alpha_i = \tau\{W_1(t_i) - W_1(t_1), W(t_i) - W(t_1)\}$. Then

$$Y_i = x_i'\gamma + z_i'\beta + (1,0)\alpha_i + e_i \tag{3.15}$$

and

$$\alpha_i = T_i\alpha_{i-1} + u_i,\ i = 1, \ldots, n, \tag{3.16}$$

where $u_i = \tau \begin{bmatrix} \int_{t_{i-1}}^{t_i} (t_i - u)\, dW(u) \\ W(t_i) - W(t_{i-1}) \end{bmatrix}$ and $T_i = \begin{bmatrix} 1 & (t_i - t_{i-1}) \\ 0 & 1 \end{bmatrix}$.

Mazzi and de Jong (1998) give the same state space formulation although they derived it differently, beginning with a continuous time version. Note that equation (5) in Mazzi and de Jong is incorrect, the exponent of e inside the integral should be $A(\delta - t)$ and the variable of the integration should be $dW(t)$.

From (3.10), we have

$$V_i = \tau^2 \int_0^{h_i} \begin{bmatrix} (h_i - u)^2 & (h_i - u) \\ (h_i - u) & 1 \end{bmatrix} du = \tau^2 \begin{bmatrix} \frac{h_i^3}{3} & \frac{h_i^2}{2} \\ \frac{h_i^2}{2} & h_i \end{bmatrix}. \tag{3.17}$$

3.6 Calculation of Σ

3.6.1 Derivation of Σ using cubic smoothing spline

We next derive the error variance–covariance matrix Σ. Let $\Gamma_i(j) = E(\alpha_i \alpha_{i-j}')$, $j = 0,1,\ldots$. Note α_i is vector autoregressive model of order one in (3.16). Thus we obtain the Yule-Walker equations

$$\begin{aligned} &\Gamma_0(0) = 0, \\ &\Gamma_i(0) = V_i + T_i \Gamma_{i-1}(0) T_i', \quad i = 1,2,\ldots,n, \\ &\Gamma_i(j) = T_i \Gamma_{i-1}(j-1), \quad j = 1,2,\ldots. \end{aligned} \tag{3.18}$$

Note that $\Gamma_i(j)=0, j \geq i$. We can use these equations to iteratively calculate the values of $\Gamma_i(j)$ for $i = 1,2,\ldots,n$ and $j = 1,2,\ldots$. Then the (i,j)th element of $\lambda^{-1}\Sigma$ is the top left element of $\Gamma_j(j-i)$ if $i \leq j$ and the top left element of $\Gamma_i(i-j)$ if $i \geq j$. For more on Yule-Walker equations, see Reinsel (1997).

Let us consider the special case of equally spaced time, so $h_i = t_i - t_{i-1} = h$. Therefore

$$T_i = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} \tag{3.19}$$

and

$$V_i = \tau^2 \begin{bmatrix} \frac{h^3}{3} & \frac{h^2}{2} \\ \frac{h^2}{2} & h \end{bmatrix} \tag{3.20}$$

for all i. By substituting (3.19) and (3.20) into (3.18), we can construct Σ. It is symmetric with the (i,j)th element on or above the diagonal given by

$$\Sigma_{ij} = \Sigma_{ji} = \sigma^2 h^3 i^2 (3j-i)/6, \ j \geq i.$$

Thus

$$\Sigma = \frac{\sigma^2 h^3}{6}\begin{bmatrix} 2 & 5 & \cdots & 3n-1 \\ 5 & 16 & & \vdots \\ \vdots & & \ddots & \vdots \\ 3n-1 & \cdots & \cdots & 2n^3 \end{bmatrix}. \tag{3.21}$$

We refer to Appendix 3 A.1 for a detailed derivation of expression (3.21).

3.6.2 Modification of Σ using higher order smoothing splines

In this section we briefly outline the method for the extension of the variance-covariance matrix Σ to a higher order smoothing spline. We need to modify the covariance matrix V_i of u_i. Recall (3.10) and for a smoothing spline of order m we have the $m \times m$ matrix V_i of the form

$$V_i = \tau^2 \begin{bmatrix} \dfrac{h^{(2m-1)}}{(2m-1)(m-1)!(m-1)!} & \dfrac{h^{(2m-2)}}{(2m-2)(m-1)!(m-2)!} & \cdots & \dfrac{h^m}{m!} \\ \dfrac{h^{(2m-2)}}{(2m-2)(m-2)!(m-1)!} & \dfrac{h^{(2m-3)}}{(2m-3)(m-2)!(m-2)!} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{h^m}{m!} & \cdots & \cdots & h \end{bmatrix} \tag{3.22}$$

and the $m \times m$ matrix T_i is modified to the form

$$T_i = \begin{bmatrix} 1 & h & \cdots & \frac{h^{(m-1)}}{(m-1)!} \\ 0 & 1 & \cdots & \frac{h^{(m-2)}}{(m-2)!} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{3.23}$$

To derive the variance-covariance Σ we apply the Yule-Walker equations (3.18). By substituting (3.24) and (3.25) into (3.18), the variance-covariance Σ can be derived for any m.

3.6.3 Further generalization

In Sections 3.6.1 and 3.6.2, we considered the special case of equally spaced time, so $h_i = t_i - t_{i-1} = h$ was constant. The further extension of the derivation of the variance-covariance matrix Σ can be made to non-equally spaced time, so $h_i = t_i - t_{i-1} \neq const$. The covariance matrix V_i of u_i, expression (3.20) or (3.22), needs to be modified for a non-constant time increment. Also matrix T_i given by expression (3.19) or (3.23) needs to be adjusted for a non constant time increment. Yule-Walker equations given by (3.18) could be applied for the derivation of the variance-covariance matrix Σ. This is however mathematically very intensive and has not been pursued here.

3.7 Prediction

In this section our major concern is in predicting the values of the outcome variable Y, for various levels of the explanatory variables.

We have been using the past n observations so far in this chapter to derive the semi-parametric regression model (3.13) and to express the model in the state space form (3.15) – (3.16). In this section, we would like focus on the prediction of n_0 future observations, which will create a vector Y_0 of $n_0 \times 1$ observations.

Recall model (3.13) and rewrite this model in the general form -

$$Y = X\gamma + \nu, \tag{3.24}$$

where $\nu = Z\beta + g + e$ so that $E(\nu) = 0$, $Var(\nu) = E(\nu\nu') = \sigma^2\Omega$ and Ω and Σ are $n \times n$ positive definite symmetric matrices. Equation (3.13) implies that the variance of disturbance vector ν can be written as

$$Var(\nu) = \sigma^2\left(cZZ' + \lambda^{-1}\Sigma + I_n\right). \tag{3.25}$$

Matrix Ω is known and we have derived its Σ component in Section 3.6 in (3.21). The ith row of Z is $z_i' = (1, t_i)$ and $i = 1, \ldots, n$.

In the prediction problem, we assume that the model

$$Y_0 = X_0\gamma + \nu_0, \tag{3.26}$$

will generate the future observations, where the $n_0 \times k$ matrix X_0 is assumed to be known and the disturbance vector ν_0 is assumed to have the same properties as the past disturbance vector ν, so $E(\nu_0) = 0$, $Var(\nu_0) = E(\nu_0\nu_0') = \sigma^2\Omega_0$ and Ω_0, Σ are now $n_0 \times n_0$ positive definite symmetric matrices, Σ is given by (3.21). The variance-covariance of disturbance vector ν_0 can be written as $Var(\nu_0) = \sigma^2\left(cZZ' + \lambda^{-1}\Sigma + I_{n_0}\right)$, where the ith row of Z is $z_i' = (1, t_i)$ and $i = n+1, \ldots, n+n_0$. Under the general disturbance variance-covariance matrix $\sigma^2\Omega$ in a general regression problem it is possible that the past observations may contain some information about future disturbance values ν_0.

In the following, we apply the finding of Judge et al. (1988, pp 343-346) to our model (3.13) in order to derive the best linear unbiased predictor for Y_0 and the variance-covariance matrix of the prediction error ν_0. First assume that parameters c and λ are known. In reality, constant c is any sufficiently large number and the smoothing parameter λ needs to be estimated from the data. In order to make our predictions, let us combine the model (3.26) and (3.24) and write the combined model as

$$\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{Y}_0 \end{bmatrix} = \begin{bmatrix} X \\ X_0 \end{bmatrix} \gamma + \begin{bmatrix} \boldsymbol{\nu} \\ \boldsymbol{\nu}_0 \end{bmatrix}, \tag{3.27}$$

where the combined variance-covariance matrix for $(\boldsymbol{\nu}, \boldsymbol{\nu}_0)'$ is given by

$$E\left[\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\nu}_0 \end{pmatrix}\begin{pmatrix} \boldsymbol{\nu}' & \boldsymbol{\nu}_0' \end{pmatrix}\right] = \sigma^2 \begin{bmatrix} \Omega & U \\ U' & \Omega_0 \end{bmatrix} = \sigma^2 \boldsymbol{G}. \tag{3.28}$$

Now the covariance matrix of the future disturbance vector $\boldsymbol{\nu}_0$ is $n_0 \times n_0$ which we denote by $\sigma^2 \boldsymbol{\Omega}_0$ and the covariances between the elements of $\boldsymbol{\nu}$ and $\boldsymbol{\nu}_0$ makes the $n \times n_0$ matrix $\sigma^2 U$. The variance-covariance matrix of the combined disturbance vector $(\boldsymbol{\nu}, \boldsymbol{\nu}_0)'$ implies

$$Var(\boldsymbol{\nu}, \boldsymbol{\nu}_0)' = \sigma^2 \left(cZZ' + \lambda^{-1}\Sigma + I_{n+n_0}\right), \tag{3.29}$$

where Σ is modified from (3.21) to

$$\Sigma = \frac{\sigma^2 h^3}{6} \begin{bmatrix} 2 & 5 & \cdots & 3n-1 & \cdots & 3(n+n_0)-1 \\ 5 & 14 & & & & \vdots \\ \vdots & & \ddots & & & \\ 3n-1 & & & 2n^3 & & \\ \vdots & & & & \ddots & \vdots \\ 3(n+n_0)-1 & \cdots & & & \cdots & 2(n+n_0)^3 \end{bmatrix},$$

therefore matrices $\boldsymbol{\Omega}$, $\boldsymbol{\Omega}_0$ and U are known in our stochastic model.

We want to derive the best linear predictor and the variance-covariance matrix of its prediction error. We introduce the transformation matrix C such that

$$CGC' = I_{n+n_0}. \tag{3.30}$$

Suppose that the matrix C is a lower triangular matrix so that (3.30) can be written as

$$\begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \Omega & U \\ U' & \Omega_0 \end{bmatrix} \begin{bmatrix} C'_{11} & C'_{12} \\ 0 & C'_{22} \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_{n_0} \end{bmatrix}. \tag{3.31}$$

Now we multiply (3.27) by C and get

$$\begin{bmatrix} \boldsymbol{Y}^* \\ \boldsymbol{Y}_0^* \end{bmatrix} = \begin{bmatrix} C_{11}\boldsymbol{Y} \\ C_{21}\boldsymbol{Y} + C_{22}\boldsymbol{Y}_0 \end{bmatrix}, \tag{3.32}$$

$$\begin{bmatrix} X^* \\ X_0^* \end{bmatrix} = \begin{bmatrix} C_{11}X \\ C_{21}X + C_{22}X_0 \end{bmatrix}, \tag{3.33}$$

$$\begin{bmatrix} \boldsymbol{\nu}^* \\ \boldsymbol{\nu}_0^* \end{bmatrix} = \begin{bmatrix} C_{11}\boldsymbol{\nu} \\ C_{21}\boldsymbol{\nu} + C_{22}\boldsymbol{\nu}_0 \end{bmatrix}. \tag{3.34}$$

therefore matrices Ω, Ω_0 and U are known in our stochastic model.

We want to derive the best linear predictor and the variance-covariance matrix of its prediction error. We introduce the transformation matrix C such that

$$CGC' = I_{n+n_0}. \tag{3.30}$$

Suppose that the matrix C is a lower triangular matrix so that (3.30) can be written as

$$\begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \Omega & U \\ U' & \Omega_0 \end{bmatrix} \begin{bmatrix} C'_{11} & C'_{12} \\ 0 & C'_{22} \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_{n_0} \end{bmatrix}. \tag{3.31}$$

Now we multiply (3.27) by C and get

$$\begin{bmatrix} \boldsymbol{Y}^* \\ \boldsymbol{Y}_0^* \end{bmatrix} = \begin{bmatrix} C_{11}\boldsymbol{Y} \\ C_{21}\boldsymbol{Y} + C_{22}\boldsymbol{Y}_0 \end{bmatrix}, \tag{3.32}$$

$$\begin{bmatrix} X^* \\ X_0^* \end{bmatrix} = \begin{bmatrix} C_{11}X \\ C_{21}X + C_{22}X_0 \end{bmatrix}, \tag{3.33}$$

$$\begin{bmatrix} \boldsymbol{\nu}^* \\ \boldsymbol{\nu}_0^* \end{bmatrix} = \begin{bmatrix} C_{11}\boldsymbol{\nu} \\ C_{21}\boldsymbol{\nu} + C_{22}\boldsymbol{\nu}_0 \end{bmatrix}. \tag{3.34}$$

Thus the transformed model is given by

$$\begin{bmatrix} \boldsymbol{Y}^* \\ \boldsymbol{Y}_0^* \end{bmatrix} = \begin{bmatrix} X^* \\ X_0^* \end{bmatrix} \gamma + \begin{bmatrix} \boldsymbol{\nu}^* \\ \boldsymbol{\nu}_0^* \end{bmatrix}, \tag{3.35}$$

where

$$E\left[\begin{pmatrix} \boldsymbol{\nu}^* \\ \boldsymbol{\nu}_0^* \end{pmatrix} \begin{pmatrix} \boldsymbol{\nu}^{*\prime} & \boldsymbol{\nu}_0^{*\prime} \end{pmatrix}\right] = \sigma^2 I_{n+n_0}. \tag{3.36}$$

Equation (3.35) implies that the best linear unbiased predictor for $\boldsymbol{Y}_0^*$ is

$$\hat{\boldsymbol{Y}}_0^* = X_0^* \hat{\gamma}, \tag{3.37}$$

where $\hat{\gamma}$ is the GLS estimator given by

$$\begin{aligned} \hat{\gamma} &= \left(X^{*\prime} X^*\right)^{-1} X^{*\prime} \boldsymbol{Y}^* \\ &= \left(X' C_{11}' C_{11} X\right)^{-1} X' C_{11}' C_{11} \boldsymbol{Y} \\ &= \left(X' \boldsymbol{\Omega}^{-1} X\right)^{-1} X' \boldsymbol{\Omega}^{-1} \boldsymbol{Y}, \end{aligned} \tag{3.38}$$

where $C_{11}' C_{11} = \boldsymbol{\Omega}^{-1}$ can be derived from (3.31).

We need to transform (3.37) to get the best linear unbiased predictor $\boldsymbol{Y}_0$. Equation (3.32) implies that $\boldsymbol{Y}_0 = C_{22}^{-1}\boldsymbol{Y}_0^* - C_{22}^{-1}C_{21}\boldsymbol{Y}$. Thus the best linear unbiased predictor for $\boldsymbol{Y}_0$ is

$$\begin{aligned}
\hat{\boldsymbol{Y}}_0 &= C_{22}^{-1}\hat{\boldsymbol{Y}}_0^* - C_{22}^{-1}C_{21}\boldsymbol{Y} \\
&= C_{22}^{-1}X_0^*\hat{\gamma} - C_{22}^{-1}C_{21}\boldsymbol{Y} \\
&= \left(C_{22}^{-1}C_{21}X + X_0\right)\hat{\gamma} - C_{22}^{-1}C_{21}\boldsymbol{Y} \\
&= X_0\hat{\gamma} - C_{22}^{-1}C_{21}\left(\boldsymbol{Y} - X\hat{\gamma}\right) \\
&= X_0\hat{\gamma} + U'\boldsymbol{\Omega}^{-1}\left(\boldsymbol{Y} - X\hat{\gamma}\right),
\end{aligned} \tag{3.39}$$

where $C_{22}'C_{21} = -U'\boldsymbol{\Omega}^{-1}$ can be derived from (3.31). The $X_0\hat{\gamma}$ component in (3.39) is the best linear unbiased estimator for $X_0\gamma$ and the $U'\Omega^{-1}\left(\boldsymbol{Y} - X\hat{\gamma}\right)$ component is the best linear unbiased predictor for $\boldsymbol{\nu}_0$.

Now we need to derive the variance-covariance matrix of the prediction error. The covariance matrix of the prediction error of the transformed data is given by the expression

$$E\left[\left(\hat{\boldsymbol{Y}}_0^* - \boldsymbol{Y}_0^*\right)\left(\hat{\boldsymbol{Y}}_0^* - \boldsymbol{Y}_0^*\right)'\right] = \sigma^2\left[I + X_0^*\left(X^{*\prime}X^*\right)^{-1}X_0^{*\prime}\right]. \tag{3.40}$$

To get the variance-covariance matrix of the prediction error $\left(\hat{\boldsymbol{Y}}_0 - \boldsymbol{Y}_0\right)$, we observe $\left(\hat{\boldsymbol{Y}}_0 - \boldsymbol{Y}_0\right) = C_{22}^{-1}(\hat{\boldsymbol{Y}}_0^* - \boldsymbol{Y}_0^*)$, so

$$E\left[\left(\hat{\boldsymbol{Y}}_0-\boldsymbol{Y}_0\right)\left(\hat{\boldsymbol{Y}}_0-\boldsymbol{Y}_0\right)'\right]=C_{22}^{-1}E\left[(\hat{\boldsymbol{Y}}_0^*-\boldsymbol{Y}_0^*)(\hat{\boldsymbol{Y}}_0^*-\boldsymbol{Y}_0^*)'\right]C_{22}'^{-1}. \qquad (3.41)$$

We substitute (3.40) into (3.41) and get

$$\begin{aligned}
&E\left[\left(\hat{\boldsymbol{Y}}_0-\boldsymbol{Y}_0\right)\left(\hat{\boldsymbol{Y}}_0-\boldsymbol{Y}_0\right)'\right]\\
&=\sigma^2\left[C_{22}^{-1}C_{22}'^{-1}+C_{22}^{-1}X_0^*\left(X^{*\prime}X^*\right)^{-1}X_0^{*\prime}C_{22}'^{-1}\right]\\
&=\sigma^2\left[\boldsymbol{\Omega}_0-U'\boldsymbol{\Omega}^{-1}U+\left(C_{22}^{-1}C_{21}X+X_0\right)\left(X^{*\prime}X^*\right)^{-1}\left(X'C_{21}'C_{22}'^{-1}+X_0'\right)\right] \qquad (3.42)\\
&=\sigma^2\left[\boldsymbol{\Omega}_0-U'\boldsymbol{\Omega}^{-1}U+\left(X_0-U'\boldsymbol{\Omega}^{-1}X\right)\left(X'\boldsymbol{\Omega}^{-1}X\right)^{-1}\left(X_0'-X'\boldsymbol{\Omega}^{-1}U\right)\right],
\end{aligned}$$

where $C_{11}'C_{11}=\boldsymbol{\Omega}^{-1}$, $C_{22}'C_{21}=-U'\boldsymbol{\Omega}^{-1}$ and $C_{22}^{-1}C_{22}'^{-1}=\boldsymbol{\Omega}_0-U'\boldsymbol{\Omega}U$ can be derived from (3.31).

In this section we have derived the best linear unbiased predictor in (3.39) and the variance-covariance matrix of its prediction error in (3.42). Knowing $\boldsymbol{\Omega}, \boldsymbol{\Omega}_0$ and U, we can create prediction intervals for predictions. The random variable

$$\frac{\left\{\left[X_0\hat{\gamma}+U'\Omega^{-1}\left(\boldsymbol{Y}-X\hat{\gamma}\right)\right]-\boldsymbol{Y}_0\right\}_i}{\sigma\sqrt{\left[\Omega_0-U'\Omega^{-1}U+\left(X_0-U'\Omega^{-1}X\right)\left(X'\Omega^{-1}X\right)^{-1}\left(X_0'-X'\Omega^{-1}U\right)\right]_{ii}}}$$

is a standard normal variable with mean 0 and a variance 1. That means that the random variable

$$\frac{\dfrac{\left\{\left[X_0\hat{\gamma}+U'\Omega^{-1}(\boldsymbol{Y}-X\hat{\gamma})\right]-\boldsymbol{Y}_0\right\}_i}{\sigma\sqrt{\left[\Omega_0-U'\Omega^{-1}U+\left(X_0-U'\Omega^{-1}X\right)\left(X'\Omega^{-1}X\right)^{-1}\left(X_0'-X'\Omega^{-1}U\right)\right]_{ii}}}}{\sqrt{\dfrac{(n-k)\hat{\sigma}^2/\sigma^2}{n-k}}}=$$

$$\frac{\left\{\left[X_0\hat{\gamma}+U'\Omega^{-1}(\boldsymbol{Y}-X\hat{\gamma})\right]-\boldsymbol{Y}_0\right\}_i}{\hat{\sigma}\sqrt{\left[\Omega_0-U'\Omega^{-1}U+\left(X_0-U'\Omega^{-1}X\right)\left(X'\Omega^{-1}X\right)^{-1}\left(X_0'-X'\Omega^{-1}U\right)\right]_{ii}}} \tag{3.43}$$

is distributed as a t random variable with $(n-k)$ degrees of freedom. Let $\left[Y_0\right]_i=Y_{n+i}$, $\hat{Y}_{n+i}=\left[X_0\hat{\gamma}+U'\Omega^{-1}(\boldsymbol{Y}-X\hat{\gamma})\right]$ and the variance $Var\left(\hat{Y}_{n+i}\right)=\hat{\sigma}^2\left[\Omega_0-U'\Omega^{-1}U+\left(X_0-U'\Omega^{-1}X\right)\left(X'\Omega^{-1}X\right)^{-1}\left(X_0'-X'\Omega^{-1}U\right)\right]$. Now expression (3.43) can be written as

$$\frac{Y_{n+i}-\hat{Y}_{n+i}}{\sqrt{Var\left(\hat{Y}_{n+i}\right)}}.$$

We may write

$$\Pr\left[-t_{\left(n-k,\alpha/2\right)}\le t_{(n-k)}\le t_{\left(n-k,\alpha/2\right)}\right]=1-\alpha,$$

or

$$\Pr[-t_{(n-k,\alpha/2)} \le \frac{Y_{n+i} - \hat{Y}_{n+i}}{\sqrt{Var\left(\hat{Y}_{n+i}\right)}} \le t_{(n-k,\alpha/2)}] = 1-\alpha,$$

which can be rewritten as

$$\left[\hat{Y}_{n+1} - t_{\left(n-k,\frac{\alpha}{2}\right)}\sqrt{Var\left(\hat{Y}_{n+1}\right)}, \hat{Y}_{n+1} + t_{\left(n-k,\frac{\alpha}{2}\right)}\sqrt{Var\left(\hat{Y}_{n+1}\right)}\right].$$

This prediction interval predicts $\left[Y_0\right]_i$ with probability $(1-\alpha)$ that the value of the random variable is contained within it.

This analysis has assumed that the parameters c and λ are known. As noted earlier, in reality λ is estimated and c is taken as any large value. Obviously λ in the above formulae would be replaced by $\hat{\lambda}$ and we might also replace Student's t percentiles with standard normal percentiles. While we can no longer claim that $\hat{Y}_{n+i}$ is the best linear unbiased predictor it should be asymptotically if $\hat{\lambda}$ is a consistent estimator of λ.

3.8 Conclusions

In this chapter we assumed the stochastic model in state space form given by Wahba (1978), Wecker and Ansley (1983) and Mazzi and de Jong (1998) for which smoothing splines provide optimal estimates. A new method of variance-covariance matrix Σ derivation using Yule – Walker equations was presented. Matrix Σ was derived for equally spaced data in time and for cubic smoothing splines. We will use this matrix for hypothesis testing namely marginal likelihood ratio based tests for linearity in Chapter 4 and Chapter 5 of this thesis. The generalization of the derivation of Σ by considering smoothing splines of higher order and not equally spaced data in time was briefly discussed. We also presented the derivation of the best linear unbiased predictor and the variance-covariance matrix of this prediction error applying the stochastic model with smoothing splines. This was used for further computation of prediction intervals.

Appendix 3 A.1: Derivation of Σ for cubic smoothing spline

We shall prove (3.21). First we show by induction that

$$\Gamma_i(0) = \tau^2 \begin{bmatrix} i^3 \dfrac{h^3}{3} & i^2 \dfrac{h^2}{2} \\ i^2 \dfrac{h^2}{2} & ih \end{bmatrix} \tag{A 3.1}$$

For $i = 0$, $\Gamma_0(0) = 0$, so (A 3.1) is true. Now assume (A 3.1) is true for $i = k$. Then from (3.18) we obtain

$$\Gamma_{k+1}(0) = \tau^2 \begin{bmatrix} \dfrac{h^3}{3} & \dfrac{h^2}{2} \\ \dfrac{h^2}{2} & h \end{bmatrix} + \tau^2 \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k^3 \dfrac{h^3}{3} & k^2 \dfrac{h^2}{2} \\ k^2 \dfrac{h^2}{2} & kh \end{bmatrix} \begin{bmatrix} 1 & 0 \\ h & 1 \end{bmatrix}$$

$$= \tau^2 \begin{bmatrix} (k+1)^3 \dfrac{h^3}{3} & (k+1)^2 \dfrac{h^2}{2} \\ (k+1)^2 \dfrac{h^2}{2} & (k+1)h \end{bmatrix}.$$

So it is true for $i = k + 1$ and by induction is true for $i = 1, 2, 3, \ldots$. Now from (3.18) we have

$$\Gamma_i(j) = T\Gamma_{i-1}(j-1) = T^2\Gamma_{i-2}(j-2) = T^j\Gamma_{i-j}(0) \quad \text{for } i \geq j,$$

and so $\Gamma_i(i-j) = T^{i-j}\Gamma_j(0)$. Thus

$$\Gamma_i(i-j) = \tau^2 \begin{bmatrix} 1 & (i-j)h \\ 0 & 1 \end{bmatrix} \begin{bmatrix} j^3\dfrac{h^3}{3} & j^2\dfrac{h^2}{2} \\ j^2\dfrac{h^2}{2} & jh \end{bmatrix} = \tau^2 \begin{bmatrix} h^3 j^2 \dfrac{(3i-j)}{6} & jh^2\dfrac{(2i-j)}{2} \\ j^2\dfrac{h^2}{2} & jh \end{bmatrix}$$

and so

$$\Sigma_{ij} = \sigma^2 h^3 j^2 \frac{(3i-j)}{6}$$

for $i \geq j$. Hence we obtain (3.21).

CHAPTER 4

Testing for the inclusion of a possibly nonlinear component

4.1 Introduction

Recall that in (3.1) we have a semiparametric regression model with one potential nonlinear component. An important problem in working with model (3.1) is testing whether there is a nonlinear component. Various methods have been proposed for this testing problem. Yanagimoto and Yanagimoto (1987) proposed a likelihood ratio test using a stochastic model. Cox and Koh (1989) proposed a locally most powerful test for linearity. Cox, Koh, Wahba, and Yandell (1988) extended this test to a semiparametric regression model. Both these tests use test statistics assuming that the error variance is known. Cox and Koh (1989) normalized the test statistic by a non-parametric estimate of the error variance. Buckley (1991) proposed an exact test using a different non-parametric estimate of the error variance. Munson and Jernigan (1989)

constructed an exact test by fitting an interpolating cubic spline to the residuals of the linear regression model. Eubank and Spiegelman (1990) proposed a test, which requires a non-parametric estimate of the error variance and a choice of smoothing parameter. Azzalini and Bowman (1993) suggested an F-type test statistic using kernel semiparametric regression. Shively, Kohn and Ansley (1994) proposed an exact point optimal invariant test for linearity, which is based on Wahba's (1978) stochastic model and is invariant to the values of the regression coefficients and error variance. Their evaluation technique is based on Kalman filtering. Their critical values have to be calculated via numerical integration, which in some circumstances, is computationally cumbersome.

In this chapter, we apply the findings of Wahba (1978) and Wecker and Ansley (1983) to cubic polynomial splines. This model is similar to that discussed by Shively et al. (1994). We assume an extended model with extra covariates. We apply the error variance-covariance matrix for this model in the case of equally spaced data, which has been derived in Chapter 3. We propose a likelihood ratio one-sided test based on the marginal likelihood function as was derived by Tunnicliffe Wilson (1989) and Ara (1995) in Section 4.2. We investigate the performance of the test via a Monte Carlo experiment in Section 4.3 and compare our marginal likelihood ratio (MLR) test with the regression specification error test by Ramsey (1969) and a related test in Section 4.4. The comparison between the mean based tests and the variance based test is a part of this section. We apply the MLR test to some real economic data in Section 4.5. Section 4.6 summarizes our findings.

4.2 Marginal likelihood ratio test

In this section we introduce the basic concepts of one-sided testing along with likelihood and marginal likelihood principles. These concepts underpin our marginal likelihood based tests.

4.2.1 One-sided testing

Many economic concepts are represented by parameters in models and the range of possible values of these parameters may be restricted by statistical, functional and logical reasons. In hypothesis testing, this leads to testing problems in which the null hypothesis is on the boundary of the parameter space.

Our MLR test, to be discussed in more detail in Section 4.2.4, and our MLR* test to be discussed in Chapter 5, are one-sided tests. The smoothing parameter λ which is controlling the "rate of exchange" between the residual errors described by the residual sum of squares and local variation is defined on a set of nonnegative numbers, hence leading to inequality restricted estimation and one-sided testing of $\lambda = 0$.

Research on one-sided testing started in the early 1950's. The main results for testing the mean of a multivariate normal distribution with a likelihood ratio procedure were presented in the book by Barlow et al. (1972). Also the likelihood-based approach is captured in Robertson et al.'s (1988) book. King and Evans (1984) observed a

significant increase in power when a one-sided Lagrange multiplier test was used instead of the two-sided test. Wu and King (1994) summarized a range of important findings on one-sided hypothesis testing and also confirmed that these types of tests result in greater power. They distinguished two main approaches to the testing problem. One is based on local optimization of power through the use of the Neyman–Pearson lemma. A variation of this approach is to construct a test that is most powerful among all the tests of the same size at a chosen point in the alternative parameter space. This is known as a point optimal (PO) test. An alternative is to construct tests by maximizing power locally at the null hypothesis. These tests are known as locally best (LB) tests. The second approach is to construct one-sided versions of conventional likelihood ratio (LR), Wald and Lagrange multiplier (LM) tests. These are likelihood-based tests and involve maximization of the likelihood function with respect to parameters subject to the inequality restrictions. Estimation of unknown parameters is a part of the test procedure.

The problem with testing against an inequality restricted alternative hypothesis is how to incorporate this additional one-sided information into the test. The conventional two-sided test statistics are typically quadratic forms of maximum likelihood estimators. When the number of parameters under test is one, appropriately signed square roots of these statistics can be applied as one-sided tests.

Consider a general density function $f(y|\theta,\phi)$ of an $n\times 1$ random vector y, where $\phi \in \Phi$ is a $q\times 1$ nuisance parameter vector, Φ is a subset of R^q, $\theta \in \Theta$ is a $p\times 1$ parameter vector under test and Θ is a subset of R^p. The problem of interest is to test

$$H_0 : \theta = \theta_0$$

against

$$H_a : \theta > \theta_0 .$$

We assume that θ_0 is an interior point of Θ. This assumption is needed as the derivation of the asymptotic distribution of the likelihood-based test statistic relies on the Taylor's expansion of the score function around θ_0.

By Wu and King (1994), with $p = 1$ and log likelihood function $L(\theta, \phi)$ induced by $f(y|\theta, \phi)$, a signed square root of the two-sided likelihood ratio statistic R is defined by

$$R = \operatorname{sgn}\left(\hat{\theta} - \theta_0\right)\left[2\left(L\left(\hat{\theta}, \hat{\phi}\right) - L\left(\theta_0, \hat{\phi}_0\right)\right)\right]^{1/2}, \tag{4.1}$$

which under H_0 is asymptotically distributed $N(0,1)$, where $\operatorname{sgn}(y) = 1, -1, 0$ if $y > 0, y < 0, y = 0$, respectively and $\hat{\theta}, \hat{\phi}$ are the maximum likelihood estimates of θ, ϕ under H_a and $\hat{\phi}_0$ is the maximum likelihood estimate of ϕ under H_0. The upper tail of the $N(0,1)$ distribution is used as the critical region.

Instead of taking the square root of the two-sided likelihood ratio test statistic as in (4.1), another approach to testing against an inequality restricted alternative is to

follow the traditional way of constructing one-sided tests. This involves comparing the log-likelihood function under H_a to that under H_0 for the likelihood ratio test. Gourieroux et al. (1980), (also see Farebrother (1988)), considered the one-sided testing problem and derived a one-sided likelihood ratio test statistic as

$$s_{LR} = 2\left(L\left(\tilde{\delta}\right) - L\left(\hat{\delta}_0\right)\right), \tag{4.2}$$

where $\hat{\delta}_0$ is the maximum likelihood estimate of δ under the H_0. They assumed that $\delta = \left(\theta', \phi'\right)'$ and $\delta_0 = \left(\theta_0', \phi_0'\right)'$ where ϕ_0 is the true unknown value of the parameter ϕ. They also assumed that ϕ_0 is an interior point of Φ. Also let φ be the score vector that corresponds to θ, scaled by dividing by sample size, that is, $\varphi = n^{-1}\,\partial L/\partial\theta$. Let $\tilde{\delta} = \left(\tilde{\theta}', \tilde{\phi}'\right)'$ be the solution of

$$\max n^{-1} L(\theta, \phi), \text{ subject to } \theta \geq \theta_0, \phi \in \Phi$$

so

$$\tilde{d} = n^{-1} \left.\frac{\partial L}{\partial \delta}\right|_{\delta=\tilde{\delta}} = \begin{bmatrix} \tilde{\varphi} \\ 0 \end{bmatrix} \tag{4.3}$$

and

$$\tilde{\varphi}'\left(\tilde{\theta}-\theta_0\right)=0, \tag{4.4}$$

where (4.4) is the Kuhn-Tucker condition, see Kuhn and Tucker (1951). This means that the components of $\tilde{\varphi}$ in (4.3) must be either zero or negative. This is because of a presence of a positive component in $\tilde{\varphi}$ implies that $L(\theta,\phi)$ has not yet been maximized. Equation (4.3) is the definition of $\tilde{\varphi}$. Gourieroux et al. (1980) noted that $\hat{\delta}_0=\left(\theta_0',\hat{\phi}_0'\right)'$ is the solution of

$$\max n^{-1}L(\theta,\phi), \text{ subject to } \theta\geq\theta_0,\phi\in\Phi,$$

so

$$\tilde{d}_0=n^{-1}\frac{\partial L}{\partial\delta}\bigg|_{\delta=\tilde{\delta}_0}=\begin{bmatrix}\tilde{\varphi}_0\\0\end{bmatrix}.$$

The asymptotic distribution of s_{LR} under H_0 is given by

$$\Pr\left(s_{LR}<c\right)=\sum_{i=0}^{p}w(p,i)\Pr\left(\chi_i^2<c\right) \tag{4.5}$$

for $c\in R$. This is a probability mixture of independent chi-squared distributions, χ_i^2, with different degrees of freedom, e.g. χ_0^2 is degenerate distribution at zero. The weights $w(p,i)$, $i=0,\ldots,p$, represent the asymptotic probability of the event that

under H_0 any i elements of $\tilde{\theta}-\theta_0$ are strictly positive, and the remaining $p-i$ elements are zeros, thus $\sum_{i=0}^{p} w(p,i)=1$. By (4.3) and (4.4), this is also the event that i corresponding elements in $\overset{\square}{\varphi}$ are zero and the remaining $p-i$ elements are strictly negative. Thus, the weights are dependent on the distribution of $\overset{\square}{\varphi}$ which is information matrix, and therefore density function, $f(y|\theta,\phi)$, specific. Also $w(p,0)$ is the asymptotic probability that strictly $\tilde{\theta}-\theta_0<0$ and χ_0^2 is the degenerate distribution with unit mass at zero. Let α be the significance level of the test. If the $w(p,i)$'s are known, the positive value c satisfying $\Pr(s_{LR}>c)=\alpha$ asymptotically under H_0 can be found by solving

$$\alpha=\sum_{i=0}^{p} w(p,i)\Pr\left(\chi_i^2>c\right). \tag{4.6}$$

Note that $w(p,0)$ is not involved so that α cannot be larger than $1-w(p,0)$ as asymptotically $\Pr(s_{LR}>0)=1-w(p,0)$ and $\Pr(s_{LR}=0)=w(p,0)$. Because $w(p,0)\le\frac{1}{2}$ in all cases, this causes no difficulty for the conventional choice of α.

Our testing approach in this thesis is based on a likelihood function, therefore the following section will describe the basic likelihood and marginal likelihood concepts.

4.2.2 Likelihood concepts

Presume that we want to estimate unknown parameter θ about which we do not have enough information. If we have a random sample represented by the vector $\boldsymbol{y}$, our intention is to reveal the appropriate information about θ from this random sample. Various information about θ will be collected from different data. The likelihood function is a function that collects all the information that the sample has about θ.

The likelihood function is the most common basis for statistical inference. Fisher (1922) introduced the concept of likelihood. He also introduced the concept of consistency, sufficiency and efficiency of parameter estimates and promoted the use of maximum likelihood methods. The maximum likelihood estimate is that value of the unknown parameter vector of interest θ, which makes the likelihood a maximum given that particular sample. Maximum likelihood estimation, under prescribed regularity conditions, (see for example Lehmann (1983) or Stuart and Ord (1991)), gives consistent, asymptotically efficient and asymptotically unbiased estimates. This implies that if a sample of n observations $\boldsymbol{y} = \left(y_1, y_2, \ldots, y_n\right)'$ is taken from a population with a probability density function $f(\boldsymbol{y};\theta)$, where θ is a scalar, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{a} N\left(0, \lim_{n\to\infty} n\left(I(\theta)\right)^{-1}\right),$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ, 'a' denotes asymptotically distributed as, and $I(\theta) = -E\left(\frac{\partial^2 \log f(y;\theta)}{\partial\theta^2}\right)$ is Fisher's information, which ensures that $\hat{\theta}$ is close to θ or minimizes information loss for sufficiently large samples.

The likelihood function usually involves several parameters, only some of which are of interest. The other parameters, known as nuisance parameters, are necessary for the model validity, however their values are not important for the conclusions to be formulated. The problem of estimating θ in the presence of a nuisance parameter ϕ requires the modification of the usual concept of Fisher's information. It is important that no information is lost in making an inference about θ in the absence or presence of ϕ. Basu (1977) and Barndoff-Nielsen (1978) gave detailed reviews of this problem. Research work by Remon (1984), Lloyd (1987), Bhapker (1989, 1991) is an illustration of many scientists who emphasized the use of marginal likelihood based procedures rather than conventional likelihood when estimating θ in the presence of the nuisance parameter ϕ.

4.2.3 Marginal Likelihood

In this section we will concentrate on methods for eliminating the nuisance parameter from the likelihood function, focussing on marginal likelihood based procedures. We are interested in eliminating ϕ from the likelihood function $f(y;\theta,\phi)$ so that inference about θ can be made. The marginal likelihood is based on a suitable transformation model, which eliminates ϕ from the likelihood function without loss

of information. The marginal distribution function therefore depends only on θ. There exist different theories on how to construct the marginal likelihood. We will focus on the 'factoring one' method, which leads to the marginal likelihood function for our testing problem to be discussed in Section 4.2.4.

Kalbfleisch and Sprott (1970) developed the marginal likelihood, which arises from density factorization. This theory is based on factoring the likelihood into two parts, one part being uninformative about θ in the absence of knowledge of ϕ and the other part containing θ only. Assume that the joint distribution $f(y;\theta,\phi)$ of the sample factorizes into two factors

$$f(y;\theta,\phi) = f(t_1;\theta) f(t_2 | t_1;\theta,\phi)$$

by a non-singular transformation of y to an $n_1 \times 1$ vector t_1 and an $n_2 \times 1$ vector t_2. The first factor $f(t_1;\theta)$ can be used for inferences concerning θ without knowing ϕ and the likelihood that arises from that factor is known as the marginal likelihood. In more general context, when t_1 is a function of θ, the quantity $dt_1(\theta)$ is a function of θ. The other factor $f(t_2 | t_1;\theta,\phi)$ is used for inferences about unknown parameter θ in the presence of knowledge of a nuisance parameter ϕ. By Kalbfleisch and Sprott (1970) the $y_i, i = 1,\ldots,n,$ are the Cartesian co-ordinates of a point in Euclidean space of dimension n with distance metric given by $(ds)^2 = \Sigma(dy_i)^2$. Assuming this the marginal likelihood of θ is proportional to

$$\frac{f(t_1;\theta)\left|K^tK\right|^{1/2}}{\left|\ddot{L}\right|}, \tag{4.7}$$

where K is the matrix defined with elements $\left(\frac{\partial y_i}{\partial t_{2j}}\right)$ and $\ddot{L}$ is the Jacobian of this transformation of y to (t_1, t_2), that is

$$\ddot{L} = \frac{\partial(y_1, y_2, \ldots, y_n)}{\partial\left(t_{11}, t_{12}, \ldots, t_{1n_1}, t_{21}, t_{22}, \ldots, t_{2n_2}\right)}.$$

Barndoff-Nielsen (1983) proposed the use of the density of a maximum likelihood estimator conditional on ancillary statistic u. His factorization is of the form

$$f(y;\theta,\phi) = f(u;\theta) f\left(\hat{\phi}_0 \middle| u;\theta,\phi\right),$$

where $\hat{\phi}_0$ is the maximum likelihood estimate of ϕ, therefore a marginal likelihood for θ, based on $f(u;\theta)$ can be approximated as

$$L(\theta) \approx L\left(\theta,\hat{\phi}_0\right)\left|J_{\phi\phi}\left(\theta,\hat{\phi}_0\right)\right|^{-1/2} J(\theta), \tag{4.8}$$

where $L\left(\theta,\hat{\phi}_0\right)$ is the profile likelihood for θ, $J_{\phi\phi}$ is the observed information on ϕ for fixed θ and $J(\theta)=\dfrac{\partial\hat{\phi}}{\partial\hat{\phi}_0}$ for fixed u. Barndoff-Nielsen (1983) called (4.8) the modified profile likelihood.

Cox and Reid (1987) modified (4.8) by eliminating the Jacobian factor and making ϕ to be orthogonal to θ. Their factorisation is of the form

$$f(y;\theta,\phi)=f\left(y\middle|\hat{\phi}_0;\theta\right)f\left(\hat{\phi}_0;\theta,\phi\right).$$

The approximate conditional likelihood of the density of $\hat{\phi}_0$ is obtained from $f\left(y\middle|\hat{\phi}_0;\theta\right)$, where $\hat{\phi}_0$ is accessible from a saddle-point approximation theory under certain assumptions.

The previous discussion was an introduction to the likelihood concepts based on the general distribution $f(y;\theta,\phi)$ for the variable y in order to give a general foundation to marginal likelihood. We are going to use the idea of marginal likelihood in the context of the linear regression model. We wish to test whether possibly nonlinear component was not left out of the model, hence we propose the MLR test in Section 4.2.4. In Chapter 5 we apply the marginal likelihood approach to the other testing problem. We wish to test whether possibly nonlinear component which should be included in the model is or is not linear, hence in Chapter 5 we propose the MLR* test.

Consider the linear regression model

$$y = X\beta + u, \tag{4.9}$$

where y is an n-dimensional random vector of the dependant variable, X is an $n \times k$ matrix of observations on k non-stochastic variables, k is the rank of matrix X, β is a k-dimensional vector of unknown parameters and u is n-dimensional vector of random errors with variance-covariance matrix $\sigma^2\Omega(\theta)$. The random error is a stochastic process depending on the parameter θ to be estimated. Let us assume that the errors in (4.9) are normally distributed, so $u \sim N\left(0, \sigma^2\Omega(\theta)\right)$. The error term can be written as follows

$$u = Xb + c,$$

where $b = \left(X'X\right)^{-1} X'u$ and $c = u - Xb = Mu$. Matrix M is a symmetric matrix with rank $n-k$ and has $n-k$ characteristic roots equal to one, the remaining k roots are zero. The marginal distribution of c is

$$h(c) = \int_{R^k} f(Xb + c)\left|X'X\right|^{1/2} db. \tag{4.10}$$

When deriving the density of $c = My$ directly from (4.9), the Jacobian factor needed for the likelihood function, involves the determinant of the singular matrix M, which

is zero. Ara (1995) found matrix M analytically by the use of the eigenvectors. Her derivation is based on the existence of orthogonal matrix P whose rows form the set of analogous eigenvectors of M, for which

$$PMP' = \begin{bmatrix} I_{n-k} & 0 \\ 0 & 0_k \end{bmatrix}$$

and $PP' = P'P = I$. Separation the P matrix into the $(n-k)\times n$ matrix P_1 and the $k \times n$ matrix P_2

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix},$$

where $P_1MP_1' = I_{n-k}$ and $P_2MP_2' = 0_k$. Also $P_2M\boldsymbol{y} = P_2M\boldsymbol{u} = 0$ because the rows of P_2 are linearly dependent on the columns of X. By Kadiyala (1970), this is valid for any value of θ and does not have any effect on the inference regarding θ. We therefore make an inference using $\boldsymbol{c} = M\boldsymbol{y}$ or $\boldsymbol{d} = P_1M\boldsymbol{y} = P_1\boldsymbol{u}$.

Now return back to the parameterization of the vector $\boldsymbol{y}$, which has distribution $N(X\beta, \sigma^2\Omega(\theta))$, $P_1\boldsymbol{y}$ therefore has distribution $N\left[P_1X\beta, \sigma^2\left(P_1\Omega(\theta)P_1'\right)\right]$ and $\boldsymbol{d} = P_1\boldsymbol{u}$ has distribution $N\left[0, \sigma^2\left(P_1\Omega(\theta)P_1'\right)\right]$. Consequently the density of $\boldsymbol{d}$ is expressed as

$$f(\boldsymbol{d}) = \left(2\pi\sigma^2\right)^{-m/2} \left|P_1\Omega(\theta)P_1'\right|^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\boldsymbol{d}'\left(P_1\Omega(\theta)P_1'\right)^{-1}\boldsymbol{d}\right], \qquad (4.11)$$

where $m = n - k$ is the number of degrees of freedom. The distribution of $\boldsymbol{d} = P_1 u$ can be expressed as a function of the distance from the origin. By King (1979), for $\boldsymbol{d}'\boldsymbol{d} = r^2$,

$$\int_0^\infty h(r) r^{\frac{m}{2}-1} dr = \Gamma\left(\frac{m}{2}\right)\pi^{\frac{m}{2}}.$$

The resulting statistic vector $\boldsymbol{v} = \dfrac{\boldsymbol{d}}{\left(\boldsymbol{d}'\boldsymbol{d}\right)^{1/2}}$ from dividing $\boldsymbol{d}$ by its distance r, is a unit vector defined in terms of $(m-1)$ co-ordinates. King (1980) showed that $\boldsymbol{v}$ is a maximal invariant with respect to transformations of the form $y^* \to \vartheta_0 y + X\vartheta$, where $\vartheta_0 > 0$ is a positive scalar and θ is $k \times 1$. The distribution of $\boldsymbol{v}$ can be expressed as

$$\begin{aligned} f(\boldsymbol{v})\,dv &= \frac{1}{2}\Gamma\left(\frac{m}{2}\right)\pi^{-\frac{m}{2}}\left|P_1\Omega(\theta)P_1'\right|^{-\frac{1}{2}}\left\{v'\left(P_1\Omega(\theta)P_1'\right)^{-1}v\right\}^{-\frac{m}{2}} dv \\ &= \frac{1}{2}\Gamma\left(\frac{m}{2}\right)\pi^{-\frac{m}{2}}\left|P_1\Omega(\theta)P_1'\right|^{-\frac{1}{2}} G^{-\frac{m}{2}} dv, \end{aligned} \qquad (4.12)$$

where $G = \boldsymbol{v}'\left(P_1\Omega(\theta)P_1'\right)^{-1}\boldsymbol{v} = \dfrac{\hat{u}'\,\Omega(\theta)^{-1}\,\hat{u}}{\boldsymbol{c}'\boldsymbol{c}}$, $\boldsymbol{c}$ is the OLS residual vector, $\hat{\boldsymbol{u}}$ is the GLS residual vector assuming variance-covariance matrix $\sigma^2\Omega(\theta)$ and dv denotes the

uniform measure on the surface of the unit m-sphere. Observe that the distribution given in (4.12) is independent of β and σ.

We need to explain that the maximal invariant is defined with respect to a set of transformations on the data. It is a statistic that is always invariant under these transformations but for data vectors not connected by a transformation, the maximal invariant takes different values. Hence it is maximal in the sense of having the largest number of different values. Any invariant test statistic can be written as a function of a maximal invariant. For more explanation on concepts of invariance and maximal invariant based tests see Lehmann (1986).

The marginal likelihood function given by Tunnicliffe Wilson (1989) used the same statistic v and is expressed as

$$f_m(\theta|y) = |\Omega(\theta)|^{-\frac{1}{2}} \left|X'\Omega(\theta)^{-1} X\right|^{-\frac{1}{2}} \left(\hat{u}' \Omega(\theta)^{-1} \hat{u}\right)^{-\frac{m}{2}}. \tag{4.13}$$

By Verbyla (1990)

$$\left|P_1\Omega(\theta)P_1\right| = \left|X'X\right|^{-1} |\Omega(\theta)| \left|X'\Omega(\theta)^{-} X\right|. \tag{4.14}$$

The expression (4.12) can be simplified by use of (4.13) and (4.14). Because any invariant test statistic can be written as a function of a maximal invariant and because of the "maximal" property of a maximal invariant all invariant tests can be constructed by treating v as the observed data and (4.12) as its density function. We

will use the simplified version of (4.12) in Section 4.2.4.for parameter κ estimation, to construct likelihood ratio MLR and MLR* based tests and for smoothing parameter λ estimation in Chapter 6.

Ara (1995) constructed the likelihood ratio test based on the marginal likelihood function (4.12) and showed that the likelihood ratio statistic is asymptotically distributed as χ_p^2 when the null hypothesis is true, where p is number of parameters being estimated. Her derivation of the asymptotic distribution of $\hat{\theta}_0$ is based on the set of regularity conditions outlined by Godfrey (1988) and used in the context of the classical likelihood as follows:

(i) If the parameter space ψ is closed, bounded and finite dimensional, the true parameter vector is an interior point of ψ.

(ii) Two different values of θ have different probability distributions for v.

(iii) The first, second and third order partial derivatives of the log likelihood function $L(\theta)$ with respect to $\theta_i, i = 1,2,...,p$, exist for all $\theta \in \psi$ and should be continuous throughout some neighbourhood of the true parameter value θ_0.

(iv) The expected value $E\left(\frac{\partial L(\theta_0)}{\partial \theta}\right) = 0$ and the variance-covariance matrix $\operatorname{var}\left(\frac{\partial L(\theta_0)}{\partial \theta}\right) = -E\left(\frac{\partial^2 L(\theta_0)}{\partial \theta \partial \theta'}\right) = I(\theta_0)$, where $I(\theta_0)$ is the information matrix and θ_0 is the true parameter value.

(v) The matrix $I(\theta_0)$ needs to be positive definitive at $\theta = \theta_0$ and locally identifiable, $m \geq p$. This requires a lack of perfect multicollinearity in X and the $\Omega(\theta)$ matrix to be identified for all θ.

(vi) The information increases as the sample sizes increases as a consequence of the eigenvalues of $I(\theta_0)$ tending to infinity as $m \to \infty$.

(vii) By a Law of Large Number applied to the second derivative matrix of the log likelihood, the null hypothesis of the true parameter value θ_0 we assume

$$-\left[I(\theta_0)\right]^{-1} \frac{\partial^2 L(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{p} I_p$$

or

$$-\left[\frac{\partial^2 L(\theta_0)}{\partial \theta \partial \theta'}\right]^{-1} \frac{\partial^2 L(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{p} I_p,$$

where 'p' means convergences in probability and θ is in a neighbourhood of θ_0.

(viii) The Central Limit Theorem for martingales applies to the score vector under rather general conditions (see Crowder (1976)) as

$$I\left(\theta_0\right)^{-\frac{1}{2}} \frac{\partial L\left(\theta_0\right)}{\partial \theta} \xrightarrow{D} N\left(0, I_p\right),$$

where 'D' means convergence in distribution and $I\left(\theta_0\right)^{-\frac{1}{2}}$ is the positive definite square root of $I\left(\theta_0\right)$.

(ix) $\lim_{m \to \infty} m^{-1} I\left(\theta_0\right)$ exists as a finite non-singular matrix.

Ara's (1995) derivation of the asymptotic distribution of $\hat{\theta}$ is based on the distribution of the score function. She showed that the asymptotic distribution of $I\left(\theta_0\right)^{-\frac{1}{2}}\left(\hat{\theta}-\theta_0\right) \xrightarrow{D} N\left(0, I_p\right)$ and $\left(\hat{\theta}-\theta_0\right)' I\left(\theta_0\right)\left(\hat{\theta}-\theta_0\right) \xrightarrow{D} \chi_p^2$ under the null hypothesis of $\theta = \theta_0$. The same author also derived the asymptotic distribution of the likelihood ratio test statistic by using a function of the score vector and the information matrix by means of a Taylor's series expansion of $L\left(\theta_0\right)$ about $\hat{\theta}$ as

$$\begin{aligned} LR &= 2\left[L\left(\hat{\theta}\right) - L\left(\theta_0\right)\right] \\ &= 2\left[L\left(\hat{\theta}\right) - L\left(\hat{\theta}\right) - \left(\theta_0 - \hat{\theta}\right)\frac{\partial L\left(\hat{\theta}\right)}{\partial \theta} - \frac{1}{2}\left(\theta_0 - \hat{\theta}\right)' \frac{\partial^2 L\left(\hat{\theta}\right)}{\partial \theta \partial \theta'}\left(\theta_0 - \hat{\theta}\right) + R\right]. \end{aligned}$$

Ara (1995) showed that the distribution of the LR statistic depends on the distribution of

$$\frac{1}{2}\left(\theta_0 - \hat{\theta}\right)' \frac{\partial^2 L\left(\hat{\theta}\right)}{\partial\theta\partial\theta'}\left(\theta_0 - \hat{\theta}\right)$$

evaluated at $\theta = \hat{\theta}$, because $\dfrac{\partial L\left(\hat{\theta}\right)}{\partial\theta}$ is zero and the residue R is asymptotically insignificant. Also the finite bound assumption on the third derivative and the consistency of $\hat{\theta}$ under the null implies that

$$\frac{\partial^2 L\left(\hat{\theta}\right)}{\partial\theta\partial\theta'} \xrightarrow{p} \frac{\partial^2 L\left(\theta_0\right)}{\partial\theta\partial\theta'},$$

as a consequence of taking the Taylor's series expansion about $\theta = \hat{\theta}$. The LR statistic is asymptotically distributed as χ^2_p under the null, because from the Weak Law of Large Numbers the use of assumption (vii) for LR statistic is justified.

4.2.4 Test for the inclusion of a possibly nonlinear component – MLR test

Suppose we have a semiparametric regression model (3.1) with one potential nonlinear component and with additional covariates as in Section 3.5 of this thesis. Recalling (3.14) and separating out the linear part of the model, we can rewrite the model as follows

$$\boldsymbol{Y} = X\gamma + \boldsymbol{Q} \tag{4.15}$$

where $\boldsymbol{Q} = Z\boldsymbol{\beta} + \boldsymbol{g} + \boldsymbol{e}$. Now $\mathrm{E}(\boldsymbol{Q}) = \boldsymbol{0}$ and the variance is $\Lambda = \mathrm{Var}(\boldsymbol{Q}) = cZZ' + \kappa\Sigma + \sigma^2 I$, with $c \to \infty$ and $\kappa = \lambda^{-1}$. We can substitute the error variance-covariance form Σ given by (3.21) into this expression in order to test for inclusion of the nonlinear component in model (3.1). We use the null hypothesis

$$H_0 : \kappa = 0, \quad \text{i.e.,} \boldsymbol{Q} \sim \mathrm{N}\left(0, cZZ' + \sigma^2 I\right)$$

against the specific alternative

$$H_a : \kappa > 0, \quad \text{i.e.,} \boldsymbol{Q} \sim N\left(0, cZZ' + \kappa\Sigma + \sigma^2 I\right).$$

If $\kappa = 0$, $g(t_i) \equiv 0$ and $s(t_i)$ is a linear function of t, recall (3.13) and (3.14). We want to test the model $Y_i = x_i'\gamma + s(t_i) + e_i$ being restricted under H_0 with $s(t_i) = \beta_0$ being constant versus an unrestricted model under H_a with $s(t_i) = \beta_0 + g(t_i)$, where $g(t_i)$ represents the non-linear component in $s(t_i)$. We therefore formulate this type of test as a test for the inclusion of a possibly nonlinear component, so the null hypothesis is that there is no need to include an additional (possibly nonlinear) component into the model. The alternative hypothesis is that there is a need to include a additional component into the model.

By Tunnicliffe Wilson (1989) and Ara (1995), the marginal likelihood function (4.12)-(4.13) is modified for our semiparametric regression (3.14) as

$$l(\kappa|y) = |\Lambda|^{-1/2} \left|X'\Lambda^{-1}X\right|^{-1/2} \left(\hat{Q}'\Lambda^{-1}\hat{Q}\right)^{-N/2} \tag{4.16}$$

where $\hat{Q}$ is the GLS residual vector from (4.15) assuming variance-covariance matrix Λ and $N = n - p$ is the number of degrees of freedom. In the special case where there are no regressors, this reduces to

$$l(\kappa|y) = |\Lambda|^{-1/2} \left(\hat{Q}'\Lambda^{-1}\hat{Q}\right)^{-N/2}. \tag{4.17}$$

4.2.4 Test for the inclusion of a possibly nonlinear component – MLR test

Suppose we have a semiparametric regression model (3.1) with one potential nonlinear component and with additional covariates as in Section 3.5 of this thesis. Recalling (3.14) and separating out the linear part of the model, we can rewrite the model as follows

$$\boldsymbol{Y} = X\gamma + \boldsymbol{Q} \tag{4.15}$$

where $\boldsymbol{Q} = Z\boldsymbol{\beta} + \boldsymbol{g} + \boldsymbol{e}$. Now $\mathrm{E}(\boldsymbol{Q}) = \boldsymbol{0}$ and the variance is $\Lambda = \mathrm{Var}(\boldsymbol{Q}) = cZZ' + \kappa\Sigma + \sigma^2 I$, with $c \to \infty$ and $\kappa = \lambda^{-1}$. We can substitute the error variance-covariance form Σ given by (3.21) into this expression in order to test for inclusion of the nonlinear component in model (3.1). We use the null hypothesis

$$H_0 : \kappa = 0, \quad \text{i.e.,} \boldsymbol{Q} \sim \mathrm{N}\left(0, cZZ' + \sigma^2 I\right)$$

against the specific alternative

$$H_a : \kappa > 0, \quad \text{i.e.,} \boldsymbol{Q} \sim N\left(0, cZZ' + \kappa\Sigma + \sigma^2 I\right).$$

If $\kappa = 0$, $g(t_i) \equiv 0$ and $s(t_i)$ is a linear function of t, recall (3.13) and (3.14). We want to test the model $Y_i = x_i'\gamma + s(t_i) + e_i$ being restricted under H_0 with $s(t_i) = \beta_0$ being constant versus an unrestricted model under H_a with $s(t_i) = \beta_0 + g(t_i)$, where $g(t_i)$ represents the non-linear component in $s(t_i)$. We therefore formulate this type of test as a test for the inclusion of a possibly nonlinear component, so the null hypothesis is that there is no need to include an additional (possibly nonlinear) component into the model. The alternative hypothesis is that there is a need to include a additional component into the model.

By Tunnicliffe Wilson (1989) and Ara (1995), the marginal likelihood function (4.12)-(4.13) is modified for our semiparametric regression (3.14) as

$$l(\kappa|y) = |\Lambda|^{-1/2} \left|X'\Lambda^{-1}X\right|^{-1/2} \left(\hat{Q}'\Lambda^{-1}\hat{Q}\right)^{-N/2} \tag{4.16}$$

where $\hat{Q}$ is the GLS residual vector from (4.15) assuming variance-covariance matrix Λ and $N = n - p$ is the number of degrees of freedom. In the special case where there are no regressors, this reduces to

$$l(\kappa|y) = |\Lambda|^{-1/2} \left(\hat{Q}'\Lambda^{-1}\hat{Q}\right)^{-N/2}. \tag{4.17}$$

Let C be the lower-triangular matrix from the Choleski decomposition of Λ and let $\left\{\hat{\omega}_i\right\}$ denote the residuals from the ordinary least squares regression of $\boldsymbol{Y}_*$ on X_* where $X_* = C^{-1}X$ and $\boldsymbol{Y}_* = C^{-1}\boldsymbol{Y}$. Then, we can write

$$|\Lambda|^{-\frac{1}{2}} = |C|^{-1}, \tag{4.18}$$

$$\left|X'\Lambda X\right|^{-\frac{1}{2}} = \left|X_*'X_*\right|^{-\frac{1}{2}}, \tag{4.19}$$

$$\left(\hat{Q}'\Lambda^{-1}\hat{Q}\right)^{-\frac{N}{2}} = \left(\sum_{i=1}^{n}\hat{\omega}_i^2\right)^{-\frac{N}{2}}. \tag{4.20}$$

By (4.18) – (4.20), equation (4.16) is simplified to the form

$$l(\kappa|y) = |C|^{-1}\left|X_*'X_*\right|^{-1/2}\left(\sum_{i=1}^{n}\hat{\omega}_i^2\right)^{-N/2}. \tag{4.21}$$

so that the marginal log-likelihood is given by

$$\log l(\kappa|y) = -\log|C| - \frac{1}{2}\log\left|X_*'X_*\right| - \frac{N}{2}\log\left(\sum_{i=1}^{n}\hat{\omega}_i^2\right). \tag{4.22}$$

We estimate κ by minimizing (4.22). Then we can formulate a test statistic by applying the marginal log-likelihood ratio (MLR) test as follows

$$\text{MLR} = 2\left[\log l\left(\hat{\kappa}|y\right) - \log l\left(0|y\right)\right]. \tag{4.23}$$

Note that MLR is independent of σ under H_0. To see this, recall that $\mathrm{E}(\boldsymbol{Q}) = 0$ and we can write $\mathrm{Var}(\boldsymbol{Q}) = \sigma^2\left(c^* ZZ' + \kappa^* \Sigma + I\right)$ where $c^* = c/\sigma^2$ and $\kappa = \kappa^*/\sigma^2$. Thus by a reparametrization, the MLR test is invariant to σ^2 but depends on c^* and is a test of $\kappa^* = 0$ and consequently of $\kappa = 0$.

By Ara (1995), under the null hypothesis and appropriate regularity conditions (i) – (ix) from Section 4.2.3, MLR follows asymptotically a mixture of the degenerate distribution with a point mass at 0 and a χ_1^2 component with equal weights, recall Section 4.2.2.

In the following section we investigate the size and power properties of this MLR test in small samples.

4.3 Simulation study

4.3.1 Size of the MLR test

We conducted a simulation study to evaluate the performance of the MLR test. In model (4.15), we alternated the $n \times k$ design matrix X to be composed of $k = 2, 3, 4, 5$ explanatory variables, hence the notation X_1, X_2, X_3, X_4. The function

$s(t) = t$ and $\sigma^2 = 0.01$; Z is a column matrix created by a sequence of numbers equally spaced between zero and one; the constant c takes the value 1,000 and the sample size n is equal to 100 or 50. We carried out 600 simulations for each combination of the matrix X and the sample size. The size of the test for each combination was obtained by comparing (4.15) with critical value from the $\chi^2(1)$ distribution for $\alpha = 10\%$ being 2.70554 which, given the mixture nature of the asymptotic null distribution, is the critical value for a test at the 5% nominal level, refer to Sections 4.2.2 – 4.2.3. Results of the simulation are listed in *Table 4.1.* We refer to Appendix 4 A.1 for more details on *S-Plus* code on MLR test.

The following four design matrices were used to generate the data:

Data set 1, $X_1(n \times 2)$, where each value in the first column is equal to one and each value in the second column is a random normal variate with mean 5 and standard deviation 1.

Data set 2, $X_2(n \times 3)$, where the first two columns are the same as in matrix X_1 and the third column is manufacturing monthly data from dX Data File, part Main Economic Indicators, Table USA 04: Manufacturing. This data from 1960 to 2000 are presented by *Figure 4.1.* We consider the latest 50 or 100 observations for the purpose of our simulation study.

Data set 3, $X_3\,(n \times 4)$, where the first three columns are the same as in matrix X_2 and the fourth column is the USA manufacturing data lagged by one period.

Data set 4, $X_4\,(n \times 5)$, where the first four columns are the same as in matrix X_3 and the fifth column is the lagged USA manufacturing data by two periods

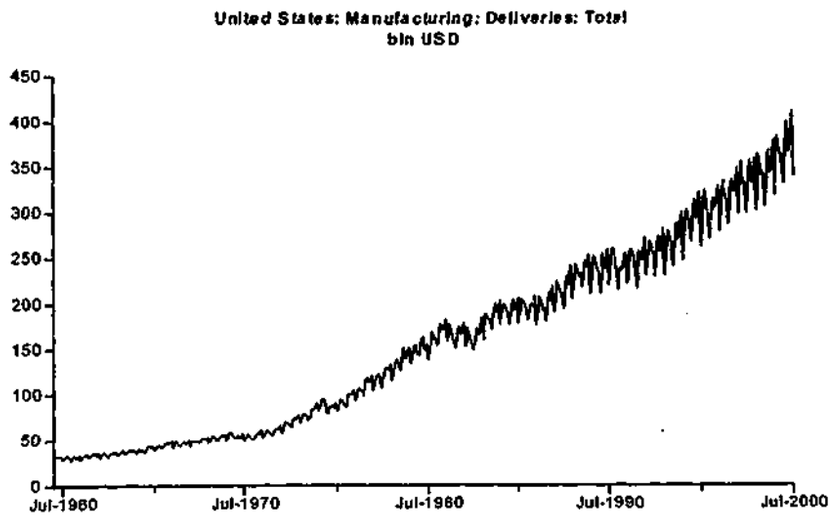

Figure 4.1: *Time plot of USA manufacturing deliveries, total in billion USD*

n	X_1	X_2	X_3	X_4
50	0.028	0.028	0.025	0.030
100	0.025	0.022	0.025	0.024

Table 4.1: *Estimated sizes of the MLR test based on the asymptotic critical values at 5% nominal level*

From the results of the simulation study on the size of the MLR test in *Table 4.1,* it is evident that the sizes are ranging between 0.022 and 0.03 (whereas the nominal size was set to 0.05). The majority of results are not significantly different from the nominal value of 0.05 at the 1% significance level. Only the value 0.022 is smaller than 0.023, which means that only in one time out of eight, the size of the test is significantly different from the nominal size. The size of the test seems to be invariant to the changes in sample size and the type of the design matrix X. Overall there is some evidence of sizes being lower than nominal but this evidence is not particularly conclusive.

We also checked whether nonzero MLR test statistics follow a $\chi^2(1)$ distribution under H_0. For the set of all calculated values of MLR > 0 we applied a collection of four plots, namely histogram, boxplot, density plot and a qq-plot. The four plot summary of the values of MLR > 0 under the null hypothesis from design matrix X_1 is depicted by *Figure 4.2.*

These graphical devices provide a quick informative guide to see whether the empirical distribution corresponds to a hypothetical theoretical distribution. Density plots are smooth versions of histograms. They provide smooth estimates of the population probability density curve. The most frequently used form to check whether a data set comes from a particular hypothesized distribution shape is a quantile-quantile plot. The qq-plot compares two set of quantiles. The first consists of the ordered set of data values, which are quantiles for the empirical distribution. The other set of quantiles are those for our hypothesized $\chi^2(1)$ distribution. If the points in this

qq-plot cluster along a straight line, the data set probably has the hypothesized $\chi^2(1)$ distribution.

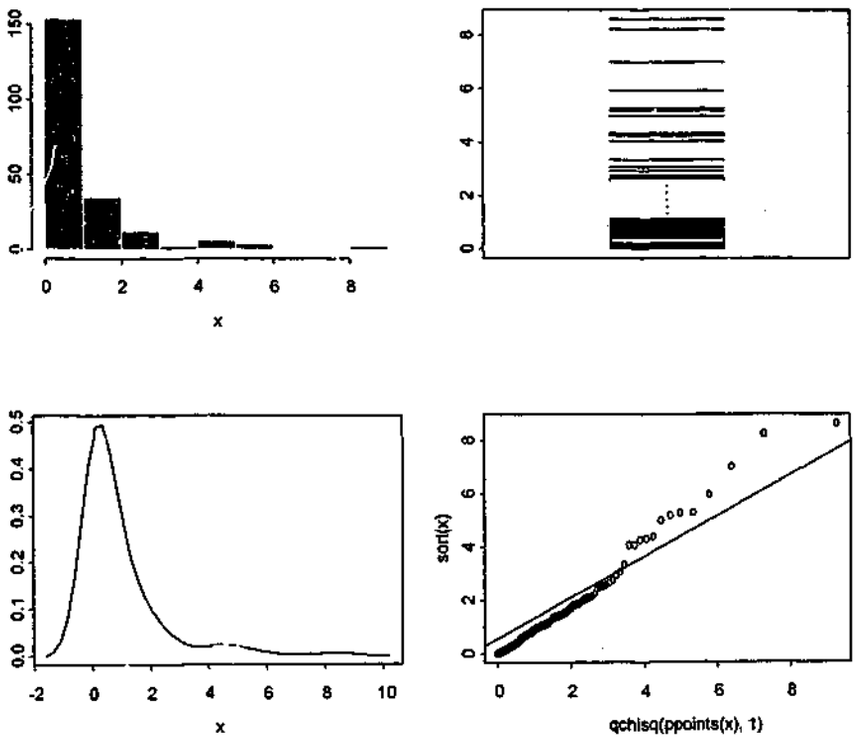

Figure 4.2: *The four plot summary of the distribution of the calculated MLR test statistics larger than zero for X_1 specification*

From *Figure 4.2* is evident that the histogram, boxplot and density plots reveal an asymmetric, skewed, like $\chi^2(1)$ distribution in shape for simulated MLR > 0. Note that the density plot of MLR > 0 should be positive. The default setting in *S-Plus* code

created some values being < 0. From the qq-plot it appears that the $\chi^2(1)$ distribution is not too far from the true distribution. There are some extreme values, but the majority of points cluster along the straight line. We obtained a similar finding for the distribution of MLR test from the simulation study on the size of the MLR test for design matrices X_2, X_3, X_4.

4.3.2 Power of the MLR test

We investigated the power of the MLR test assuming that $s(t_i)$ is one of several non-linear test functions given by (4.25). The observations were generated by

$$Y_i = \boldsymbol{x}_i'\gamma + \theta s(t_i) + e_i, \tag{4.24}$$

where the errors e_i are independent and normally distributed with variance $\sigma^2 = 0.01$. For the sample size $n = 50$, the t_i are equally spaced on the interval (0,1) with $t_1 = 0$ and $t_n = 50$, and the power is computed at the points $\theta = 0.0, 0.1, \ldots, 0.6$. We considered the following ten nonlinear functions

$$
\begin{aligned}
f_1(t) &= e^{-3e^{-3t}} \\
f_2(t) &= t^2 \\
f_3(t) &= e^{-5(1-t)} \\
f_4(t) &= \sin(4t)^2 + 3t^8 \\
f_5(t) &= \sin^{-1}(3t) \\
f_6(t) &= e^{\sin^2(3t)} \\
f_7(t) &= \sin(4\pi t) \\
f_8(t) &= t + e^{-(t-0.5)^2/0.04} \\
f_9(t) &= e^{\sin(3t)^3} + e^{\cos(3t)^2} + t^4 \\
f_{10}(t) &= e^{\sin(3t)^8} + 4t^4 \,.
\end{aligned}
\tag{4.25}
$$

These functions are shown in *Figure 4.3(a)-(e)* and were chosen to have the same range but varying degrees of nonlinearity. Here, $s(t) = f_0(t) = t$ is the linear function used for the size calculations in the preceding section. The function $f_1(t)$ is a Gompertz function, which is often used to model growth curves; $f_2(t)$ is a quadratic and a natural alternative to the assumption of linearity and $f_3(t)$ is an exponential function. Functions $f_4(t) - f_{10}(t)$ represent combination of polynomial, exponential and trigonometric functions. By this selection of test functions, we covered various trends, shapes, amplitudes and frequency of oscillation in possible non-linear relationships.

Let X be the two-column matrix initially with each value in the first column equal to one and each value in the second column a random normal variate with mean 5 and standard deviation 1. We ran our simulation study for 600 replications for different values of θ; the constant c taking the value 1,000; and the regressors being the same

as in Section 4.3.1 Results of this simulation are presented in *Table 4.2*. We stoped simulating the power of MLR test for some values of θ for some functions because the increment in the power by increasing the θ value was very small compared to other non-linear functions.

Figure 4.4 shows a power comparison of the MLR test for the ten non-linear test functions. The results show that the MLR test performs well. The lowest power is obtained for functions f_1 and f_2. The MLR test does not perform satisfactorily for the least nonlinear functions of the set of ten deterministic curves, see *Figure 4.3 (a) – (e)*. The highest power is obtained for function f_6. The MLR test also performs well also for functions f_7 and f_8. The power of the MLR test depends upon the degree of non-linearity of the test function $f(t_i)$.

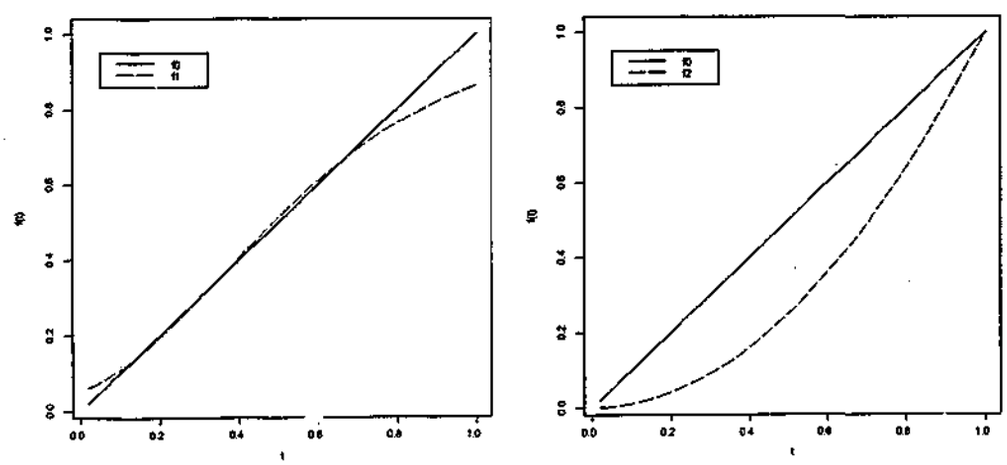

Figure 4.3(a): *Functions* $f_0(t), f_1(t), f_2(t)$, *used for power calculations*

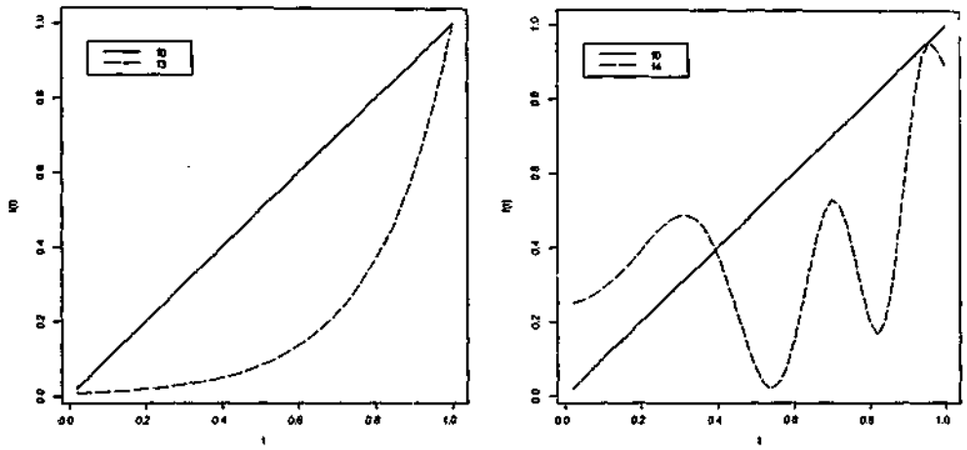

Figure 4.3(b): *Functions* $f_0(t), f_3(t), f_4(t)$, *used for power calculations*

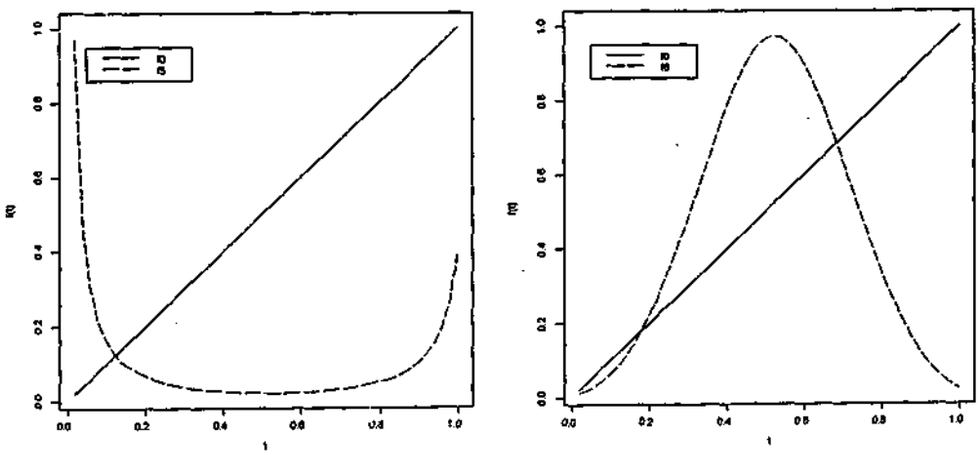

Figure 4.3(c): *Functions* $f_0(t), f_5(t), f_6(t)$, *used for power calculations*

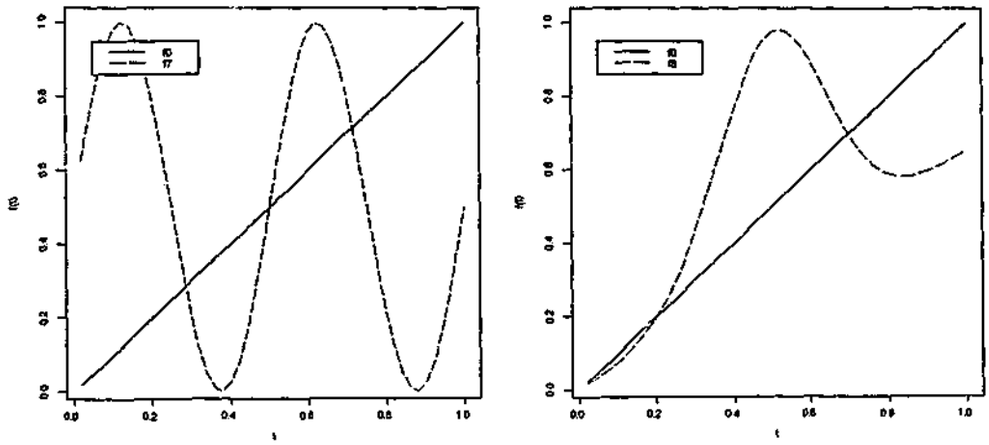

Figure 4.3(d): *Functions* $f_0(t), f_7(t), f_8(t)$, *used for power calculations*

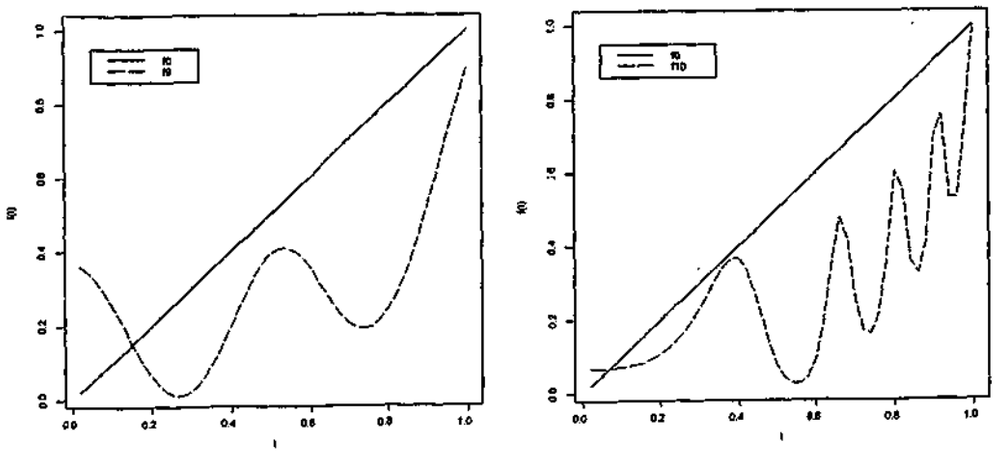

Figure 4.3(e): *Functions* $f_0(t), f_9(t), f_{10}(t)$, *used for power calculations*

Theta	Function	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10
0.00		0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
0.05		0.030	0.035	0.059	0.048	0.050	0.112	0.075	0.087	0.049	0.049
0.10		0.032	0.058	0.092	0.091	0.083	0.452	0.132	0.176	0.075	0.075
0.15		0.037	0.083	0.171	0.198	0.183	0.815	0.375	0.375	0.155	0.104
0.20		0.041	0.130	0.342	0.371	0.280	0.980	0.675	0.672	0.235	0.130
0.30		0.047	0.280	0.703	0.697	0.533	1.000	0.980	0.976	0.506	0.290
0.40		0.055	0.453	0.898	0.918	0.783		1.000	1.000	0.823	0.476
0.60					0.987	0.985				0.952	0.773
0.80					1.000	1.000				1.000	
1.00											
1.20											

Table 4.2: *Power estimates of the MLR test for ten different nonlinear functions at the 5% nominal level using asymptotic critical values*

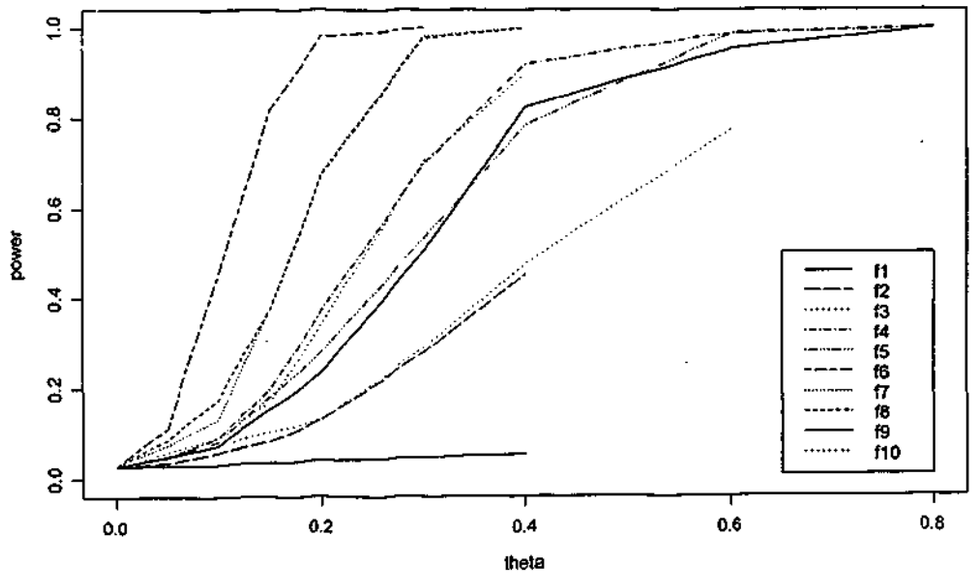

Figure 4.4: *Estimated power of the MLR test for ten non-linear test functions*

4.4 Three tests comparisons

Now we briefly describe one of the regression specification error tests (RESET) by Ramsey (1969) commonly known as the F test. We will compare the power of our MLR test with the F test and its modification the F spline test.

4.4.1 Regression specification error test – F and F spline test

The specification error considered in this thesis is that of omitted variables. The following explanation comes from Ramsey and Gilbert (1972). Suppose that the basic model relevant to the development of the test statistics is given by

$$y = X\beta + u, \tag{4.26}$$

where y is an $n \times 1$ regressand vector, X is an $n \times k$ nonstochastic matrix of rank k, β is a $k \times 1$ vector of coefficients, and u is an $n \times 1$ vector of independent disturbance terms each distributed normally with mean zero and variance σ^2. The specification of the null hypothesis is given by (4.26). The alternative hypothesis is defined by specifying that the true model has some specification other than that given in (4.26). Thus, the use of regression (4.26) to analyze data generated by some other model leads to specification error. Under the alternative hypothesis, the specification of the true model that would give rise to specification error for the case of omitted variables is given by

$$y = X\beta + w\eta + v. \tag{4.27}$$

In model (4.27), y, X, β are the same as in (4.26), w is a $n \times 1$ non-stochastic regressor vector, η is the corresponding coefficient, and v is distributed as $N(0, \sigma^2 I)$. The specification error test proposed by Ramsey (1969) was defined with respect to Theil's BLUS (best linear unbiased scalar covariance matrix predictor) residual vector $\tilde{u}$ given by $\tilde{u} = A'y$, where the $n \times (n-k)$ matrix A has the properties

$$A'A = I_{n-k}, AA' = \left[I - X\left(X'X\right)^{-1} X'\right] = M.$$

In model (4.27) the use of (4.26) as a true regression model leads to a vector $\tilde{u}$ distributed as $(n-k)$-variate normal with mean vector given by $A'\xi$, where ξ is a non-stochastic vector. The null hypothesis H_0 can be expressed as

$$H_0 : \tilde{u} \sim N\left(0, \sigma^2 I_{n-k}\right)$$

and alternative hypothesis H_1 as

$$H_1 : \tilde{u} \sim N\left(A'\xi, \bar{\sigma}^2 I_{n-k}\right),$$

where $\bar{\sigma}^2$ is the variance of $\tilde{u}_i$. The fundamental idea underlying the test is that the effect of specification error is to alter the distribution of the residuals from that

postulated under the null hypothesis. In the omitted variable problem, the vector ξ is given by ηw, the vector w can be approximated by a polynomial. Ramsey (1969) also found that, in practice the polynomial of degree $k = 3$ is large enough to obtain reasonable power against the alternative hypothesis.

We want to apply the Ramsey error misspecification test to our model. Recall (4.15), the specification of the null hypothesis is given by the restriction

$$Q = e \tag{4.28}$$

where e is a vector of independent disturbance terms each distributed normally with mean zero and variance σ^2. This leads to the restricted model

$$Y = X\gamma + e. \tag{4.29}$$

The alternative hypothesis is defined by specifying that the true model has some specification other than that given by (4.29). Thus, the use of the restricted regression to analyze data generated by some other model leads to specification error. We want to test the non-linear component in Q, therefore the specification error considered here is that of omitted variable in (4.15). Under the alternative hypothesis, the specification of the true model is given by (3.13).

For a test of significance level α, the null hypothesis is rejected if the test statistic

$$T_F > F_{m,n-m-k,\alpha} \tag{4.30}$$

in which

$$T_F = \frac{\left(SSE_R - SSE_U\right)/m}{SSE_U/(n-m-k)}, \tag{4.31}$$

where SSE_R stands for error sum of squares resulting from restricted model (4.29) and SSE_U represents the error sum of squares resulting from unrestricted model (3.13). The number of parameters being estimated is denoted by k and the symbol m is used for the number of restrictions. The test statistic T_F follows the F-distribution under H_0. In (3.13), function $s(t_i)$ can be estimated by any nonlinear function. Our arbitrary choice was a cubic polynomial function. We denote this type of the regression specification error test as the F test.

We use a modification to the above mentioned F test and denote this modified test as the F *spline* test. If in (3.13), function $s(t_i)$ is estimated non-parametrically by a smooth spline function instead of the cubic polynomial function, we assume that the true model in this case will be more accurate, because of the flexibility of the smoothing splines. We denote the test statistic as $T_{Fspline}$ for this alternation. For more details on *S-Plus* codes on F and F *spline* see Appendixes 4A.2 and 4A.3.

We assume that the F *spline* test works exactly like the F test. The unrestricted model is specified by (3.13) and restricted model by (4.29) as for the F test. The only

difference in the F and F *spline* tests is that the function $g(t_i)$ in (3.13) is approximated by the smoothing spline instead of by the polynomial of degree $k = 3$.

The following brief discussion is about how to effectively use the smoothing spline feature from non-parametric regression in *S-Plus*. To get the test statistic $T_{Fspline}$ for the F *spline* test, we fit a generalized additive model to the data as the unrestricted model, using the function "gam" in *S-Plus*. Then with analysis of variance on "gam" we get the required test statistic $T_{Fspline}$. The number of restrictions m is estimated from the smoothing matrix (see Hastie and Tibshirani, 1990). Using "anova.gam" in *S-Plus* gives a breakdown of the degrees of freedom for all the terms in the model into a single degree of freedom for its linear component, and the reminder for non-parametric component. In addition, a type of F test is performed for each of the non-parametric terms in *S-Plus*

4.4.2 Simulation comparison

We carried out 600 replications with $n = 50$. Let matrix X be the two-column matrix with each value in the first column equal to one and each value in the second column a random normal variate with mean 5 and standard deviation 1. Results are listed in *Table 4.3*. Power of the MLR, F and F spline tests differ according to different deterministic nonlinear functions being used in the study. All three tests are most powerful for function f_6. Our MLR test performed well for all test functions besides function f_1. When comparing the MLR test with the F test, the former

outperformed the latter for functions f_5, f_6, f_8 and both are very close to each other for functions f_5, f_6. When comparing the MLR test with the F spline test, the former outperformed the latter for functions f_2, f_3, f_6. The power of both F and F spline tests is close to each other for all functions. In *Table 4.3* the best power is highlighted for each test and nonlinear test function with respect to θ.

The simulation study comparing the MLR, F and F spline tests revealed an interesting finding that the F test appeared to perform superior to the MLR and F spline tests. The F test had the best power for functions $f_1, f_2, f_3, f_4, f_8, f_9, f_{10}$. In fact, for function f_7 when the F test performed poorly in comparison to MLR, the F spline test outperformed MLR.

Thus the F test or its modification the F spline test are better alternatives to the MLR test. Note that the size of the MLR test is the smallest of all three tests therefore the power of the MLR test might be instantly behind by this initial handicap.

Figures 4.5 (a) – (j) plot the power of the three tests for each individual deterministic test function. These plots give better graphical comparison of all three tests for every nonlinear test function.

	Theta	0.00	0.05	0.10	0.15	0.20	0.30	0.40	0.60	0.80	1.00
Function	Test										
f1	MLR	0.028	0.030	0.032	0.037	0.041	0.047	0.055			
	F	0.055	**0.100**	**0.301**	**0.593**	**0.851**	**0.991**	**1.000**			
	F spline	0.056	0.052	0.052	0.053	0.065	0.070	0.095	0.151		
f2	MLR	0.028	0.035	0.058	0.083	0.130	0.280	0.453			
	F	0.055	**0.115**	**0.393**	**0.715**	**0.938**	**0.990**	**1.000**			
	F spline	0.056	0.060	0.071	0.095	0.125	0.255	0.385	0.713		
f3	MLR	0.028	0.059	0.092	0.171	0.342	0.703	0.898			
	F	0.055	**0.100**	**0.288**	**0.593**	**0.840**	**0.990**	**1.000**			
	F spline	0.056	0.071	0.085	0.167	0.336	0.653	0.860	1.000		
f4	MLR	0.028	0.048	0.091	0.198	0.371	0.697	0.918	0.987	1.000	
	F	0.055	**0.076**	**0.156**	**0.271**	**0.458**	**0.838**	**0.977**	**1.000**		
	F spline	0.056	0.076	0.132	0.238	0.415	0.795	0.971	1.000		
f5	MLR	0.028	0.050	0.083	**0.183**	0.280	0.533	0.783	0.985	1.000	
	F	0.055	0.060	0.098	0.170	0.253	0.475	0.720	0.976	1.000	
	F spline	0.056	**0.065**	**0.090**	0.180	**0.302**	**0.593**	**0.850**	**0.990**	**1.000**	
f6	MLR	0.028	0.112	0.452	**0.815**	**0.980**	**1.000**				
	F	0.055	0.116	0.378	0.760	0.935	1.000				
	F spline	0.056	**0.135**	**0.458**	0.800	0.978	1.000				
f7	MLR	0.028	0.075	0.132	0.375	0.675	0.980	1.000			
	F	0.055	0.055	0.105	0.176	0.303	0.528	0.713			
	F spline	0.056	**0.090**	**0.253**	**0.516**	**0.870**	**0.990**	**1.000**			
f8	MLR	0.028	0.087	0.176	0.375	0.672	0.976	1.000			
	F	0.055	**0.120**	**0.312**	**0.570**	**0.850**	**0.990**	**1.000**			
	F spline	0.056	0.091	0.218	0.450	0.712	0.905	0.987	1.000		
f9	MLR	0.028	0.049	0.075	0.155	0.235	0.506	0.823	0.952	1.000	
	F	0.055	**0.080**	**0.113**	**0.230**	**0.393**	**0.712**	**0.910**	**0.981**	**1.000**	
	F spline	0.056	0.070	0.108	0.200	0.317	0.705	0.933	0.975	1.000	
f10	MLR	0.028	0.049	0.075	0.104	0.130	0.290	0.476	0.773		
	F	0.055	**0.090**	**0.162**	**0.300**	**0.500**	**0.780**	**0.975**	**1.000**		
	F spline	0.056	0.060	0.086	0.120	0.165	0.350	0.538	0.880		

Table 4.3: *Simulated power of the MLR, F and F spline tests for ten non-linear test functions*

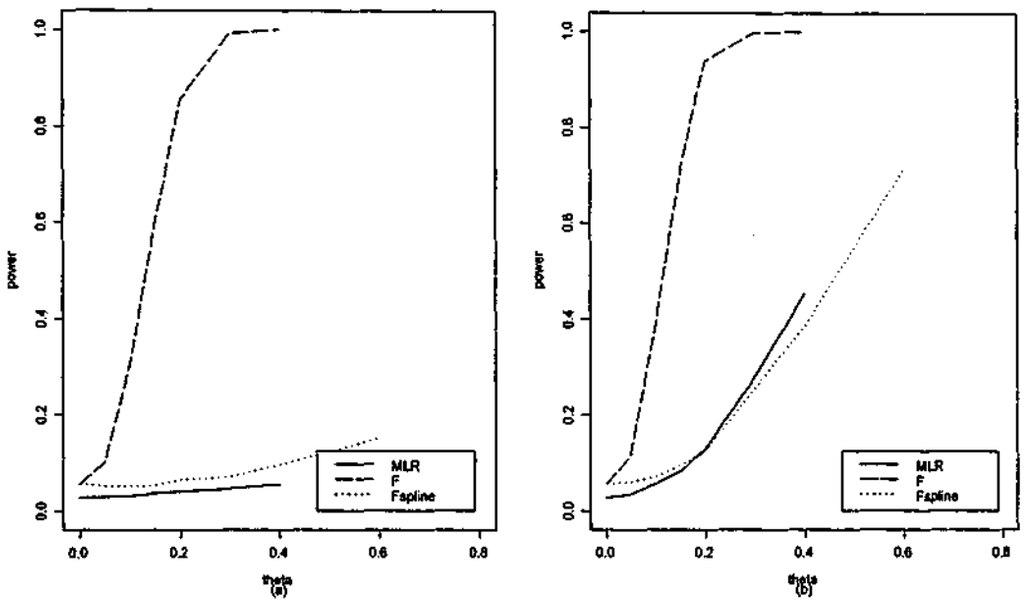

Figure 4.5 (a) – (b): *Power of the MLR, F and F spline tests for the function f_1 part (a) and f_2 part (b)*

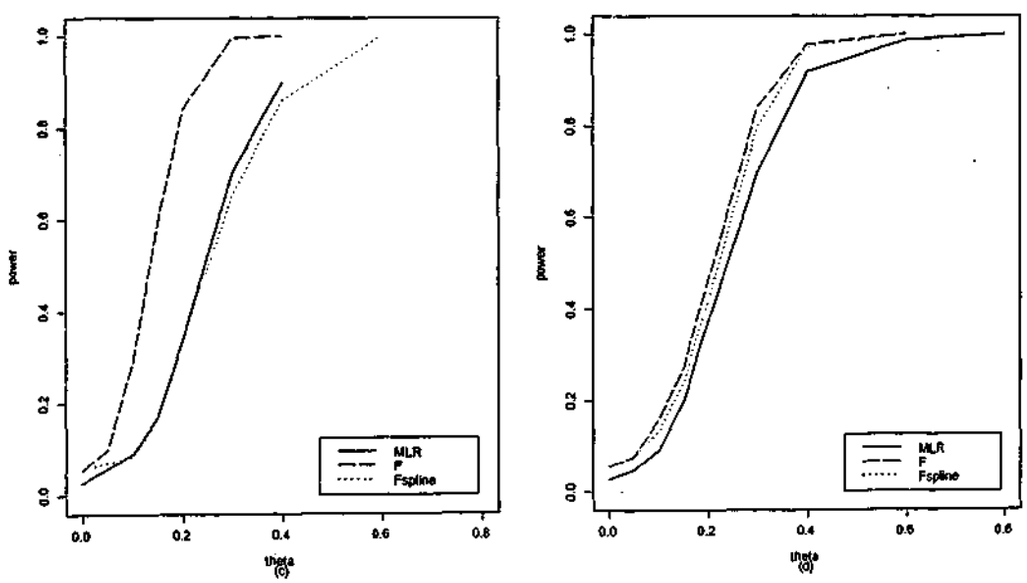

Figure 4.5 (c) – (d): *Power of the MLR, F and F spline tests for the function f_3 part (c) and f_4 part (d)*

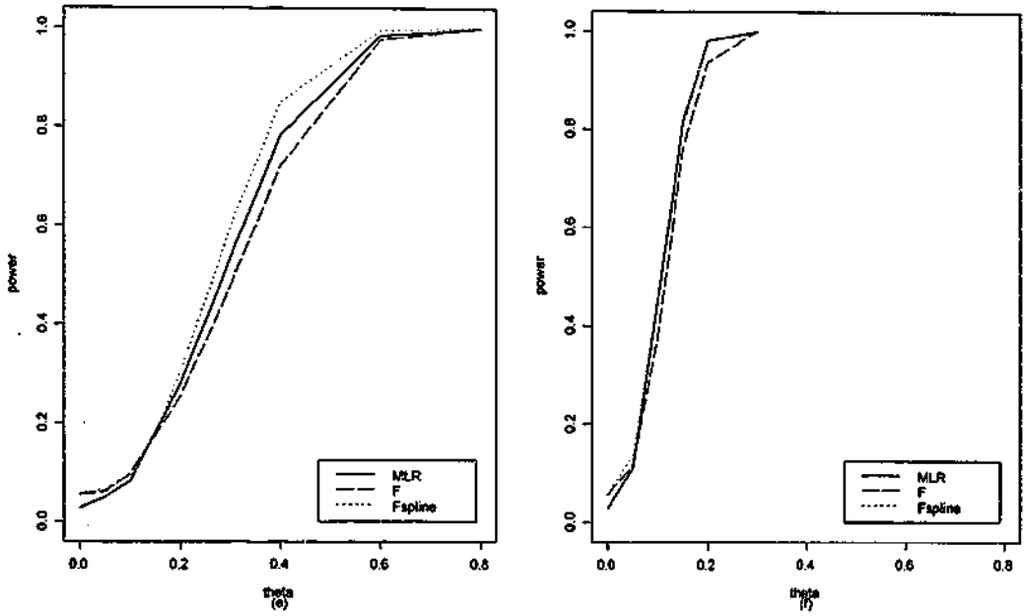

Figure 4.8 (e) – (f): *Power of the MLR, F and F spline tests for the function f_5 part (e) and f_6 part (f)*

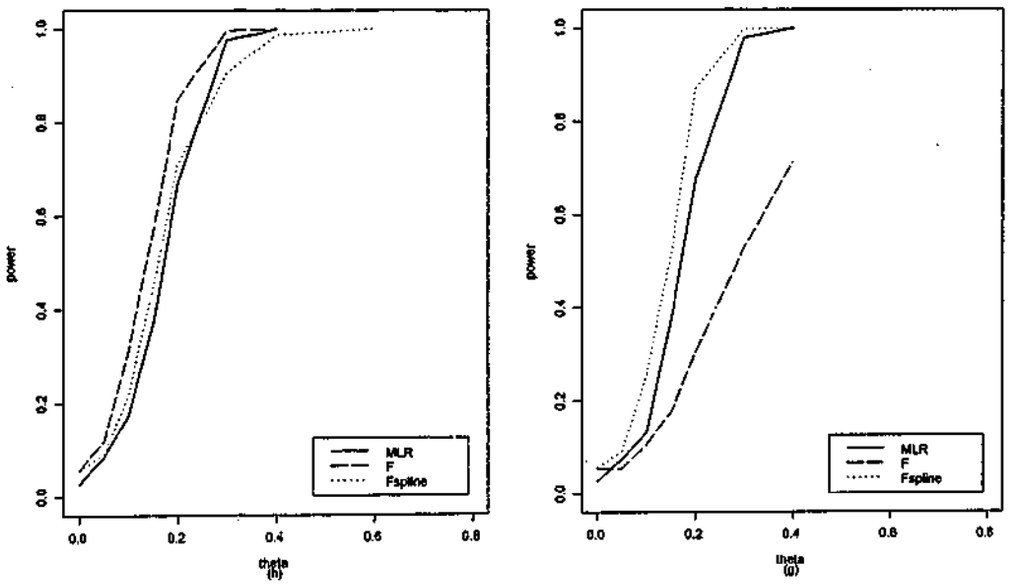

Figure 4.5 (g) – (h): *Power of the MLR, F and F spline tests for the function f_7 part (g) and f_8 part (h)*

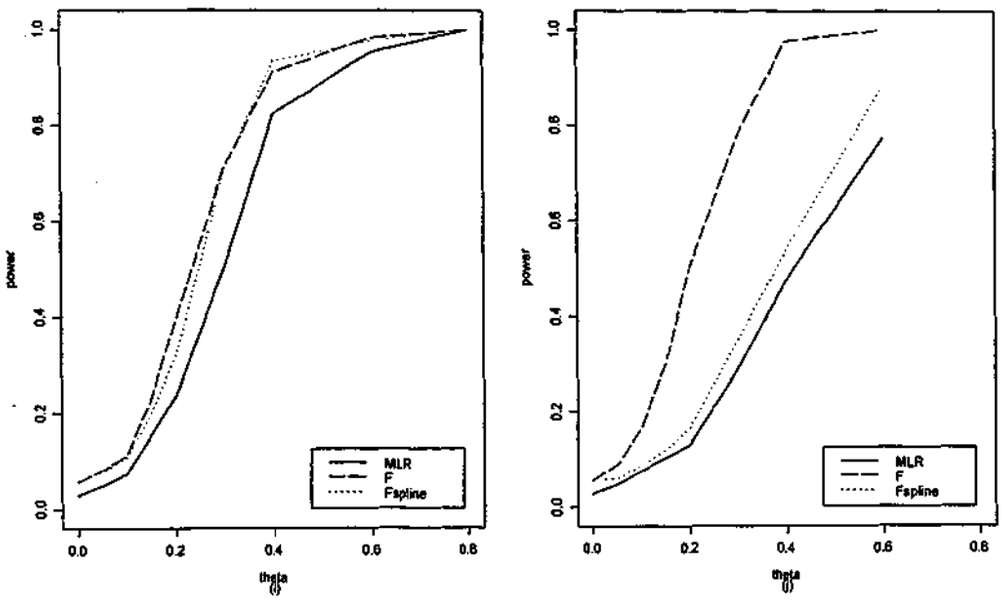

Figure 4.5 (i) – (j): *Power of the MLR, F and F spline tests for the function f_9 part (i) and f_{10} part (j)*

Plots *(d), (e), (f), (h)* and *(i)* related to functions f_4, f_5, f_6, f_7 and f_9 show very small deviations of the MLR, F and F spline tests results. Plots *(c), (g),* and *(j)* related to functions f_3, f_8 and f_{10} show moderate deviations of all three tests results. Plots *(a), (b)* show the largest deviations mostly on MLR and F spline test results for functions f_1, f_2. These results are most likely due to a different degree of nonlinearity of the ten deterministic functions being used in the study. Most importantly the MLR test failed for the function, which shows the least nonlinear shape, i.e. function f_1 or Gompertz curve. The F test was least powerful for function f_8, which is possibly the most nonlinear.

The F spline test is a mean based test and the MLR test is a variance based test. It is interesting that similar power curves resulted for both tests from the simulation study mostly for functions $f_4, f_5, f_6, f_7, f_8, f_9$.

4.5 The MLR test application

In this section we apply the MLR test to *the nominal yields to maturity data* (hereafter yields data) and *the holding period return data* (hereafter hold data) from the Fama Twelve-Month Treasury Bill Term Structure File of the U.S. Government Securities File of the Centre for Research in Securities Prices (CRSP) at the University of Chicago. The file contains twelve yields series on Treasury bills; one series for bills with one month to maturity, another for bills with two months to maturity, and so on, to a series with twelve months to maturity.

The MLR test was conducted on the series for bills with 3 months to maturity and with 6 months to maturity for both yields and hold data. The underlying model, refer to (3.13), is assumed to be a semiparametric model

$$Y_i = x_i'\gamma + s(t_i) + e_i,$$

where $\mathbf{Y}$ is $(n \times 1)$ vector of Y_i which are 6 months hold (yields) to maturity, $\mathbf{X}$ is a $(n \times 2)$ matrix of constant intercept term and the 3 months hold (yields) to maturity x_i, e_i are errors that are iid $N(0, \sigma_e^2)$ and $n = 120$. Using a common notation from

the term structure theory, for example as used by Hall et al. (1992), this semiparametric model can be rewritten in the form

$$R(6,t_i) = \gamma R(3,t_i) + L(6,t_i) + e_i,$$

where $L(6,t_i) \equiv s(t_i)$ is a risk premia, which may account for risk considerations or investors' preferences about liquidity, $R(6,t_i) \equiv Y_i$ and $R(3,t_i) \equiv x_i'$. According to some expectation theories, risk premia is assumed to be zero, while other versions assume that it is constant over time. For more about the theory of the term structure of interest rates, see Hall, Anderson and Granger (1992). The risk premia is likely to be time varying in a linear or non-linear fashion, because of uncertainty caused by the volatility in monetary growth, interest rates and other economic environmental explanatory variables, which are believed to be non-linear in time. The MLR test is a test for the inclusion of a possibly non-linear component, in this case a risk premia $L(6,t_i) \equiv s(t_i)$ term in the model without testing its specific form.

Figure 4.6 shows a scatter matrix of the Holding period return data, where the notation 'tmths' denotes 3 months hold to maturity and 'smths' denotes 6 months hold to maturity. From Figure 4.6, a stochastic linear relationship of the 6 months hold to maturity data to the 3 months hold maturity data is visible. Dependence of both data sets on time is however not constant and shows a visible departure from linearity. It does make a sense to test whether to include a risk premia $L(6,t_i) \equiv s(t_i)$ component into the model in a form different from a constant. Moreover Figure 4.7 illustrates the time plots of the 6 months hold to maturity data and the 3 months hold maturity data.

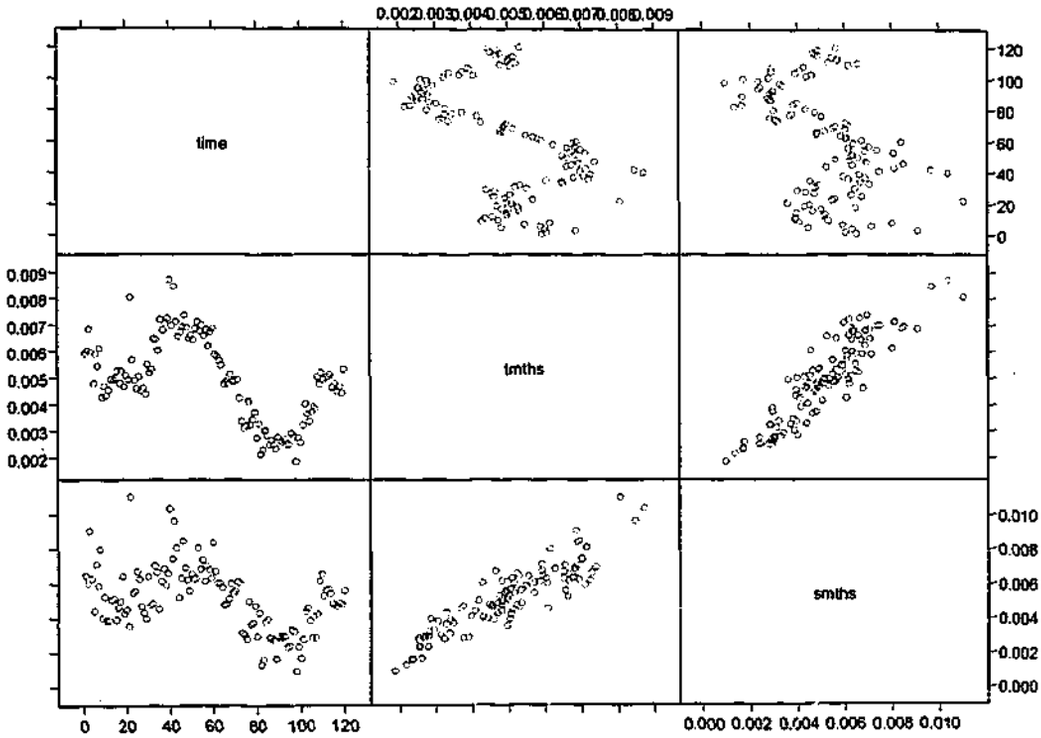

Figure 4.6: *Scatter matrix of the Holding period return data*

It is evident that holds of different maturity appear to move together over time, which is typical for cointegrated time series, for more see Granger (1981).

The MLR test has $H_0 : \kappa = 0$, which means that the risk premia is constant over time, therefore its value will be estimated by the parameter vector γ. The specific alternative $H_A : \kappa > 0$ means that the risk premia is time varying and therefore it should stay in the model. Result of the MLR test for hold data is $MLR = 23.29$ is

supported by the *p-value* being 0.00. We have significant evidence to reject $H_0 : \kappa = 0$ at 5% level when we compare MLR test statistic with $\chi^2(1) = 2.70554$. The MLR test suggests inclusion of the time varying risk premia component into the model for hold maturity data.

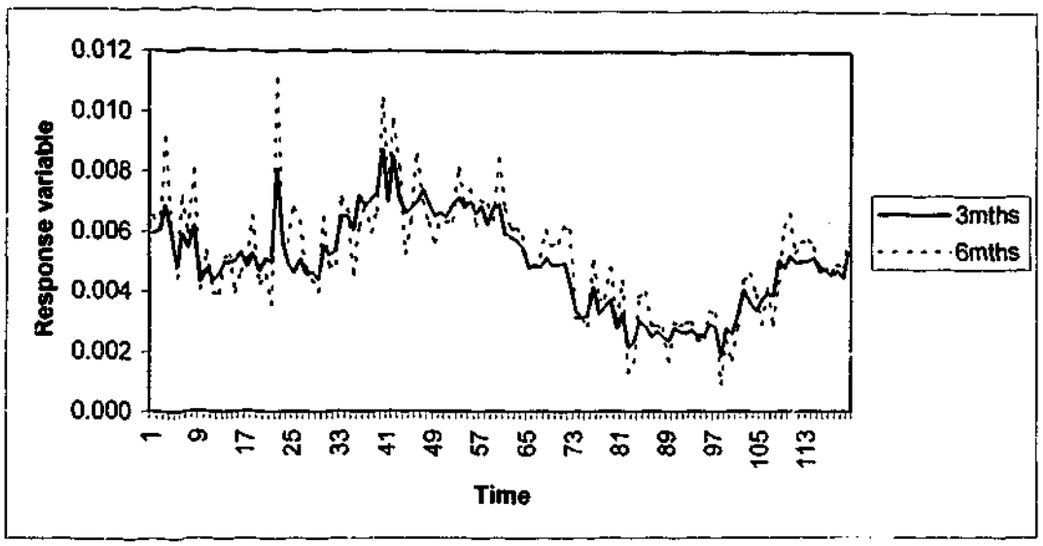

Figure 4.7: *Holds to maturity data from 31/12/85 to 29/12/95*

The logical further step is to estimate the semiparametric model, to clarify whether the estimated model is statistically significant and to show its potential applicability in the forecasting yields of Treasury bills.

We also applied the MLR test to yields maturity return for 3 and 6 month data from the same source. The MLR test confirmed the previous finding and rejected H_0 at the

5% nominal significance level. The MLR test suggests including the time varying risk premia component into the model for yields maturity data.

This analysis demonstrates the use of MLR test and suggests gains from its use in applied research. It is appropriate to think about the form of the time varying risk premium, which will be tested by MLR* test to be discussed in Chapter 5. At this stage we can say that it is advisable to include the time varying risk premium as an exogenous variable in the model.

4.6 Results summary

In this chapter we proposed the marginal likelihood ratio test for the inclusion of a possibly nonlinear component in a linear regression model. Our simulation study on the MLR test revealed a slightly smaller than the nominal size of the test. The size of the MLR test was invariant to the change in the design matrix. The study on the power of the MLR test using ten nonlinear deterministic functions was promising. The simulation study comparing our MLR test with the F and F spline tests failed to confirm superiority of the proposed MLR test. Either the F test or F spline test outperformed the MLR test in power for all ten nonlinear deterministic test functions. Results from this chapter support a mean based F or F spline tests instead of a variance based MLR test when testing for the inclusion of a possibly nonlinear component. The mean based F spline test seems to be worthy of further exploration. It had the best power for function f_8 and outperformed the F test for functions f_5, f_6.

We demonstrated the application of the MLR test on the nominal yields to maturity data and the holding period return data from the Fama Twelve-Month Treasury Bill Term Structure File of the U.S. Government Securities.

Appendix 4A.1 *S-Plus* code of MLR test

```
################################
# MLR test, size and power simulations#
################################

### Generate x data
n <- 50
set.seed(1)
x <- rnorm(n,5,1)
tt <- (1:n)/n

## Set up equation for y data
gamma1 <- 0 #intercept
gamma2 <- 1 #slope
sigma <- 0.1 #stdev

### Set up X and S matrices
ones <- seq(1,1,l=n)
X <- cbind(ones,x)
S <- cbind(ones,tt)
halfN <- 0.5*(n-ncol(X))

### Set up Sigma matrix
h3 <-1/(n^3)
Sigma <- matrix(0,nrow=n,ncol=n)
for (i in 1:n)
        Sigma[i,i:n] <- Sigma[i:n,i] <- (i*i*(i+1.5*(0:(n-i))))/3 *h3

### Set up nonlinear functions
f0 <- tt
f1 <- exp(-3*exp(-3*tt))
f2 <- tt^2
f3 <- exp(-5*(1-tt))
f4 <- 0.71*(sin((4*tt)^2)+3*tt^8)/3+.25
f5 <- (0.909*(1/sin(3*tt))/15)-0.04
f6 <- 1.4*(exp(sin(3*tt)^2)/2.5)-0.55
f7 <-0.5*sin(4*pi*tt)+0.5
f8 <-((5*tt+5*exp(-((tt-0.5)^2)/0.04))/7)*0.909
f9 <- 2.5*((exp(sin(3*tt)^3)+exp(cos(3*tt)^2)+tt^4)/4.5)-1.7
f10 <-((exp(sin((3*tt)^3))+4*tt^4)/6)-0.1

### Functions to compute likelihood and MLR statistic (here lr)
logdet <- function(x)
        sum(log(eigen(x)$values))

fl<-function(ll)
{
        C <- chol(SS + exp(ll)*Sigma)
        Cinv <- t(solve(C))
        Xstar <- Cinv %*% X
        Ystar <- Cinv %*% Y
        tX <- t(Xstar)
        XX <- tX %*% Xstar
        beta <- solve(XX) %*% tX %*% Ystar
        res <- Ystar - Xstar %*% beta
        return(logdet(C) + 0.5*logdet(XX) + halfN*log(sum(res*res)))
}

minfl <- function()
```

```
{
        cat(".")
        y <-  fmean + rnorm(n,0,sigma)
        Y <- matrix(y,ncol=1)
        assign("Y",Y,frame=1)
        fl0 <- fl(-1e20) # Value at zero
        fredall <- nlmin(fl,1) # Minimize starting near zero
        fl.fred <- fl(fredall$x) # Best fl
        if(fl0 < fl.fred)
                lstar <- lr <- 0
        else
        {
                lr <- 2*(fl0 - fl.fred)
                lstar <- exp(fredall$x)
        }
        return(c(lr,lstar))
}

## Simulation
theta <- 0.6
fmean <- gamma1 + gamma2*x + theta*f10
cc <- 1000
SS <- cc*(S %*% t(S)) + diag(n)
#set.seed(2)
nn <- 600
lmat <- matrix(0,nrow=nn,ncol=2)
for (k in 1:10)     lmat[k,] <- minfl()
.
.
.
for (k in 591:600)lmat[k,] <- minfl()
lr <- lmat[,1]
lstar <- lmat[,2]
count2 <- length(lr[lr>2.70554])
print(count2)
power<-count2/nn
print(power)
```

Appendix 4A.2 *S-Plus* code of F test

```
######################
# Ftest, power  simulation#
######################

### Generate x data
n <- 50
set.seed(1)
x <- rnorm(n,5,1)
tt <- (1:n)/n

### Set up equation for y data
gamma1 <- 0 #intercept
gamma2 <- 1 #slope
sigma <- 0.1 #stdev

### Set up nonlinear functions
f0 <- tt
f1 <- exp(-3*exp(-3*tt))
f2 <- tt^2
f3 <- exp(-5*(1-tt))
f4 <- 0.71*(sin((4*tt)^2)+3*tt^8)/3+.25
f5 <- (0.909*(1/sin(3*tt))/15)-0.04
f6 <- 1.4*(exp(sin(3*tt)^2)/2.5)-0.55
f7 <-0.5*sin(4*pi*tt)+0.5
f8 <-((5*tt+5*exp(-((tt-0.5)^2)/0.04))/7)*0.909
f9 <- 2.5*((exp(sin(3*tt)^3)+exp(cos(3*tt)^2)+tt^4)/4.5)-1.7
f10 <-((exp(sin((3*tt)^3))+4*tt^4)/6)-0.1

### Calculate F statistic
Ftestsim<-function()
{
        y<-fmean+rnorm(n,0,sigma)
        ma<-lm(y ~ x +tt+tt^2+tt^3)
        m0<-lm(y ~ x )
        Ftest<-((deviance(m0)-deviance(ma))/3)/(deviance(ma)/43)
        return(Ftest)
}

#######################################
## Simulation to compute power ##
#############################

theta <-0.3
fmean <- gamma1 + gamma2*x + theta*f8
set.seed(23)
nsim <- 600
Ftest <- numeric(nsim)
for(i in 1:nsim)
        Ftest[i] <- Ftestsim()

count <- length(Ftest[Ftest>2.83])
print(count)
power <- count/nsim
print(power)

#######################
## Calculate critical value ##
#######################
```

```
fmean <- gamma1 + gamma2*x
set.seed(2)
nsim <- 600
Ftest <- numeric(nsim)
for(i in 1:nsim)
        Ftest[i] <- Ftestsim()
Ftest.critical.value <- quantile(Ftest,0.95)
print(Ftest.critical.value)
```

Appendix 4A.3 *S-Plus* of F spline test

```
##########################################################
# Ftest assuming that the additional nonlinear component is estimated#
# by smoothing spline, power simulation                                #
##########################################################

### Generate x data
n <- 50
set.seed(1)
x <- rnorm(n,5,1)
tt <- (1:n)/n

### Set up equation for y data
gamma1 <- 0 #intercept
gamma2 <- 1 #slope
sigma <- 0.1 #stdev

### Set up nonlinear functions
f0 <- tt
f1 <- exp(-3*exp(-3*tt))
f2 <- tt^2
f3 <- exp(-5*(1-tt))
f4 <- 0.71*(sin((4*tt)^2)+3*tt^8)/3+.25
f5 <- (0.909*(1/sin(3*tt))/15)-0.04
f6 <- 1.4*(exp(sin(3*tt)^2)/2.5)-0.55
f7 <-0.5*sin(4*pi*tt)+0.5
f8 <-((5*tt+5*exp(-((tt-0.5)^2)/0.04))/7)*0.909
f9 <- 2.5*((exp(sin(3*tt)^3)+exp(cos(3*tt)^2)+tt^4)/4.5)-1.7
f10 <-((exp(sin((3*tt)^3))+4*tt^4)/6)-0.1

theta <-0.2
fmean <- gamma1 + gamma2*x + theta*f8

nsim<-600
Fspl<-numeric(nsim)
        for (i in 1:nsim)
        {

                y<-fmean+rnorm(n,0,sigma)
                ma<-gam(y ~ x + s(tt))
                junk<-anova(ma,test="F")
                Fspl[i]<-(junk$"Npar F"[3])

        }

count<-length(Fspl[Fspl>3.105614])
power<-count/nsim
print(power)

# Fcritical by bootstrap from 1000 obs is 3.105614 #
# F critical 4,40 d.f. alpha 0.05 is 2.61 #
```

CHAPTER 5

Testing for linearity of a possibly nonlinear component

5.1. Introduction

This chapter is an extension of Chapter 4. Recall that in (3.1) we have a semiparametric regression model with one potential nonlinear component. We apply the same finding as in Section 4.1 and propose a likelihood ratio one-sided test based on the marginal likelihood function as was derived by Tunnicliffe Wilson (1989) and Ara (1995). We use the error variance-covariance matrix (3.21) for this model in the case of equally spaced data. We will modify the null hypothesis when testing for linearity of a possibly nonlinear component, hence the MLR* notation.

We need to distinguish between the MLR test from Chapter 4 and the MLR* test to be proposed in this chapter. The MLR test is a test of the existence of a nonlinear component whereas the MLR* test is a test of linearity of a possibly nonlinear component. Recall Section 4.6, where we illustrated the use of the MLR test on holds and yields maturity data. The MLR test suggested including a time varying risk premia component into the model. In this chapter we test by the MLR* test, whether the time varying risk premia component should be included in a linear form or a nonlinear form. Hence the MLR* is a test of linearity of a possibly nonlinear component.

We propose the marginal likelihood based test (MLR*) for testing for linearity of a possibly nonlinear component in Section 5.2 by applying the error variance-covariance matrix for this model in the case of equally spaced data, which has been derived in Chapter 3. We investigate the performance of the test via a Monte Carlo experiment in Section 5.3. We compare the MLR* test with the regression specification error test by Ramsey (1969), known as F test, and with the locally most powerful invariant test by King and Hillier (1985) in Section 5.4. The comparison between the mean based tests and the variance based tests is one of the themes of this section. We apply the MLR* test to real economic data in Section 5.5. Section 5.6 summarises our findings.

5.2 Test for linearity of a possibly nonlinear component – MLR* test

We have a semi-parametric regression model with one potentially non-linear component. Recalling (3.14), we can rewrite the model as follows

$$\boldsymbol{Y} = \boldsymbol{X}\gamma + \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{Q}_*, \tag{5.1}$$

or

$$\boldsymbol{Y} = \boldsymbol{G}\boldsymbol{\delta} + \boldsymbol{Q}_*,$$

where $\boldsymbol{Q}_* = \boldsymbol{g} + \boldsymbol{e}$, matrix $\boldsymbol{G} = [\boldsymbol{X} : \boldsymbol{Z}]$ involves matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$, so $\boldsymbol{\delta}' = (\gamma', \boldsymbol{\beta}')$, $\boldsymbol{e}_i$ are iid $\mathrm{N}(0, \boldsymbol{\sigma}^2)$ and the vector $\boldsymbol{g}$ is the same as in Section 4.2.4. Now $\mathrm{E}(\boldsymbol{Q}_*) = \boldsymbol{0}$, the variance is $\Lambda_* = \mathrm{Var}(\boldsymbol{Q}_*) = \kappa\Sigma + \sigma^2 I$ and $\kappa = \lambda^{-1}$. We can substitute the error covariance form Σ given by (3.21) into this expression in order to test the linearity of component f, referring to (3.13). We use the null hypothesis

$$H_0 : \kappa = 0 \quad \text{i.e.,} \boldsymbol{Q}_* \sim \mathrm{N}(0, \sigma^2 I)$$

against the specific alternative

$$H_a : \kappa > 0 \quad \text{i.e.,} \boldsymbol{Q}_* \sim N(0, \kappa\Sigma + \sigma^2 I).$$

We therefore want to test the model $Y_i = x_i'\gamma + s(t_i) + e_i$ being restricted under H_0 with $s(t_i) = \beta_0 + \beta_1 t_i$ versus an unrestricted model under H_a with $s(t_i) = z_i'\beta + g(t_i)$ being non-linear. Therefore under H_0 if $\kappa = 0$, $g(t_i) \equiv 0$ and $s(t)$ is a linear function of t.

We will construct the test statistic in the same way as in Section 4.2. By Tunnicliff Wilson (1989) and Ara (1995), the marginal likelihood function (4.12) - (4.13) is modified for our testing problem as follows

$$l(\kappa|y)_* = |\Lambda_*|^{-1/2} |G'\Lambda_*^{-1}G|^{-1/2} \left(\hat{Q}_*'\Lambda_*^{-1}\hat{Q}_*\right)^{-N/2}, \tag{5.2}$$

where $\hat{Q}_*$ is the GLS residual vector from (5.1) assuming variance-covariance matrix Λ_*. Equation (5.2) can be simplified into the form

$$l(\kappa|y)_* = |C_*|^{-1} |G_*'G_*|^{-1/2} \left(\sum_{i=1}^{n} \hat{\omega}_i^2\right)^{-N/2}, \tag{5.3}$$

so that the marginal log-likelihood is given by

$$\log l(\kappa|y)_* = -\log|C_*| - \frac{1}{2}\log|G_*'G_*| - \frac{N}{2}\log\left(\sum_{i=1}^{n} \hat{\omega}_i^2\right), \tag{5.4}$$

where C_* is the lower-triangular matrix from the Choleski decomposition of Λ_* and $\left(\sum_{i=1}^{n}\hat{\omega}_i^2\right)$ denotes the sum of squared residuals from the ordinary least squares regression of $\boldsymbol{Y}_*$ on $\boldsymbol{G}_*$ where $\boldsymbol{G}_* = C_*^{-1}\boldsymbol{G}$ and $\boldsymbol{Y}_* = C_*^{-1}\boldsymbol{Y}$. We estimate κ by minimising (5.4). Then the marginal log-likelihood ratio MLR* test statistic follows

$$\mathrm{MLR}^* = 2\left[\log l\left(\hat{\kappa}|\boldsymbol{y}\right)_* - \log l\left(0|\boldsymbol{y}\right)_*\right]. \tag{5.5}$$

Note that MLR* is independent of σ under H_0. To see this, recall that $\mathrm{E}\left(\boldsymbol{Q}_*\right) = 0$ and we can write $\mathrm{Var}\left(\boldsymbol{Q}_*\right) = \sigma^2\left(\kappa^*\Sigma + I\right)$ where $\kappa = \kappa^*/\sigma^2$. Thus by a reparametrization, the MLR test is invariant to σ^2 and is a test of $\kappa^* = 0$ and consequently of $\kappa = 0$.

By Ara (1995), under the null hypothesis and some appropriate regularity conditions, namely (i) – (ix) from Section 4.2.3, MLR* follows asymptotically a mixture of the degenerate distribution with a point mass at 0 and a χ_1^2 component with equal weights, recall Section 4.2.2.

In the following section, we investigate the size and power properties of this MLR* test in small samples.

5.3 Simulation study

5.3.1 Size of the MLR* test

We conducted a simulation study to evaluate the performance of the marginal likelihood based MLR* test. In model (5.1), we alternated the $n \times k$ design matrix $\boldsymbol{G} = [\boldsymbol{X} : \boldsymbol{Z}]$ to be composed of $k = 3,4,5,6$ explanatory variables, hence the respective notation G_1, G_2, G_3, G_4. The sample size is equal to 100 or 50. The size of the test for each sample size was obtained by comparing (5.5) with the critical value from the $\chi^2(1)$ distribution for $\alpha = 10\%$ being 2.70554 which, given the mixture nature of the asymptotic null distribution, is the critical value for a test at the 5% nominal level, refer to Sections 4.2.2 – 4.2.3. We carried out 600 replications in the simulation experiment for each combination of the matrix $\boldsymbol{G}$ and the sample size. Results are listed in *Table 5.1*. We refer to Appendix 5 A.1 for more details on the *S-Plus* code for the MLR* test.

The following four design matrices were used to generate the data:

Data set 1, $G_1\,(n \times 3)$, where each value in the first is column equal to one, in the second column is a random normal variate with mean 5 and standard deviation 1 and the third column is created by a sequence of numbers equally spaced between zero and one.

Data set 2, $G_2\,(n \times 4)$, where the first three columns are the same as in matrix G_1 and the fourth column is the USA manufacturing data from the dX Data File, from part Main Economic Indicators, Table USA 04: Manufacturing. The raw manufacturing monthly data from 1960 to 2000 are presented in *Figure 4.1*, refer to Section 4.3.1. We consider the latest 50 or 100 observations for the purpose of our simulation study.

Data set 3, $G_3\,(n \times 5)$, where the first four columns are the same as in matrix G_2 and the fifth column is the USA manufacturing data lagged by one period.

Data set 4, $G_4\,(n \times 6)$, where the first five columns are the same as in matrix G_3 and the sixth column is the USA Manufacturing data lagged by two periods.

n	G_1	G_2	G_3	G_4
50	0.031	0.028	0.025	0.022
100	0.023	0.022	0.020	0.016

Table 5.1: *Estimated size of the MLR_* test based on asymptotic critical values at the 5% nominal level*

The results of the simulation study on the size of the MLR* test given in *Table 5.1* reveal that the estimated test sizes are in a range between 0.016 and 0.031 (whereas the nominal size was set to 0.05). The majority of results for sample size $n = 50$ are not significantly different from the nominal value at the 1% significance level. All

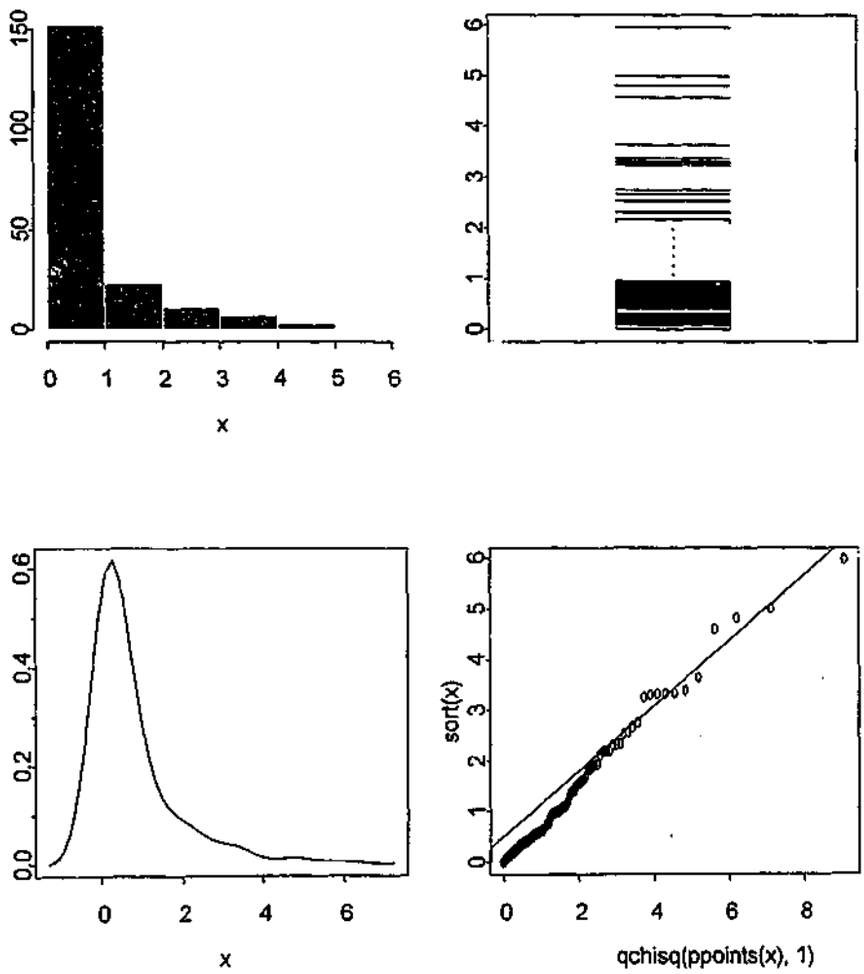

Figure 5.1: *The four plot summary of the distribution of MLR* test statistic larger than zero for* G_1

results for the sample size $n = 100$ are not significantly different from the nominal value at the 5% significance level. The size of the test seems not to be invariant to changes in sample size. The size of the MLR* test is larger for the smaller sample

size. The size of the MLR* test is smaller for larger dimensions of the G matrix for both sample sizes.

We applied the same collection of four summary plots, namely the histogram, the boxplot, the density curve and the qq-plot as in Section 4.2.4 to check whether the non-zero values of the MLR* test statistic follow a $\chi^2(1)$ distribution. The four plot summary of the MLR* > 0 with design matrix G_1 is depicted by *Figure 5.1.*

From *Figure 5.1* the histogram, boxplot and density plots reveal an asymmetric, skewed, $\chi^2(1)$ like distribution for simulated MLR* > 0. From the qq-plot it appears that the $\chi^2(1)$ distribution is not too far from the true distribution. There are some extreme values, but the majority of points cluster along the straight line.

The four plot summaries of the size MLR* > 0 with design matrixes G_2, G_3, G_4 revealed an asymmetric, skewed, $\chi^2(1)$ like distribution in shape similar to that presented by *Figure 5.1.*

5.3.2 Power of the MLR* test

We investigated the power of the MLR* test using the same ten nonlinear test functions as in the Section 4.3.2.

Function	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10
Theta										
0.00	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031
0.05	0.035	0.042	0.065	0.058	0.055	0.129	0.097	0.087	0.063	0.058
0.10	0.037	0.059	0.104	0.107	0.089	0.467	0.175	0.208	0.078	0.076
0.15	0.043	0.086	0.178	0.209	0.169	0.833	0.283	0.397	0.124	0.103
0.20	0.051	0.125	0.297	0.332	0.252	0.974	0.463	0.683	0.178	0.120
0.30	0.059	0.239	0.625	0.661	0.502	1.000	0.810	0.917	0.446	0.265
0.40	0.071	0.420	0.867	0.903	0.745		0.962	0.995	0.758	0.445
0.60	0.108	0.740	0.995	0.994	0.975		1.000		0.935	0.756
0.80	0.178	0.943	1.000		1.000				0.998	0.950
1.00									1.000	0.997
1.20										

Table 5.2: *Power estimates of the MLR* test for ten different nonlinear functions at the 5% nominal level using asymptotic critical value*

Let the matrix G in (5.1) be the three-column matrix with each value in the first column equal to one, the second column being a random normal variate with mean 5 and standard deviation 1, and the third column being created by a sequence of numbers equally spaced between zero and one. We carried out 600 replications on the same set of test nonlinear functions given by (4.25). Results of this simulation are presented in *Table 5.2.* We stoped simulating the power of MLR* test for some values of θ for some functions because the increment in the power by increasing the θ value was very small compared to other non-linear functions.

Figure 5.2 shows the powers of the MLR* test for ten non-linear test functions. The results show that the MLR* test performs well for the majority of the test functions. The best power was obtained for function f_6. There is also visible a distinctive group

of functions f_7, f_8 and f_3, f_4, f_5, f_9 which exhibit similar power. Also functions f_2, f_{10} result in low power for smaller θ values and have a quick increase in power for larger θ. The results are not satisfactory for the Gompertz function f_1 which is the least nonlinear of the set of ten deterministic curves.

Figure 5.2 shows a power comparison of the MLR* test for the ten non-linear test functions.

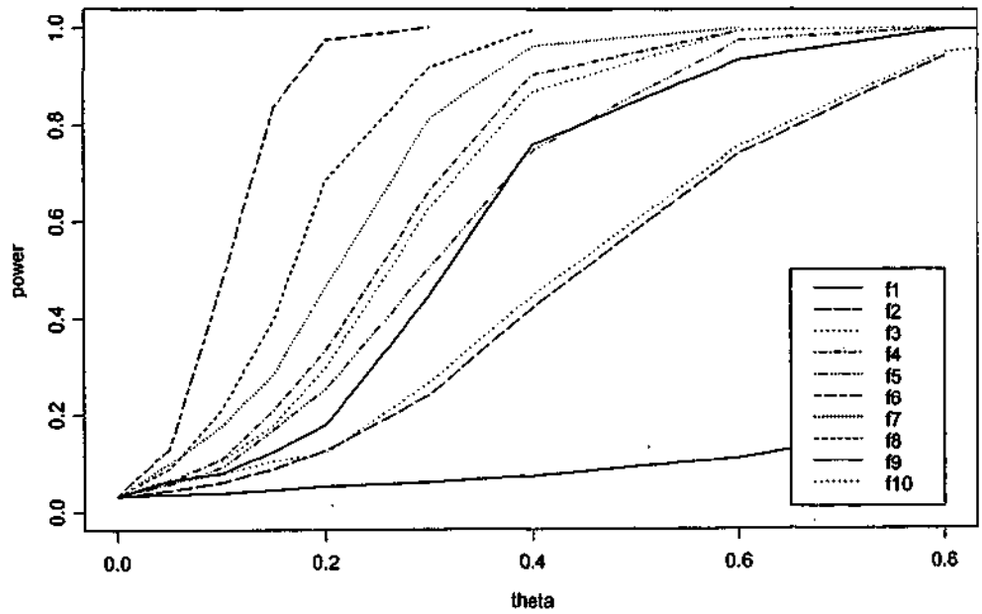

Figure 5.2: *Power of the MLR* test for all ten non-linear test functions*

Overall, the results show that the MLR* test performs well. The lowest power is obtained for the function f_1. The MLR* test does not perform satisfactorily for the least nonlinear functions of the set of ten deterministic curves, refer to *Figure 4.6 (a)*

– *(e)* in Chapter 4. The highest power is obtained for function f_6. The power of the MLR* test depends upon the degree of non-linearity of the test function $f(t_i)$ as might be expected.

5.4 Comparison of four tests

We now compare our MLR* test with the locally most powerful invariant (LMPI) test proposed by King and Hillier (1985), the modified regression specification error F* test by Ramsey (1969) and F*spline test.

5.4.1 Locally most powerful invariant test

The LMPI test for the null hypothesis $H_0 : \kappa = 0$ against the specific alternative $H_a : \kappa = \kappa_1 > 0$ rejects H_0 for large values of the test statistic

$$T_{LMPI} = \frac{e_0' \Sigma e_0}{e_0' e_0} \tag{5.6}$$

where e_0 is the ordinary least squares residual vector for the regression of y on X and Z as given by (3.14) and Σ is given by (3.21). The critical value for our case was calculated using standard numerical methods as outlined in King (1981) and is 0.00826 at 0.95. (The author wants to thank to Zeng-Hua Lu from Department of

Econometrics and Business Statistics, Monash University for his calculation of the critical value for the LMPI test using the Imhof (1961) algorithm.)

5. 4. 2 Regression specification error test – F* and F* spline test

The specification error considered in this chapter is that of the misspecified type of regression. We have a linear regression of $s(t)$ that might be nonlinear. We want to apply the Ramsey error misspecification test to our model, refer to Section 4.4.1 in Chapter 4.

Recall (5.1), the specification of the null hypothesis is given by the restriction

$$Q_* = e \tag{5.7}$$

where e is a vector of independent disturbance terms each distributed normally with mean zero and variance σ^2. This leads to the restricted model

$$Y = X\gamma + Z\beta + e. \tag{5.8}$$

The alternative hypothesis is defined by specifying that the true model has some specification other than that given by (5.8). Thus, the use of the restricted regression to analyse data generated by some other model leads to specification error. We want to test the non-linear component in Q_*, therefore the specification error considered

here is that of misspecified type of $s(t)$ in (5.1). Under the alternative hypothesis, the specification of the true model is given by (3.13).

For a test of significance level α, the null hypothesis is rejected if the test statistic

$$T_{F^*} > F_{m,n-m-k,\alpha} \tag{5.9}$$

in which

$$T_{F^*} = \frac{\left(SSE_R - SSE_U\right)/m}{SSE_U/(n-m-k)}, \tag{5.10}$$

where SSE_R is the error sum of squares resulting from restricted model (5.8), and SSE_U represents the error sum of squares resulting from unrestricted model (3.13). The number of parameters being estimated in the unrestricted model is denoted by k and the symbol m is used for the number of restrictions, which is 2 in this case. The test statistic T_{F^*} follows the F-distribution under the null hypothesis. The difference between the F test from Chapter 4 and the F^* test is in the matrix X specification. In the F test, the matrix X is a two-column matrix with each value in the first column equal to one, with a random normal variate with mean 5 and standard deviation 1 in the second column. In the F^* test, the same matrix X is combined with the matrix Z to make the matrix $\boldsymbol{G} = [\boldsymbol{X} : \boldsymbol{Z}]$. The matrix $\boldsymbol{G}$ is the three-column matrix with the first two columns the same as those for X and the third column is created by a

sequence of numbers equally spaced between zero and one. For more details on the difference between the F test and the F^* test refer to Appendix 4A.2 and 5A.3.

Similarly as in Chapter 4, function $s(t_i)$ in (3.13) can be estimated by any nonlinear function. Our choice was a cubic polynomial function. We denote this type of regression specification error test as the F^* test. If in (3.13), function $s(t_i)$ is estimated non-parametrically by a smooth spline function, we denote this type of test as the F^* *spline* test, hence the test statistic is denoted as $T_{F^*spline}$. For more details on the *S-Plus* codes on F^* and F^* *spline* see Appendix 5A.3 and 5A.4.

The same discussion applies for the test statistic $T_{F^*spline}$ in F^* *spline* test as in Section 4.4.1 regarding how to get the number of degrees freedom. The difference between F *spline* test from Chapter 4 and F^* *spline* test is in the design matrix X and G specification, similar to the difference between the F and F^* tests. For more details of the difference refer to Appendix 4A.3 and 5A.4.

5.4.3 Simulation comparison

We carried out 600 replications with $n = 50$ with the regressors and nonlinear functions being the same as in Section 5.3.2 See Appendices 5A.1 – 5A.4 for *S-Plus* codes of all four tests. The results of the simulations are listed in *Table 5.3*.

	Theta	0.00	0.05	0.10	0.15	0.20	0.30	0.40	0.60	0.80	1.00
Function	Test										
f1	MLR*	0.031	0.035	0.037	0.043	0.051	0.059	0.071	0.108	0.178	
	LMPI	0.046	0.050	0.056	0.056	0.056	0.065	0.080	0.123	0.193	
	F*	0.052	0.054	0.056	0.060	**0.071**	**0.085**	**0.102**	**0.150**	**0.216**	
	F* spline	0.056	**0.058**	**0.061**	**0.062**	0.063	0.070	0.085	0.136	0.208	
f2	MLR*	0.031	0.042	0.059	0.086	0.125	0.239	0.420	0.740	0.943	
	LMPI	0.046	**0.065**	**0.095**	**0.150**	**0.216**	**0.375**	**0.566**	**0.873**	**0.980**	1.000
	F*	0.052	0.060	0.073	0.110	0.133	0.250	0.393	0.751	0.948	0.995
	F* spline	0.056	**0.065**	0.073	0.100	0.118	0.235	0.378	0.713	0.941	
f3	MLR*	0.031	0.065	0.104	0.178	0.297	0.625	0.867	**0.995**	1.000	
	LMPI	0.046	**0.100**	**0.168**	**0.263**	**0.443**	**0.780**	**0.931**	0.990	1.000	
	F*	0.052	0.090	0.120	0.220	0.341	0.650	0.888	0.990	1.000	
	F* spline	0.056	0.080	0.108	0.180	0.287	0.630	0.867	0.990	1.000	
f4	MLR*	0.031	0.058	0.107	0.209	0.332	0.661	0.903	0.994		
	LMPI	0.046	**0.078**	0.151	0.235	0.370	0.663	0.850	0.987	1.000	
	F*	0.052	**0.078**	**0.158**	**0.250**	**0.433**	0.760	0.948	1.000		
	F* spline	0.056	0.076	0.108	0.230	0.415	**0.795**	**0.971**	1.000		
f5	MLR*	0.031	0.055	0.089	0.169	0.252	0.502	0.745	0.975	1.000	
	LMPI	0.046	**0.080**	**0.152**	**0.240**	**0.363**	0.650	0.860	0.990	1.000	
	F*	0.052	0.063	0.115	0.200	0.285	0.600	0.792	0.986	1.000	
	F* spline	0.056	0.063	0.090	0.170	0.302	**0.670**	**0.870**	**0.995**	1.000	
f6	MLR*	0.031	0.129	0.467	0.833	0.974	1.000				
	LMPI	0.046	**0.168**	**0.543**	**0.873**	**0.995**	1.000				
	F*	0.052	0.146	0.447	0.823	0.960	1.000				
	F* spline	0.056	0.135	0.458	0.836	0.971	1.000				
f7	MLR*	0.031	0.097	0.175	0.283	0.463	0.810	0.962	1.000		
	LMPI	0.046	0.040	0.038	0.034	0.032	0.020	0.010	0.003		
	F*	0.052	0.080	0.086	0.120	0.162	0.250	0.315	0.443	0.521	
	F* spline	0.056	**0.120**	**0.250**	**0.620**	**0.833**	**0.950**	1.000			
f8	MLR*	0.031	0.087	0.208	0.397	0.683	0.917	0.995			
	LMPI	0.046	**0.113**	**0.242**	**0.500**	**0.743**	**0.981**	1.000			
	F*	0.052	0.100	0.210	0.390	0.618	0.915	0.985	1.000		
	F* spline	0.056	0.090	0.217	0.490	0.728	0.973	1.000			
f9	MLR*	0.031	0.063	0.078	0.124	0.178	0.446	0.758	0.935	0.998	1.000
	LMPI	0.046	**0.090**	**0.112**	0.176	0.270	0.450	0.648	0.912	0.983	0.998
	F*	0.052	0.070	0.086	0.134	0.178	0.310	0.493	0.823	0.953	0.995
	F* spline	0.056	0.080	0.108	**0.180**	**0.317**	**0.700**	**0.938**	**0.997**	1.000	
f10	MLR*	0.031	0.058	0.076	0.103	0.120	0.265	0.445	0.756	0.950	0.997
	LMPI	0.046	0.060	0.088	0.110	0.175	0.312	0.460	0.733	0.898	0.967
	F*	0.052	**0.080**	**0.091**	**0.143**	**0.190**	**0.360**	**0.540**	0.860	0.980	1.000
	F* spline	0.056	0.070	0.086	0.130	0.163	0.354	0.527	**0.886**	**0.990**	1.000

Table 5.3: *Simulated power of the MLR*, LMPI, F* and F* spline tests for the ten nonlinear test functions*

From *Table 5.3* it is evident that the power of the MLR*, LMPI, F* and F* spline tests differ according to the different deterministic nonlinear functions being used in the study. All four tests are most powerful for function f_6. Our MLR* test performed well for all of the test functions. However it does not have the best power for any function. Note that the size of MLR* test is the smallest of all four tests therefore the power of the MLR* test is lower as a result of this.

When comparing the MLR* test with LMPI test (both are variance-covariance based tests), the latter outperformed the former for functions $f_1, f_2, f_3, f_5, f_6, f_8$. The MLR* test performs better than the LMPI test for larger θ for functions f_9, f_{10}. The LMPI test failed for function f_7, for which the MLR* test revealed very good power. The mean based F* or F* spline tests showed the best power for functions $f_1, f_4, f_7, f_9, f_{10}$. In fact the F* spline test was always more powerful than the F* test. This indicates the better flexibility of the spline based F test to the polynomial based F test. Overall the LMPI test had the best power for functions f_2, f_3, f_6, f_8 and partly for function f_5. Therefore when testing for linearity of a possibly nonlinear component, the mean based tests performed on average equally well compared to the variance based tests. In *Table 5.3* the best power was highlighted for each test and nonlinear test function with respect to θ.

Figures 5.3 (a) – (j) plot the power of the four tests for each individual deterministic test function. These plots give a better graphical comparison of all four tests for every nonlinear test function.

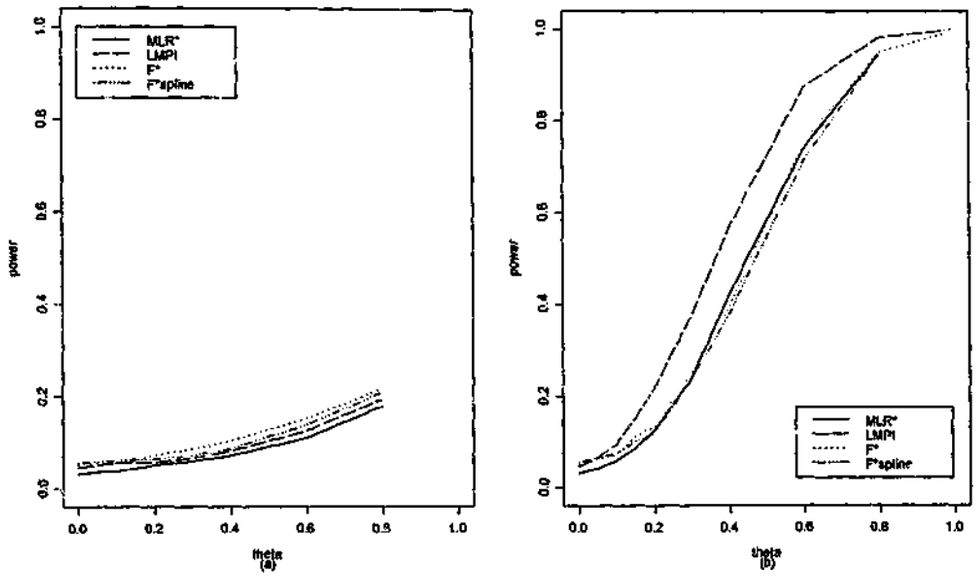

Figure 5.3 (a) – (b): *Power of the MLR*, LMPI, F* and F* spline tests for function* f_1 *part (a) and* f_2 *part (b)*

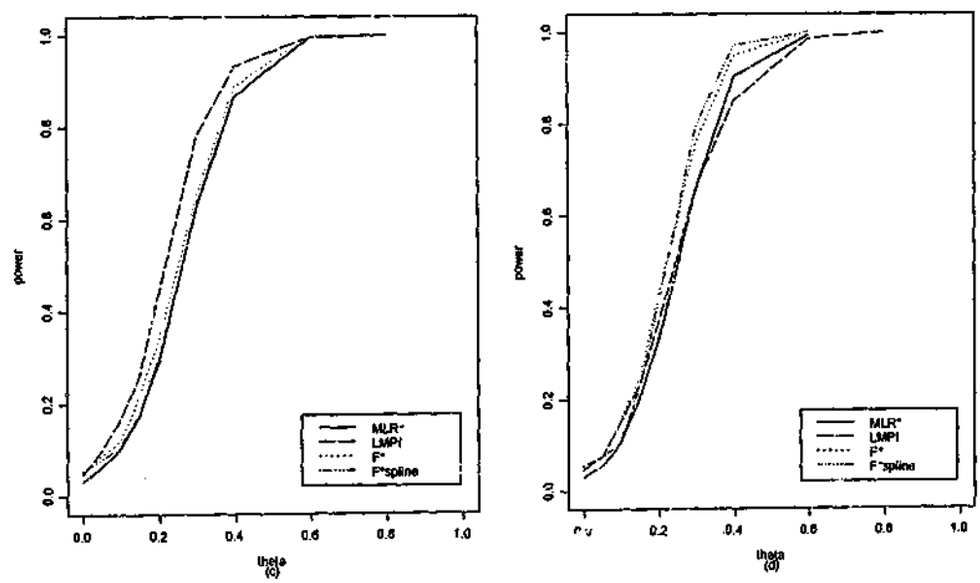

Figure 5.3 (c) – (d): *Power of the MLR*, LMPI, F* and F* spline tests for function* f_3 *part (c) and* f_4 *part (d)*

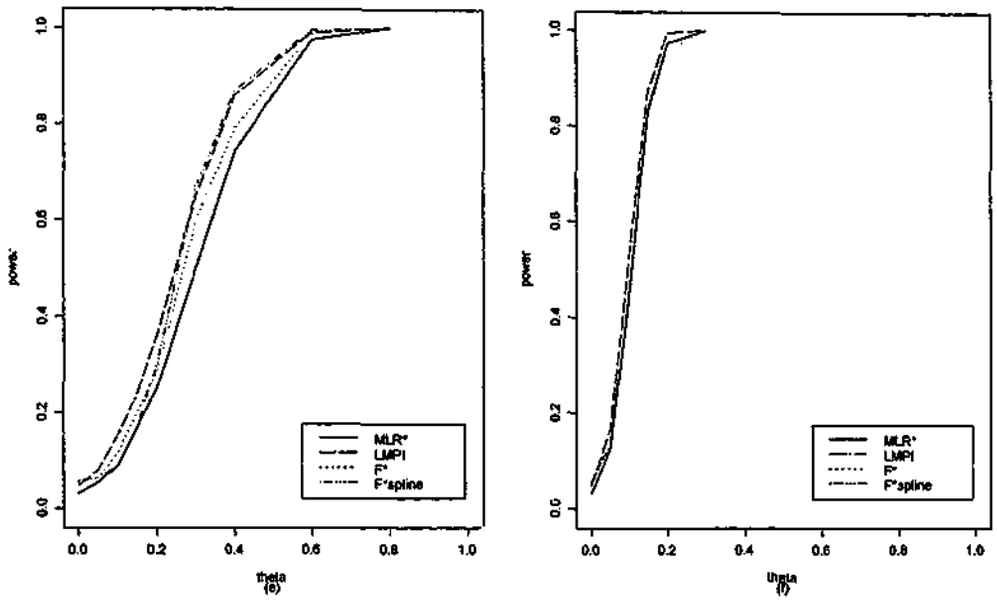

Figure 5.3 (e) – (f): *Power of the MLR*, LMPI, F* and F* spline tests for function* f_6 *part (e) and* f_6 *part (f)*

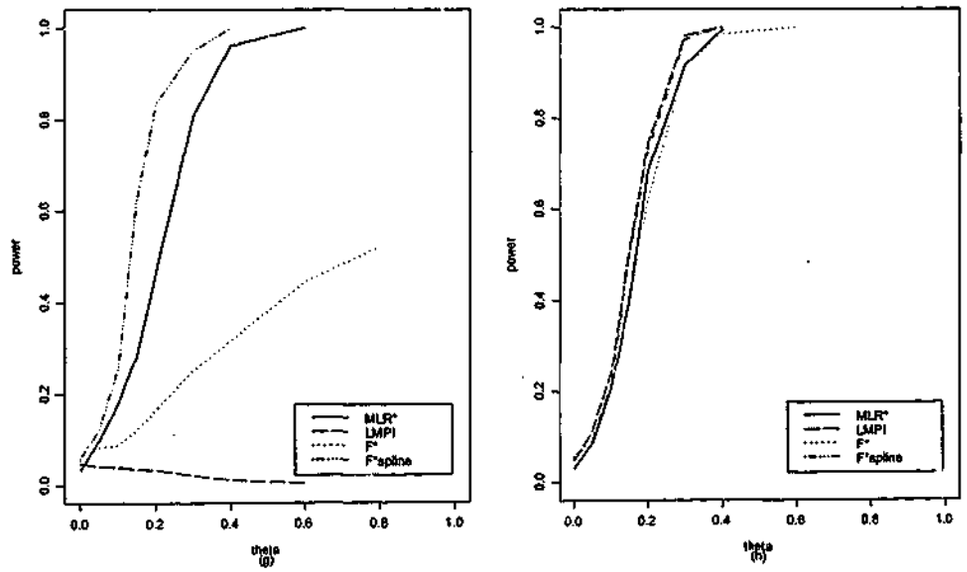

Figure 5.3 (g) – (h): *Power of the MLR*, LMPI, F* and F* spline tests for function* f_7 *part (g) and* f_8 *part (h)*

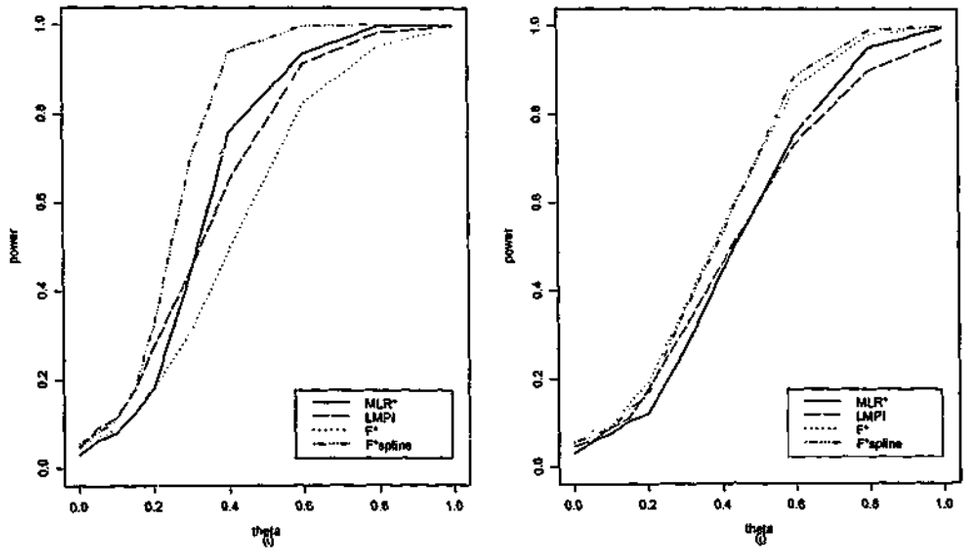

Figure 5.3 (i) – (j): *Power of the MLR*, LMPI, F* and F* spline tests for function f_9 part (g) and f_{10} part (h)*

Plots *(f), (h), (c), (d), (e),* and *(a),* related to functions $f_6, f_8, f_3, f_4, f_5, f_1$ show very small deviations on results of all four tests. Plots *(b)* and *(i)* related to functions f_2 and f_9 show moderate deviations in the results of all four tests. Plot *(g)* is very different to the others because the LMPI test failed and the F* test shows very poor power compared to the F* spline and the MLR* tests. These results are most likely due to the different degree of nonlinearity of each of the ten deterministic functions being used in the study. Most importantly the MLR* test did not fail for any test function.

5.5 The MLR* test application

Recall the semiparametric model (3.1) and the data sets outlined in Section 4.5. The conclusion of MLR test from Chapter 4 was to include the time varying risk premia into the model. It is appropriate to know the functional form in which this time varying risk premia should be included into the model. Recall Figure 4.6. The 6-months hold time dependency is graphed on Figure 5.4

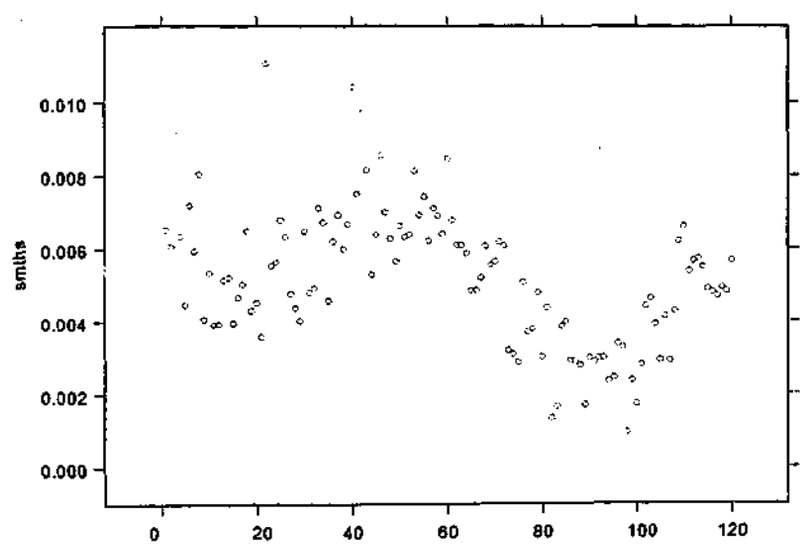

Figure 5.4: *6-months holds to maturity data from 31/12/85 to 29/12/95*

There is a visible time varying dependance in some non-linear fashion. We can try to model this dependency by a polynomial function, however we are interested to apply

the cubic smoothing spline approach. The MLR* test has $H_0 : \kappa = 0$, which means that the risk premia is linear over the time, refer to Figure 5.5. The specific alternative $H_A : \kappa > 0$ means that the risk premia is non-linear, refer to Figure 5.6.

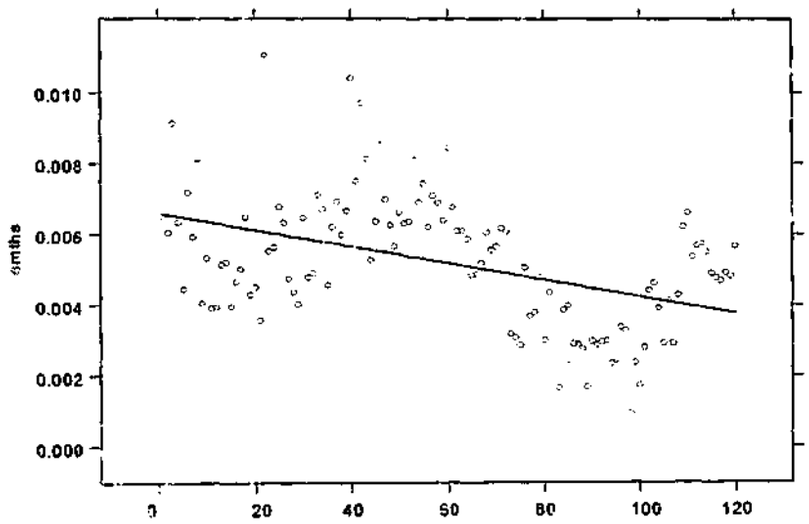

Figure 5.5: *6-months holds to maturity data with fitted linear trend*

Result of the MLR* test for hold data is $MLR^* = 23.197$ is supported by the *p-value* being 0.00. We have a significant evidence to reject $H_0 : \kappa = 0$ at 5% level when we compare MLR* test statistic with $\chi^2(1) = 2.70554$.

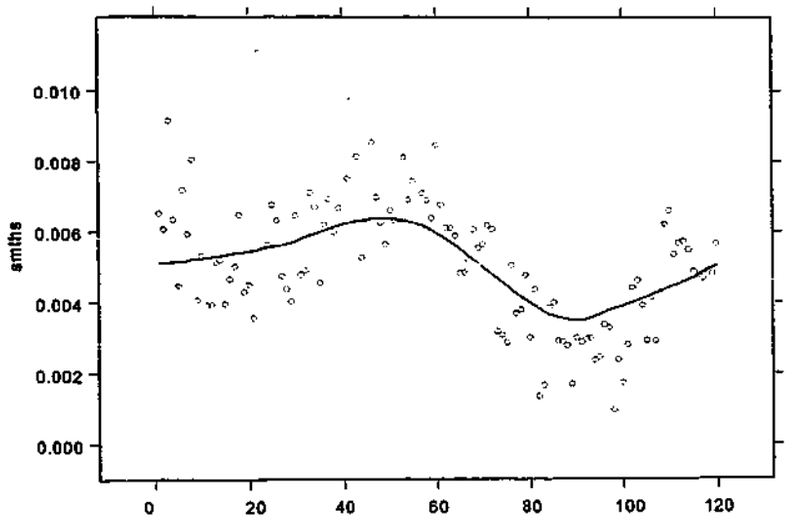

Figure 5.6: *6-months holds to maturity data with cubic spline superimposed (λ was estimated by GCV)*

We also applied the MLR* test to yields maturity return for 3 and 6 month data from the same source. The MLR* test confirmed the previous finding and rejected H_0 at the 5% nominal significance level.

According to MLR test from Chapter 4, we have significant evidence to include the time varying risk premia component $s(t)$ for both hold and yields data into the semiparametric model. Moreover the MLR* test results suggest that the time varying risk premia component $s(t)$ is non-linear. One possibility is to model $s(t)$ nonparametrically, which depends on its non-linear complexity.

Our finding confirmed the intuitive idea that the time varying risk premium should be included in the model as a non-linear component. Apart of the risk premia, perhaps other economic environmental explanatory variables of the 6 months hold (yields) to maturity such as interest rates are believed to be non-linear as well, because of uncertainty caused by the volatility in monetary growth. Their inclusion in the model in a nonlinear form is worth of exploring because of the benefits resulting from improved term structure models with extra nonlinear variables.

5.6 Results summary

In this chapter we proposed the modified version of the marginal likelihood ratio MLR* test for testing for linearity of a possible nonlinear component in a semiparametric regression model. The simulation study on the MLR* test revealed a good size of the test. However the size of the MLR* test depends on the sample size and the number of regressors in the model. The simulated size of the MLR* test is larger for smaller sample sizes. When the number of regressors in the model is increased, the simulated size of the MLR* test tends to decrease for both sample sizes.

Also the study of the power of the MLR* test using the ten nonlinear deterministic test functions was successful. The simulation study comparing our MLR* test with the LMPI, F* and F* spline tests showed that MLR* did not outperform the other tests. Nevertheless our MLR* test extends the family of possible tests for regression misspecification.

We also compared the mean based F and F* spline tests with the variance based MLR* and LMPI tests. The variance based test, namely the LMPI test, was the best alternative for the five out of the ten test functions. The mean based tests, either the F* or the F* spline test, were the best alternative for the reminding five test functions. The comparison of the powers are not particularly conclusive to be favourable to the mean or the variance based tests. Considering the parsimony of the testing procedure, the mean based tests are more convenient to formulate and apply.

We demonstrated the MLR* test performance on the nominal yields to maturity data and the holding period return data from the Fama Twelve-Month Treasury Bill Term Structure File of the U.S. Government Securities.

Appendix 5A.1 *S-Plus* code of MLR * test

```
#####################################
# MLR star test, size and power simulations#
#####################################

### Generate x data
n <- 50
set.seed(1)
x <- rnorm(n,5,1)
tt <- (1:n)/n

## Set up equation for y data
gamma1 <- 0 #intercept
gamma2 <- 1 #slope
sigma <- 0.1 #stdev

### Set up X and S matrices
ones <- seq(1,1,l=n)
XX <- cbind(ones,x,tt)
halfN <- 0.5*(n-ncol(XX))

### Set up Sigma matrix
h3 <-1/(n^3)
Sigma <- matrix(0,nrow=n,ncol=n)
for (i in 1:n)
        Sigma[i,i:n] <- Sigma[i:n,i] <- (i*i*(i+1.5*(0:(n-i))))/3 *h3

### Set up nonlinear functions
f0 <- tt
f1 <- exp(-3*exp(-3*tt))
f2 <- tt^2
f3 <- exp(-5*(1-tt))
f4 <- 0.71*(sin((4*tt)^2)+3*tt^8)/3+.25
f5 <- (0.909*(1/sin(3*tt))/15)-0.04
f6 <- 1.4*(exp(sin(3*tt)^2)/2.5)-0.55
f7 <-0.5*sin(4*pi*tt)+0.5
f8 <-((5*tt+5*exp(-((tt-0.5)^2)/0.04))/7)*0.909
f9 <- 2.5*((exp(sin(3*tt)^3)+exp(cos(3*tt)^2)+tt^4)/4.5)-1.7
f10 <-((exp(sin((3*tt)^3))+4*tt^4)/6)-0.1

### Functions to compute likelihood and MLR star statistic (here lr)
logdet <- function(x)
        sum(log(eigen(x)$values))

fl<-function(ll)
{
        C <- chol(diag(n) + exp(ll)*Sigma)
        Cinv <- t(solve(C))
        Xstar <- Cinv %*% XX
        Ystar <- Cinv %*% Y
        tX <- t(Xstar)
        XXX <- tX %*% Xstar
        beta <- solve(XXX) %*% tX %*% Ystar
        res <- Ystar - Xstar %*% beta
        return(logdet(C) + 0.5*logdet(XXX) + halfN*log(sum(res*res)))
}

minfl <- function()
{
```

```
        cat(".")
        y <-  fmean + rnorm(n,0,sigma)
        Y <- matrix(y,ncol=1)
        assign("Y",Y,frame=1)
        fl0 <- fl(-1e20) # Value at zero
        fredall <- nlmin(fl,1) # Minimize starting near zero
        fl.fred <- fl(fredall$x) # Best fl
        if(fl0 < fl.fred)
                lstar <- lr <- 0
        else
        {
                lr <- 2*(fl0 - fl.fred)
                lstar <- exp(fredall$x)
        }
        return(c(lr,lstar))
}

##########################################
## Simulation to compute size and power ##
##########################################

theta<- 1
set.seed(23)
fmean<-gamma1 + gamma2*x + theta*f10
nn <- 600
lmat <- matrix(0,nrow=nn,ncol=2)

for (k in 1:10)     lmat[k,] <- minfl()
.
.
.
for (k in 591:600)lmat[k,] <- minfl()

lr <- lmat[,1]
lstar <- lmat[,2]
count2 <- length(lr[lr>2.70554])
print(count2)
power<-count2/nn
print(power)
```

Appendix 5A.2 *S-Plus* code of LMPI test

```
##################################
## LMPI test, size and power  simulation#
##################################

### Generate x data
n <- 50
set.seed(1)
x <- rnorm(n,5,1)
tt <- (1:n)/n

### Set up equation for y data
gamma1 <- 0 #intercept
gamma2 <- 1 #slope
sigma <- 0.1 #stdev

### Set up Sigma matrix
h3 <- 1/(n^3)
Sigma <- matrix(0,nrow=n,ncol=n)
for (i in 1:n)
        Sigma[i,i:n] <- Sigma[i:n,i] <- (i*i*(i+1.5*(0:(n-i))))/3*h3

### Set up nonlinear functions
f0 <- tt
f1 <- exp(-3*exp(-3*tt))
f2 <- tt^2
f3 <- exp(-5*(1-tt))
f4 <- 0.71*(sin((4*tt)^2)+3*tt^8)/3+.25
f5 <- (0.909*(1/sin(3*tt))/15)-0.04
f6 <- 1.4*(exp(sin(3*tt)^2)/2.5)-0.55
f7 <-0.5*sin(4*pi*tt)+0.5
f8 <-((5*tt+5*exp(-((tt-0.5)^2)/0.04))/7)*0.909
f9 <- 2.5*((exp(sin(3*tt)^3)+exp(cos(3*tt)^2)+tt^4)/4.5)-1.7
f10 <-((exp(sin((3*tt)^3))+4*tt^4)/6)-0.1

### Calculate statistic
lmpisim <- function()
{
        y <- fmean + rnorm(n,0,sigma)
        e0 <- residuals(lm(y ~ x + tt))
        tmp <- Sigma %*% e0
        lmpi <- sum(e0*tmp) / sum(e0*e0)
        return(lmpi)
}

#############################
## Simulation to compute power ##
#############################

theta <- 0.6
fmean <- gamma1 + gamma2*x + theta*f3
set.seed(2)
nsim <- 500
lmpi <- numeric(nsim)
for(i in 1:nsim)
        lmpi[i] <- lmpisim()

count <- length(lmpi[lmpi>0.0082611788])
print(count)
```

```
power <- count/nsim
print(power)


##################################
## Calculate critical value by bootstrap##
#######################################

fmean <- gamma1 + gamma2*x
set.seed(2)

nsim <- 600
lmpi <- numeric(nsim)
for(i in 1:nsim)
        lmpi[i] <- lmpisim()
lmpi.critical.value <- quantile(lmpi,0.95)
print(lmpi.critical.value)
# 0.008200231
```

Appendix 5A.3 *S-Plus* code of F * test

```
##########################
## F* test, power simulations #
##########################

### Generate x data
n <- 50
set.seed(1)
x <- rnorm(n,5,1)
tt <- (1:n)/n

### Set up equation for y data
gamma1 <- 0 #intercept
gamma2 <- 1 #slope
sigma <- 0.1 #stdev

### Set up nonlinear functions
f0 <- tt
f1 <- exp(-3*exp(-3*tt))
f2 <- tt^2
f3 <- exp(-5*(1-tt))
f4 <- 0.71*(sin((4*tt)^2)+3*tt^8)/3+.25
f5 <- (0.909*(1/sin(3*tt))/15)-0.04
f6 <- 1.4*(exp(sin(3*tt)^2)/2.5)-0.55
f7 <-0.5*sin(4*pi*tt)+0.5
f8 <-((5*tt+5*exp(-((tt-0.5)^2)/0.04))/7)*0.909
f9 <- 2.5*((exp(sin(3*tt)^3)+exp(cos(3*tt)^2)+tt^4)/4.5)-1.7
f10 <-((exp(sin((3*tt)^3))+4*tt^4)/6)-0.1

### Calculate F statistic
Ftestsim<-function()
{
        y<-fmean+rnorm(n,0,sigma)
        ma<-lm(y ~ x +tt+tt^2+tt^3)
        m0<-lm(y ~ x +tt)
        Ftest<-((deviance(m0)-deviance(ma))/2)/(deviance(ma)/44 )
        return(Ftest)
}

#############################
## Simulation to compute power ##
#############################

theta <-0.3
fmean <- gamma1 + gamma2*x + theta*f8
set.seed(23)
nsim <- 600
Ftest <- numeric(nsim)
for(i in 1:nsim)
        Ftest[i] <- Ftestsim()

count <- length(Ftest[Ftest>3.22])
print(count)
power <- count/nsim
print(power)

#########################
## Calculate critical value ##
#################'#######
```

```
fmean <- gamma1 + gamma2*x
set.seed(2)

nsim <- 600
Ftest <- numeric(nsim)
for(i in 1:nsim)
        Ftest[i] <- Ftestsim()
Ftest.critical.value <- quantile(Ftest,0.95)
print(Ftest.critical.value)
```

Appendix 5A.4 *S-Plus* code of F* spline test

```
##############################################################
# F* test assuming that the additional nonlinear component is estimated #
# by smoothing spline, power simulation                                  #
##############################################################

### Generate x data
n <- 50
set.seed(1)
x <- rnorm(n,5,1)
tt <- (1:n)/n

### Set up equation for y data
gamma1 <- 0 #intercept
gamma2 <- 1 #slope
sigma <- 0.1 #stdev

### Set up nonlinear functions
f0 <- tt
f1 <- exp(-3*exp(-3*tt))
f2 <- tt^2
f3 <- exp(-5*(1-tt))
f4 <- 0.71*(sin((4*tt)^2)+3*tt^8)/3+.25
f5 <- (0.909*(1/sin(3*tt))/15)-0.04
f6 <- 1.4*(exp(sin(3*tt)^2)/2.5)-0.55
f7 <-0.5*sin(4*pi*tt)+0.5
f8 <-((5*tt+5*exp(-((tt-0.5)^2)/0.04))/7)*0.909
f9 <- 2.5*((exp(sin(3*tt)^3)+exp(cos(3*tt)^2)+tt^4)/4.5)-1.7
f10 <-((exp(sin((3*tt)^3))+4*tt^4)/6)-0.1

theta <-0.3
fmean <- gamma1 + gamma2*x + theta*f8

nsim<-600
Fspl<-numeric(nsim)
        for (i in 1:nsim)
        {

                y<-fmean+rnorm(n,0,sigma)
                ma<-gam(y ~ x + s(tt))
                m0<-gam(y ~ x + tt)
                anova(ma,m0)
                junk<-anova(m0,ma,test="F")
                Fspl[i]<-(junk$"F Value"[2])

        }

count<-length(Fspl[Fspl>3.106669])
power<-count/nsim
print(power)

#crit.val by bootstrap 1000 repl. is 3.106669#
#Fcrit 4,39 d.f. alpha 0.05 is 2.62 #
```

CHAPTER 6

Choosing the smoothing parameter – bandwidth selection

6.1 Introduction

We mentioned the problem of choosing an appropriate value of the smoothing parameter λ in Section 2.4. The "correct" value of λ is important in smoothing spline techniques in order to avoid over or under smoothing. This chapter examines the methods of choosing the smoothing parameter λ in univariate smoothing spline techniques. The chapter is organised as follows. Section 6.2 describes the estimation methods being used for estimating the smoothing parameter. Section 6.3 defines our approach to smoothing parameter λ selection, which is based on the *marginal likelihood function.* We proposed the *likelihood ratio* test for testing for the inclusion

of a possibly nonlinear component in Chapter 4 and for testing for linearity of a possibly nonlinear component in Chapter 5 of this thesis. The smoothing parameter λ is a by-product of the maximising procedure used in our tests. We now investigate the properties of the smoothing parameter λ obtained by this approach and compare it with estimates from existing methods. In Section 6.4 we report on an extensive simulation study of the unknown function estimation when the smoothing parameter is estimated by generalised cross-validation and our marginal likelihood and compare the estimates with the value of λ which minimizes the true squared error. Section 6.5 compares our marginal likelihood method with those by Kohn et al. (1991). Section 6.6 summarises the results. The appendices provide details of the *S-Plus* code for the simulations and results obtained from them.

6.2 Description of the estimation methods proposed in the literature

We introduced some methods for estimating the smoothing parameter λ in Chapter 2. We explain the most frequently used methods in more detail in this section. Methods based on cross-validation are discussed in Section 6.2.1 and methods based on the maximum likelihood are presented in Section 6.2.2.

6.2.1 Methods based on cross-validation

Consider the general regression (6.1). In this regression model where we have observations y_i at the design points t_i , $i = 1,\ldots,n$, and the observations are assumed to satisfy

$$y_i = s(t_i) + \varepsilon_i \,, \tag{6.1}$$

with $s(t)$ being a smoothed curve and $\varepsilon_i, i = 1,\ldots,n$, iid $\mathrm{N}(0,\sigma^2)$. The design points $t_1,\ldots,t_n$ are assumed to be distinct and ordered so that without loss of generality $0 \le t_1 < t_2 < \ldots < t_n \le 1$. A smoothing spline of order m that estimates $s(t)$ is obtained by minimising the penalised squares

$$S(\lambda) = \sum_{i=1}^{n} \left(y_i - s(t_i)\right)^2 + \lambda \int_0^1 \left(s^{(m)}(t)\right)^2 dt \tag{6.2}$$

over all functions $s(t)$ having square integrable mth derivative, where $\int \left(s^{(m)}(t)\right)^2$ represents a roughness measure. For given λ, the function $\hat{s}(t;\lambda)$ is a spline of degree $2m-1$, that is a piecewise polynomial of degree $2m-1$ between the design points t_i with $2m-2$ continuous derivatives across the points. Cubic smoothing splines are derived as a minimisation of the penalised square problem with respect to $s(t_i)$ and λ for $m = 2$ as follows

$$S(\lambda)=\sum_{i=1}^{n}\left(y_i-s\left(t_i\right)\right)^2+\lambda\int_0^1\left(s''(t)\right)^2 dt. \tag{6.3}$$

Cross-validation is the standard data-driven method for determining the smoothing parameter λ in equation (6.2) and (6.3). This method is also known as "leave-out-one" estimation. The principle is to leave the data points out one at a time and to select that value of λ under which the missing data points are best predicted by the remainder of the data. Assuming that the random error in (6.1) has zero mean, the true regression curve s has the property that, if an observation y is taken at the point t, the value of $s(t)$ is the best predictor of y in terms of mean square error. Let $\hat{s}_\lambda^{(-i)}$ be the smoothing spline calculated from all the data pairs expect $\left(t_i,y_i\right)$, for the value λ, then by Wahba and Wold (1975), the cross-validation score is defined by

$$CV(\lambda)=n^{-1}\sum_{i=1}^{n}\left(y_i-\hat{s}_\lambda^{(-i)}\left(t_i\right)\right)^2. \tag{6.4}$$

The smoothing parameter can be chosen objectively from the data by minimising the cross-validation criterion. It cannot be guaranteed that (6.4) has a unique minimum. By Green et al. (1994), expression (6.4) can be simplified using the hat matrix $A(\lambda)$ as follows

$$CV(\lambda)=n^{-1}\sum_{i=1}^{n}\left(\frac{y_i-\hat{s}\left(t_i\right)}{1-A_{ii}(\lambda)}\right)^2, \tag{6.5}$$

where $\hat{s}$ is the spline smoother calculated from the full data set $\{(t_i, y_i)\}$ with smoothing parameter λ and $A_{ii}(\lambda)$ is the ith diagonal element of the hat matrix. $A_{ii}(\lambda)$ is called the leverage value, because it measures the potential for the observed response at $t_i(y_i)$ to employ influence on the fitted value at $t_i(\hat{y}_i)$. An observation with a high leverage is potentially problematic. The values of the smoothing spline $\hat{s}$ depend linearly on the data $\{y_i\}$ through the equation $\hat{s} = A(\lambda)\mathbf{y}$. If the diagonal entries $A_{ii}(\lambda)$ are known, the cross-validation score can be calculated from the residuals $\{y_i - \hat{s}(t_i)\}$ about the spline smoother calculated from the entire data set. Therefore no additional smoothing is required. For more about $A(\lambda)$, $A_{ii}(\lambda)$ and the Reinsch algorithm for their computation, see Green et al. (1994).

Generalised cross-validation (GCV) is a modified version of cross-validation. It is an improved version of CV replacing $A_{ii}(\lambda)$ by its average value $n^{-1}traceA(\lambda)$. The GCV score is constructed, by analogy with CV, by summing the squared residuals corrected by the square of $\{1 - n^{-1}trA(\lambda)\}$. Craven and Wahba (1979) introduced the GCV as follows

$$GCV(\lambda) = n^{-1}\frac{\sum_{i=1}^{n}\{y_i - \hat{s}(t_i)\}^2}{\{1 - n^{-1}trA(\lambda)\}^2}. \tag{6.6}$$

The GCV choice of smoothing parameter is found by minimising the function $GCV(\lambda)$ over λ. By Green et al. (1994) an equivalent form of (6.6) makes clearer the connection between $GCV(\lambda)$ and $CV(\lambda)$, because

$$GCV(\lambda) = n^{-1}\sum_{i=1}^{n}\left\{\left(\frac{1-A_{ii}(\lambda)}{1-n^{-1}trA(\lambda)}\right)^2\left[\frac{y_i-\hat{s}(t_i)}{1-A_{ii}(\lambda)}\right]^2\right\}.$$

If all the $A_{ii}(\lambda)$ were equal, i.e. if t_i were equally spaced, then the $\hat{\lambda}_{GCV}$ would be identical to $\hat{\lambda}_{CV}$. Usually, there will be some differences between the two approaches and $\hat{\lambda}_{GCV}$ will generally be preferable for unequally spaced data. Simonoff (1996) noted that both GCV and CV could cause under-smoothing, data interpolation being the worst case. The same author commented on the effect of outliers on the smoothing spline, particularly at high leverage design points. One solution to this problem is to robustify the roughness penalty criterion to downweight the effect of outliers. He also pointed out that $\hat{\lambda}_{GCV}$ is predisposed to autocorrelation effects, with positive autocorrelation leading to under-smoothing. Moving average type autocorrelation has little effect.

Robinson and Moyeed (1989) suggested a *robustification* of generalised cross-validation (RGCV) given by

$$RGCV(\lambda) = n^{-1}\cdot\frac{1+n^{-1}trA^2(\lambda)}{\left\{1-n^{-1}trA(\lambda)\right\}^2}\sum_{i=1}^{n}\left\{y_i-\hat{s}(t_i)\right\}^2. \qquad (6.7)$$

Neubauer and Schimek (1994) found in their simulation study that robustified generalised cross validation performs better than generalised cross-validation, but may still yield values for λ that under-smooths the data. It is therefore appropriate to use a plot of estimated results and original data in data analysis for visual inspection.

6.2.2 Methods based on maximum likelihood

Wahba (1978) derived the smoothing spline estimate of $s(t)$ in model (6.1) using an integrated Wiener process, recall Section 3.2 in Chapter 3. Wahba (1985) discusses a modified maximum likelihood (what she calls generalized maximum likelihood) for estimating λ (GML). She defined an estimator of λ as the minimiser of the following expression

$$GML(\lambda) = \frac{y'(I - A(\lambda))y}{\left[\det{}^{+}(I - A(\lambda))\right]^{1/n-m}}, \tag{6.8}$$

where $\det^{+}(I - A(\lambda))$ is the absolute value of the product of the $n - m$ non-zero eigenvalues of $(I - A(\lambda))$, and $A(\lambda)$ is the hat matrix. For a cubic smoothing spline, expression (6.8) is

$$GML(\lambda) = \frac{y'(I - A(\lambda))y}{\left[\det{}^{+}(I - A(\lambda))\right]^{1/n-2}}. \tag{6.9}$$

Wahba (1985) also pointed out that if the unknown function $s(t)$ being estimated is smooth then $\hat{\lambda}_{GML}$ under-smooths relative to $\hat{\lambda}_{GCV}$. Also the predictive mean square error using the GML estimate goes to zero at a slower rate than the mean square error using the GCV estimate. She also found that if the true function is "rough" then $\hat{\lambda}_{GML}$ and $\hat{\lambda}_{GCV}$ have asymptotically similar behaviour.

Wahba's (1985) generalized maximum likelihood is different to our method to be discussed in Section 6.2. Her method is based on the Bayesian smoothing spline estimate.

Kohn, Ansley and Tharm (1991) in their simulation study compared the finite-sample performance of generalized cross-validation, cross-validation, and marginal likelihood estimators of the smoothing parameter λ. They obtained the marginal likelihood estimate of the smoothing parameter λ by minimising the likelihood function

$$ML(\lambda) = \prod_{i=m+1}^{n} \left\{ \frac{\omega^{(0)}(i)^2}{R^{(0)}(i)} \right\} T(\lambda)^{1/n-m} \tag{6.10}$$

using the modified Kalman filter. Their derivation of (6.10) is based on the work of Kohn and Ansley (1987). They also defined a fourth estimator, which they called the true squared error (TSE) estimator, which minimises $\sum_{i=1}^{n} \left\{ s(t_i) - \hat{s}(t_i;\lambda) \right\}^2$ over all λ. Note that the TSE estimator is not strictly an estimator, because the function $s(t)$ is unknown. However, the TSE estimator provides some information on the performance of the other three estimators, namely $\hat{\lambda}_{CV}$, $\hat{\lambda}_{GCV}$ and $\hat{\lambda}_{GML}$. We are going

to use the TSE estimator in Section 6.3 as a benchmark for the other estimators. The same authors found that quintic splines ($m = 3$) based on $\hat{\lambda}_{GML}$ outperformed cubic splines. Their $\hat{\lambda}_{CV}$ results for equally spaced data were very similar to those obtained using $\hat{\lambda}_{GCV}$, while for unequally spaced data, $\hat{\lambda}_{GCV}$ outperformed $\hat{\lambda}_{CV}$. They confirmed the theoretical results of Craven and Wahba (1979).

Marginal likelihood estimation of the smoothing parameter λ by Kohn, Ansley and Tharm (1991) is different from our method to be discussed in Section 6.3 Their density function depends on λ and σ^2. The problem of estimation in the presence of a nuisance parameter (in this case σ^2) remains. Our marginal likelihood function is independent of σ^2, see Section 4.2, and so estimation in our case is independent of all nuisance parameters.

6.3 Our method – Marginal likelihood method

Let us assume the *cubic* smoothing spline, $m = 2$. In Section 3.5 we modified Wahba's (1978) model by adding some covariates to (6.1); refer to expression (3.13). In Chapter 4 we constructed the marginal likelihood function $l(\kappa|\boldsymbol{y})$ by (4.16). In this Chapter we estimate κ and consequently $\lambda = 1/\kappa$ by maximizing (4.16) or (4.22) after some simplification. Now we do not need to consider extra regressors in (4.15), so (4.16) reduces to the form (4.17), which can be simplified as

$$l(\kappa|y) = |C|^{-1}\left(\sum_{i=1}^{n}\hat{\omega}_i^2\right)^{-N/2},$$

so that the marginal log-likelihood is given by

$$\log l(\kappa|y) = -\log|C| - \frac{N}{2}\log\left(\sum_{i=1}^{n}\hat{\omega}_i^2\right) \tag{6.11}$$

and λ is estimated by minimising (6.11). The smoothing constant λ is obtained by maximising the marginal likelihood function, hence the notation MML is used for our method hereafter. The *S-Plus* code for this task is given by *Figure 6.1.* Our MML method of the smoothing parameter λ estimation is simple to compute and easy to implement.

In the next section we compare the three estimation methods applied to our set of ten nonlinear test functions given by (4.25), with various σ values. We also evaluate the accuracy of estimating the unknown test functions .

```
mml<-function(y,tt)
{
        n<-length (y)
        ones <- seq(1,1,l=n)
        Z <- cbind(ones,tt)
        halfN <- 0.5*n

        ### Set up Sigma matrix
        h3 <-1/(n^3)
        Sigma <- matrix(0,nrow=n,ncol=n)
        for (i in 1:n)
                Sigma[i,i:n] <- Sigma[i:n,i] <- (i*i*(i+1.5*(0:(n-i))))/3*h3

        ### Functions to compute likelihood
        logdet <- function(x)
                sum(log(eigen(x)$values))
        fl<-function(ll)

        {
                C <- chol(ZZ + exp(ll)*Sigma)
                Cinv <- t(solve(C))
                Ystar <- Cinv %*% Y
                return(logdet(C) + halfN*log(sum(Ystar*Ystar)))
        }

        cc <- 1000
        ZZ <- cc*(Z %*% t(Z)) + diag(n)

        Y<-matrix(y,ncol=1)
        assign("Y",Y,frame=1)
        fl0 <- fl(-1e20) # Value at zero
        fredall <- nlmin(fl,1) # Minimize starting near zero
        fl.fred <- fl(fredall$x) # Best fl
        if(fl0 < fl.fred)
                lstar <- 0.000000001
        else
        {

                lstar <- exp(fredall$x)
        }
        return(1/lstar)
}
```

Figure 6.1: *S-Plus code for λ estimation by our MML method and square root of ISE*

6.4 Simulation study

6.4.1 Procedure

We performed a simulation study in order to compare the different methods of estimating the smoothing parameter λ. We considered three smoothing parameter estimates, namely the generalised cross-validation estimator $\hat{\lambda}_{GCV}$, the maximum marginal likelihood estimator $\hat{\lambda}_{MML}$ (our method) and the true squared error "estimator" $\hat{\lambda}_{TSE}$. Although the TSE estimator is not a feasible one because the function $s(t)$ is unknown, it provides the benchmark on the performance of the other two estimators. The true squared error "estimator" is chosen to minimise integrated squared error given full knowledge of the function $s(t)$ to be estimated.

To determine how well all three estimators perform in estimating the unknown function $s(t_i)$, we compute the square root of integrated squared error for each estimator,

$$\sqrt{ISE(\lambda)} = \sqrt{\sum_{i=1}^{n}\left\{s(t_i) - \hat{s}(t_i;\lambda)\right\}^2}. \tag{6.12}$$

The simulation study considered cubic splines ($m = 2$). We used the same set of 10 nonlinear test functions $s(t_i) = f_1(t_i),\ldots,f_{10}(t_i)$, $i = 1,\ldots,n$ as in Chapters 4 and 5,

refer to the expression (4.25) and to their graphical presentation in *Figures 4.3 (a) – (e)*. These were used for the sample size $n = 50$ and the standard deviations $\sigma = 0.1, 0.2, 0.5$ and 1.0. Six hundred replications were performed for each combination of the test function $s(t_i)$, sample size n and standard deviation σ. The simulation program for each estimation method was written in *S-Plus*. For the *S-Plus* codes for the GCV method and the TSE benchmark method see Appendices 6A.1, 6A.2

We used *"smooth.spline"* and *"predict.smooth.spline"* built in *S-Plus* routines for calculating $\hat{s}(t_i;\lambda)$ in (6.14). We also saved programming time on $\hat{\lambda}_{GCV}$ computation by assigning *"spar"* = 0 in *S-Plus*, so that cross-validation is used to automatically select λ.

We present a graphical comparison of our MML method with GCV method in *Figure 6.2*. In *Figure 6.2*, there is a plot of the test function $s(t) = \sin(4t)^2 + 3t^8$ along with corresponding data points being generated by adding an error term so that $y_i = s(t_i) + \varepsilon_i$, where ε_i, $i = 1,\ldots,n$ iid $N(0, 0.5^2)$. We estimate the function $s(t)$ non-parametrically and get the smoothing parameter by the GCV and MML methods.

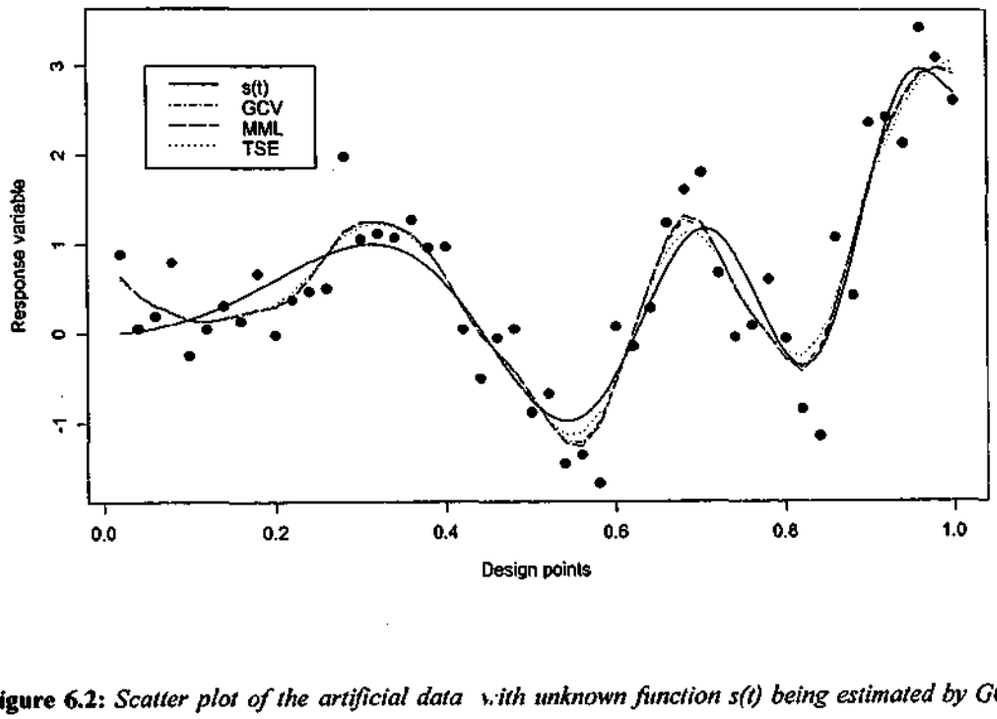

Figure 6.2: *Scatter plot of the artificial data with unknown function s(t) being estimated by GCV, MML and TSE.*

From *Figure 6.2* is evident that all three methods estimate the non-linear function $s(t)$ very well and visually they do not differ very much. The estimates of smoothing constant λ are marginally different, $\hat{\lambda}_{GCV} = 0.0000239956$, $\hat{\lambda}_{MML} = 0.0000181961$ and $\hat{\lambda}_{TSE} = 0.0000467454$.

When evaluating the results of the simulation, we compare the median of $\hat{\lambda}$ estimates rather than using the mean, as the median of $\hat{\lambda}$ is not affected by extreme values.

6.4.2 Results evaluation - the smoothing parameter estimation

A summary of the simulation results of $\hat{\lambda}$ for GCV, MML and TSE are presented in *Table 6.1* for every test function. We present the results of the simulation study by the quartiles and the interquartile range.

Table 6.1 shows the median estimates of the three estimation methods for all ten nonlinear test functions for $\sigma = 0.1, 0.2, 0.5, 1.0$. We arrived at the following conclusions from the simulation study on the smoothing parameter λ estimation.

- The median values of $\hat{\lambda}_{GCV}, \hat{\lambda}_{MML}, \hat{\lambda}_{TSE}$ increase as the value of the standard deviation increases. This was expected because larger standard deviations enlarge the spread of the data, making it difficult to detect the nonlinearity in $s(t)$. The maximum values were effectively infinite (denoted by ">100"). This means that the same functions will often be estimated by a smoothing spline equivalent to the linear function. This represents the limiting case of over-smoothing.

- The rate that λ increases as σ increases is different for different test functions. This is due to the different shapes of the nonlinear test functions. For example, function $f_6(t)$ is less wiggly compared to $f_{10}(t)$, refer to *Figure 4.3 (c)* and *4.3 (e)*. Thus the estimates of λ increase more rapidly with σ for $f_{10}(t)$ than for $f_6(t)$.

- The median value of the smoothing parameter λ is very different for different nonlinear functions for constant σ. Assuming $\sigma = 0.1$, the median for the GCV method ranges from 0.000002 to a maximum value of 0.039900. For the MML method, the minimum median value is 0.000026 and the maximum 0.045680. The minimum for the TSE benchmark method is 0.000002 and the maximum is 0.145000. Such a big difference in results is again due to the different degrees of non-linearity.

- The GCV method returns the highest $\hat{\lambda}$ for all values of the standard deviation for functions f_2, f_3, f_6. For these functions, the GCV method might have a tendency to over-smooth data compared to the MML method. The MML method returns the highest $\hat{\lambda}$ for all values of the standard deviation for functions $f_1, f_4, f_5, f_7, f_8, f_9$ and f_{10}. For these functions the MML method might have a tendency to over-smooth data compared to the GCV method.

		Sigma			
	Method	0.1	0.2	0.5	1
f1	GCV	0.0399	0.6179	>100	>100
	MML	0.0457	>100	>100	>100
	TSE	0.0145	0.0846	0.4065	0.9846
f2	GCV	0.0127	0.0302	0.3726	>100
	MML	0.0062	0.0216	0.2190	>100
	TSE	0.0098	0.0193	0.0617	0.1880
f3	GCV	0.0019	0.0069	0.0545	>100
	MML	0.0014	0.0058	0.0494	>100
	TSE	0.0014	0.0046	0.0230	0.0762
f4	GCV	0.0000	0.0001	0.0108	0.7311
	MML	0.0000	0.0004	0.0348	>100
	TSE	0.0000	0.0006	0.0017	0.0468
f5	GCV	0.0000	0.0005	0.0561	>100
	MML	0.0005	0.0039	0.0712	>100
	TSE	0.0000	0.0002	0.0143	0.0847
f6	GCV	0.0005	0.0012	0.0092	0.0328
	MML	0.0003	0.0009	0.0061	0.0243
	TSE	0.0005	0.0010	0.0048	0.0199
f7	GCV	0.0001	0.0001	0.0004	0.0197
	MML	0.0000	0.0001	0.0015	>100
	TSE	0.0001	0.0001	0.0004	0.0018
f8	GCV	0.0004	0.0010	0.0120	0.0687
	MML	0.0026	0.0010	0.0121	0.0708
	TSE	0.0004	0.0008	0.0039	0.0347
f9	GCV	0.0003	0.0007	0.0501	>100
	MML	0.0002	0.0007	0.1956	>100
	TSE	0.0003	0.0006	0.0028	0.1122
f10	GCV	0.0000	0.0001	0.1312	>100
	MML	0.0012	0.0072	0.9867	>100
	TSE	0.0000	0.0001	0.0182	0.1801

Table 6.1: *Median λ estimates by GCV, MML and TSE for the ten nonlinear test functions*

From *Table 6.1* it is evident that there are not big differences in estimates of the smoothing parameter λ for some test functions. However for some other test functions, such as $f_2(t), f_9(t)$ for $\sigma = 0.5$ and $f_7(t)$ for $\sigma = 1.0$ the differences in $\hat{\lambda}$ are large. For example, we refer to *Figure 6.3* for a visual difference in estimating

test function $f_9(t)$ if $\hat{\lambda}_{GCV} = 0.050060$, $\hat{\lambda}_{MML} = 0.195600$ and $\hat{\lambda}_{TSE} = 0.002765$ for $\sigma = 0.5$.

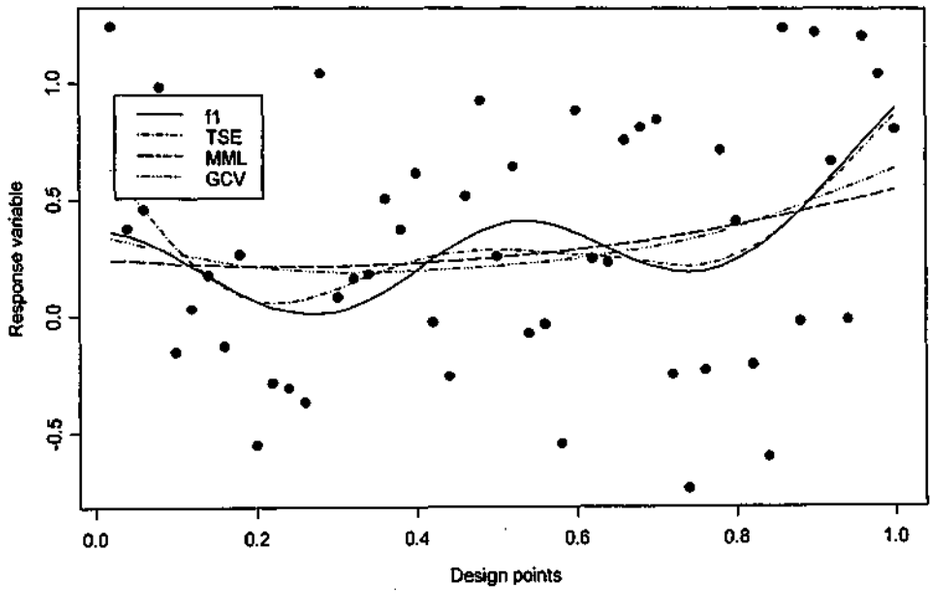

Figure 6.3: *Scatter plot of the artificial data and test function $f_9(t)$ being estimated by GCV, MML and TSE for $\sigma = 0.5$*

From *Figure 6.3* it is evident that the GCV and MML methods both have a tendency to over-smooth the data. Their graphs are flatter compared to that for TSE. GCV picked up the curvature only marginally better than MML. The TSE benchmark method picked up the test function shape better, however there are still big differences between the underlying test function and its estimate.

We evaluated the variation of the smoothing parameter λ estimates by the interquartile range (IQR). A summary of IQR for all three methods of estimation and all test functions with respect to the standard deviation is presented in *Table 6.2.*

Similar to the median λ estimates, results on interquartile range are very different for different estimation methods and test functions. There does not seem to be any visible pattern of dominancy of one method of estimation. Results vary with the degree of nonlinearity of the test function. The interquartile range for the MML method was smaller than for GCV in 14 out of 24 cases. We can therefore conclude that method MML returned slightly less variable estimates of smoothing constant λ compared to GCV.

6.4.3 Results evaluation - the unknown function estimation

In this section we focus on the accuracy of the estimation method of the unknown nonlinear function $s(t)$. To determine how well the GCV, MML and TSE methods perform to estimate the test functions, we computed the square root of integrated squared error of the differences between $s(t_i)$ and $\hat{s}(t_i;\lambda)$ as the performance criterion, refer to (6.12). We collected quartiles, medians and computed interquarile range of $\sqrt{ISE}$, assuming the three estimation methods, $\sigma = 0.1, 0.2, 0.5, 1.0$. and the same set of ten nonlinear test functions as in Section 6.4.1 We present results from the computation for every test function in *Tables 6.3 (a) – (j).*

	Method	Sigma 0.1	0.2	0.5	1
f1	GCV	0.2721	>100	>100	>100
	MML	1.3514	>100	>100	>100
	TSE	0.0418	0.1986	56.9813	118.9073
f2	GCV	0.0111	0.0395	>100	>100
	MML	0.0035	0.0260	>100	>100
	TSE	0.0114	0.0226	0.1094	96.4840
f3	GCV	0.0025	0.0126	0.3271	>100
	MML	0.0009	0.0055	0.3939	>100
	TSE	0.0017	0.0071	0.0342	0.1686
f4	GCV	0.0000	0.0001	0.1045	>100
	MML	0.0000	0.0013	0.7920	>100
	TSE	0.0000	0.0000	0.0102	0.1789
f5	GCV	0.0001	0.0020	1.4046	>100
	MML	0.0004	0.0065	>100	>100
	TSE	0.0000	0.0004	0.0439	0.2239
f6	GCV	0.0003	0.0012	0.0144	0.0569
	MML	0.0001	0.0006	0.0050	0.0426
	TSE	0.0003	0.0009	0.0133	0.0307
f7	GCV	0.0000	0.0001	0.0004	>100
	MML	0.0000	0.0000	>100	>100
	TSE	0.0000	0.0002	0.0003	0.3648
f8	GCV	0.0003	0.0010	0.0280	2.5800
	MML	0.0001	0.0009	0.0173	>100
	TSE	0.0002	0.0007	0.0128	0.0645
f9	GCV	0.0002	0.0006	>100	>100
	MML	0.0001	0.0007	>100	>100
	TSE	0.0000	0.0000	0.0251	72.5851
f10	GCV	0.0000	0.0086	>100	>100
	MML	0.0020	0.0165	>100	>100
	TSE	0.0000	0.0002	0.1073	157.7612

Table 6.2: *Interquartile range of λ estimates by GCV, MML and TSE methods for all the test function*

Figures 6.4 (a) – (j) represent a graphical summary of the interquartile range of $\sqrt{ISE}$ for each method and the test function. Note that TSE method is the benchmark, and represents the best possible achievable results. It is a level of accuracy we would like to achieve by other estimation procedures. Each graph shows the quartiles of $\sqrt{ISE}$ obtained for each estimation method.

	Method	Sigma			
		0.1	0.2	0.5	1.0
Q1	GCV	0.1546	0.2412	0.4785	0.9004
	MML	0.1533	0.3146	0.5996	0.9908
	TSE	0.1206	0.1977	0.4013	0.7554
Median	GCV	0.1971	0.3160	0.7001	1.3600
	MML	0.2028	0.5548	0.9092	1.5910
	TSE	0.1578	0.2679	0.6001	1.1780
Q3	GCV	0.2401	0.4241	0.9987	1.9820
	MML	0.3513	0.8546	1.3080	2.1850
	TSE	0.1974	0.3541	0.8334	1.6530
IQR	GCV	0.0855	0.1829	0.5202	1.0816
	MML	0.1980	0.5400	0.7084	1.1942
	TSE	0.0768	0.1564	0.4321	0.8976

Table 6.3 (a): *Simulated results of $\sqrt{ISE}$ estimated by GCV, MML and TSE for the test function f_1*

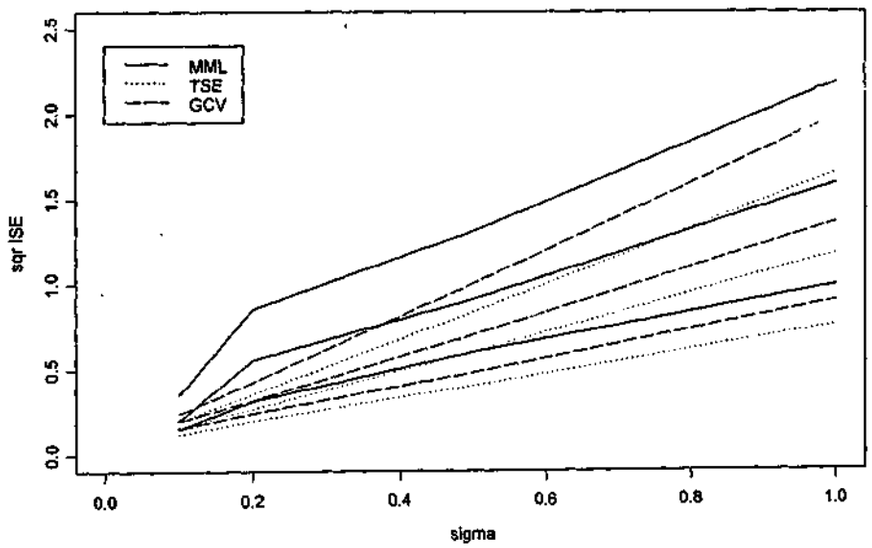

Figure 6.4 (a): *Quartiles of $\sqrt{ISE}$ for GCV, MML, TSE with the test function f_1*

	Method	Sigma 0.1	0.2	0.5	1
Q1	GCV	0.1492	0.2784	0.6231	1.0130
	MML	0.1417	0.2672	0.6686	1.0730
	TSE	0.1226	0.2236	0.4993	0.8456
Median	GCV	0.1935	0.3776	0.7926	1.4240
	MML	0.1821	0.3526	0.9029	1.5730
	TSE	0.1606	0.2901	0.6458	1.2240
Q3	GCV	0.2442	0.4746	1.0560	2.0100
	MML	0.2272	0.4505	1.2660	2.1620
	TSE	0.2040	0.3881	0.8853	1.6640
IQR	GCV	0.0950	0.1962	0.4329	0.9970
	MML	0.0855	0.1833	0.5974	1.0890
	TSE	0.0814	0.1645	0.3860	0.8184

Table 6.3 (b): *Simulated results of* $\sqrt{ISE}$ *estimated by GCV, MML and TSE for the test function* f_2

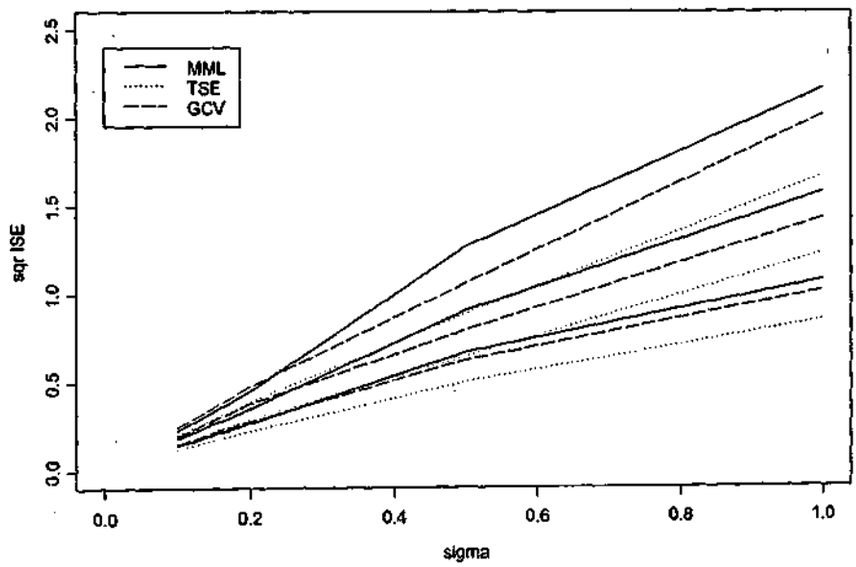

Figure 6.4 (b): *Quartiles of* $\sqrt{ISE}$ *for GCV, MML, TSE with the test function* f_2

	Method	Sigma 0.1	0.2	0.5	1
Q1	GCV	0.2001	0.3456	0.7397	1.2140
	MML	0.1892	0.3321	0.7141	1.2470
	TSE	0.1735	0.2957	0.5882	1.0220
Median	GCV	0.2445	0.4328	0.9725	1.5460
	MML	0.2266	0.4066	0.9509	1.6370
	TSE	0.2135	0.3700	0.7540	1.2910
Q3	GCV	0.2956	0.5323	1.1780	2.0970
	MML	0.2718	0.4912	1.1620	2.1900
	TSE	0.2540	0.4483	0.9728	1.7550
IQR	GCV	0.0955	0.1867	0.4383	0.8830
	MML	0.0826	0.1591	0.4479	0.9430
	TSE	0.0805	0.1526	0.3846	0.7330

Table 6.3 (c): *Simulated results of* $\sqrt{ISE}$ *estimated by GCV, MML and TSE for the test function* f_3

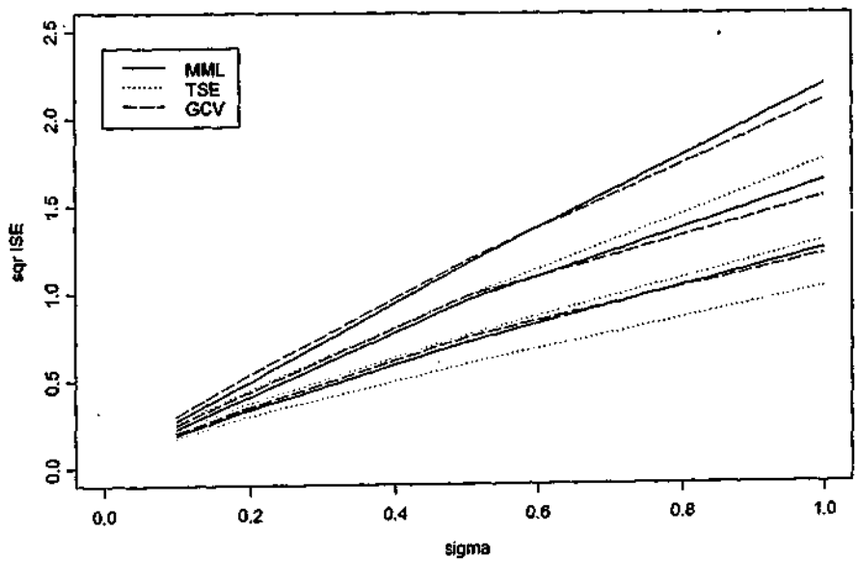

Figure 6.4 (c): *Quartiles on IQR of* $\sqrt{ISE}$ *for GCV, MML, TSE with the test function* f_3

	Method	Sigma			
		0.1	0.2	0.5	1
Q1	GCV	0.3198	0.5907	1.2220	1.6610
	MML	0.3151	0.6428	1.2280	1.6610
	TSE	0.3028	0.5415	1.0630	1.4610
Median	GCV	0.3608	0.6684	1.4230	1.9180
	MML	0.3580	0.7707	1.4130	1.9700
	TSE	0.3430	0.6181	1.1960	1.6750
Q3	GCV	0.4098	0.7717	1.5920	2.4000
	MML	0.4088	0.9024	1.5940	2.4180
	TSE	0.3870	0.6962	1.3470	2.0480
IQR	GCV	0.0900	0.1810	0.3700	0.7390
	MML	0.0937	0.2596	0.3660	0.7570
	TSE	0.0842	0.1547	0.2840	0.5870

Table 6.3 (d): *Simulated results of* $\sqrt{ISE}$ *estimated by GCV, MML, TSE for the test function* f_4

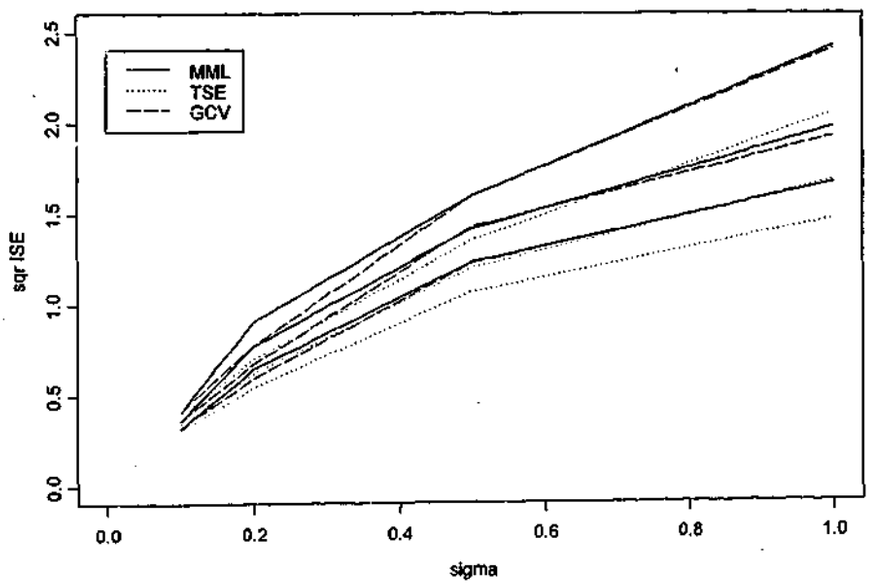

Figure 6.4 (d): *Quartiles of* $\sqrt{ISE}$ *for GCV, MML, TSE with the test function* f_4

	Method	Sigma 0.1	0.2	0.5	1
Q1	GCV	0.3731	0.5896	0.9958	1.3540
	MML	0.4342	0.6260	0.9723	1.3630
	TSE	0.3450	0.5314	0.8356	1.1640
Median	GCV	0.4108	0.6703	1.1560	1.6750
	MML	0.4737	0.6942	1.1310	1.7270
	TSE	0.3805	0.5907	0.9671	1.4390
Q3	GCV	0.4530	0.7548	1.3410	2.1940
	MML	0.5157	0.7564	1.3200	2.2060
	TSE	0.4124	0.6576	1.1340	1.8340
IQR	GCV	0.0799	0.1652	0.3452	0.8400
	MML	0.0815	0.1304	0.3477	0.8430
	TSE	0.0674	0.1262	0.2984	0.6700

Table 6.3 (e): *Simulated results on* $\sqrt{ISE}$ *by GCV, MML and TSE for the test function* f_5

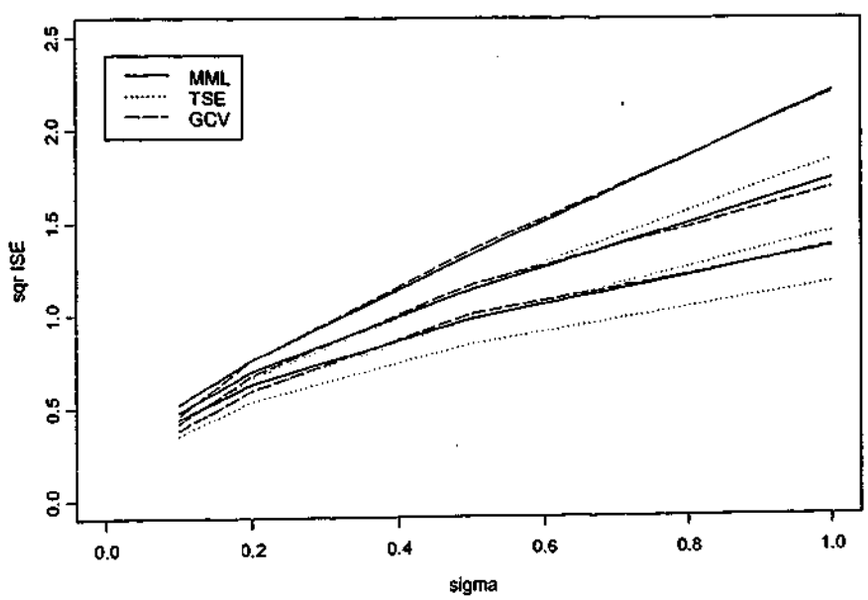

Figure 6.4 (e): *Quartiles of* $\sqrt{ISE}$ *for GCV, MML, TSE with the test function* f_5

	Method	Sigma 0.1	0.2	0.5	1
Q1	GCV	0.2073	0.3880	0.8270	1.4170
	MML	0.2077	0.3755	0.8025	1.4020
	TSE	0.1868	0.3418	0.7227	1.1630
Median	GCV	0.2560	0.4872	1.0390	1.8890
	MML	0.2488	0.4533	0.9869	1.8150
	TSE	0.2317	0.4264	0.9198	1.5180
Q3	GCV	0.3086	0.6005	1.2880	2.4340
	MML	0.2903	0.5381	1.1940	2.2780
	TSE	0.2729	0.5096	1.0830	1.9670
IQR	GCV	0.1013	0.2125	0.4610	1.0170
	MML	0.0826	0.1626	0.3915	0.8760
	TSE	0.0861	0.1678	0.3603	0.8040

Table 6.3 (f): *Simulated results of* $\sqrt{ISE}$ *by GCV, MML, TSE for the test function* f_6

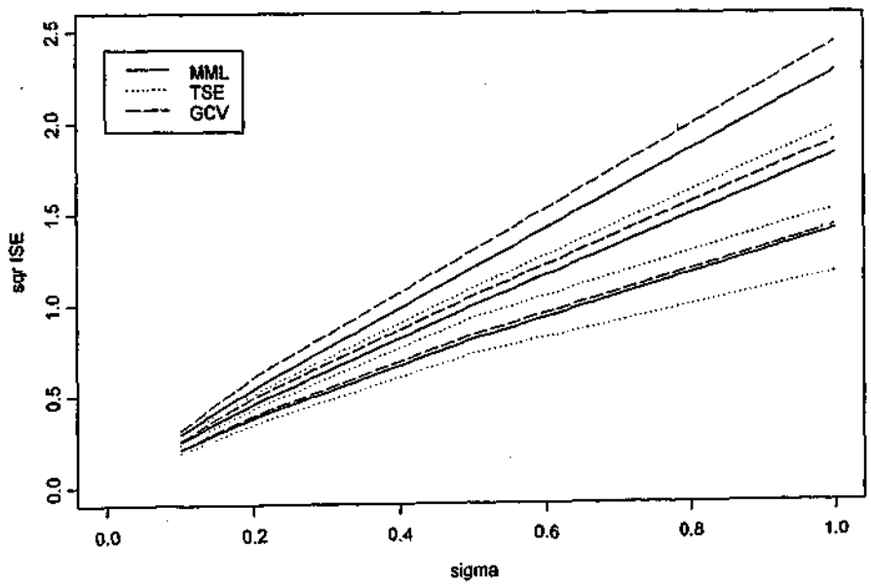

Figure 6.4 (f): *Quartiles of* $\sqrt{ISE}$ *for GCV, MML , TSE with the test function* f_6

	Method	Sigma 0.1	0.2	0.5	1
Q1	GCV	0.2728	0.5007	1.1390	2.3550
	MML	0.2876	0.5018	1.2860	2.4500
	TSE	0.2565	0.4687	1.0360	1.8690
Median	GCV	0.3158	0.5853	1.3540	2.5670
	MML	0.3241	0.5738	1.7310	2.6560
	TSE	0.2937	0.5413	1.2100	2.2240
Q3	GCV	0.3654	0.6931	1.6570	2.9430
	MML	0.3630	0.6596	2.3820	3.0050
	TSE	0.3405	0.6342	1.4370	2.5180
IQR	GCV	0.0926	0.1924	0.5180	0.5880
	MML	0.0754	0.1578	1.0960	0.5550
	TSE	0.0840	0.1655	0.4010	0.6490

Table 6.3 (g): *Simulated results of* $\sqrt{ISE}$ *by GCV, MML, TSE for the test function* f_7

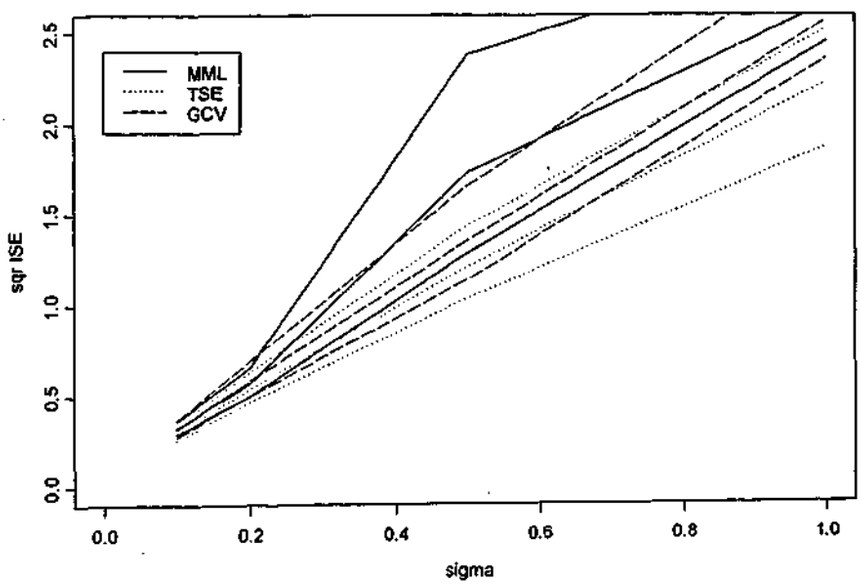

Figure 6.4 (g): *Quartiles of* $\sqrt{ISE}$ *for GCV, MML, TSE with the test function* f_7

	Method	Sigma			
		0.1	0.2	0.5	1
Q1	GCV	0.2124	0.3945	0.8696	1.4910
	MML	0.2130	0.3856	0.8518	1.4510
	TSE	0.1938	0.3484	0.7531	1.1750
Median	GCV	0.2600	0.4949	1.0710	1.8430
	MML	0.2535	0.4674	1.0320	1.8260
	TSE	0.2350	0.4345	0.9130	1.5130
Q3	GCV	0.3137	0.6053	1.3350	2.3030
	MML	0.2944	0.5625	1.2480	2.2750
	TSE	0.2766	0.5168	1.0950	1.9000
IQR	GCV	0.1013	0.2108	0.4654	0.8120
	MML	0.0814	0.1769	0.3962	0.8240
	TSE	0.0828	0.1684	0.3419	0.7250

Table 6.3 (h): *Simulated results of* $\sqrt{ISE}$ *by GCV, MML, TSE for the test function* f_8

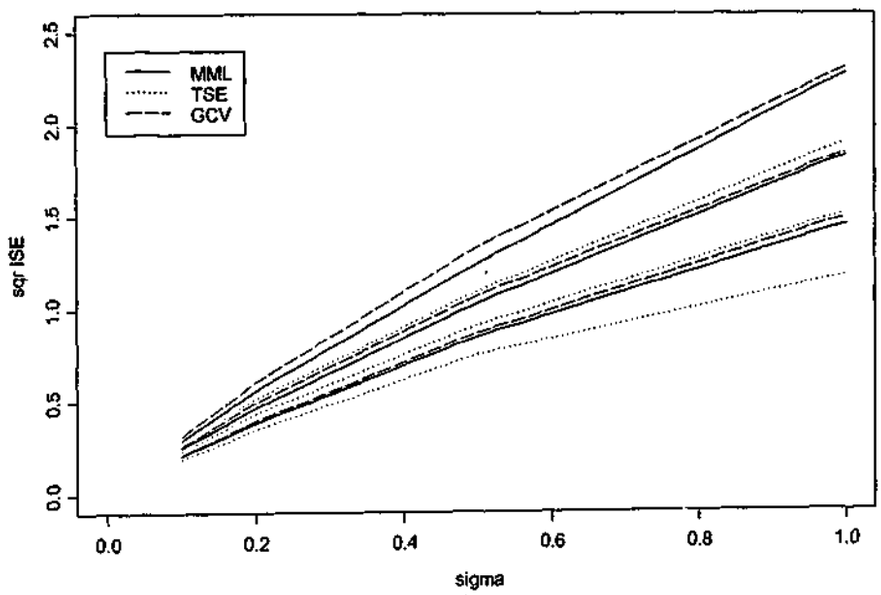

Figure 6.4 (h): *Quartiles of* $\sqrt{ISE}$ *for GCV, MML, TSE with the test function* f_8

	Method	Sigma 0.1	0.2	0.5	1
Q1	GCV	0.2262	0.4146	1.0690	1.4010
	MML	0.2233	0.4056	1.0840	1.4330
	TSE	0.2067	0.3686	0.8113	1.2230
Median	GCV	0.2691	0.5037	1.1990	1.7280
	MML	0.2614	0.4874	1.2220	1.7840
	TSE	0.2464	0.4522	0.9877	1.4900
Q3	GCV	0.3198	0.6126	1.3950	2.2650
	MML	0.3055	0.5860	1.4440	2.2690
	TSE	0.2911	0.5375	1.1530	1.8680
IQR	GCV	0.0936	0.1980	0.3260	0.8640
	MML	0.0822	0.1804	0.3600	0.8360
	TSE	0.0844	0.1689	0.3417	0.6450

Table 6.3 (i): *Simulates results of* $\sqrt{ISE}$ *by GCV, MML, TSE for the test function* f_9

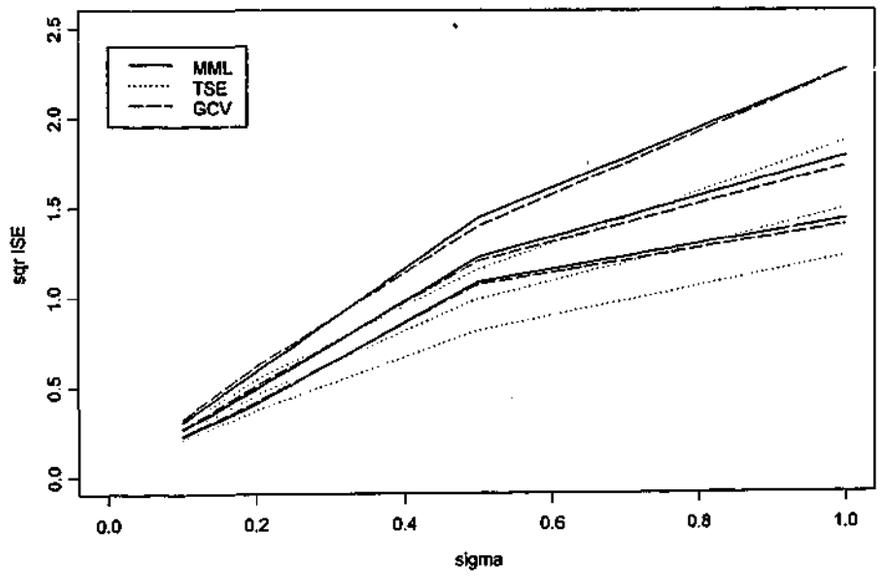

Figure 6.4 (i): *Quartiles of* $\sqrt{ISE}$ *for GCV, MML, TSE with the test function* f_9

	Method	Sigma 0.1	0.2	0.5	1
Q1	GCV	0.4512	0.7997	1.1100	1.3860
	MML	0.6967	0.8352	1.1210	1.4290
	TSE	0.4516	0.7050	0.9875	1.2470
Median	GCV	0.4986	0.8621	1.2190	1.7060
	MML	0.7507	0.8871	1.2540	1.7850
	TSE	0.5083	0.7617	1.0940	1.5130
Q3	GCV	0.5654	0.9461	1.4040	2.2270
	MML	0.7860	0.9567	1.4580	2.2710
	TSE	0.5918	0.8232	1.2430	1.9090
IQR	GCV	0.1142	0.1464	0.2940	0.8410
	MML	0.0893	0.1215	0.3370	0.8420
	TSE	0.1402	0.1182	0.2555	0.6620

Table 6.3 (j): *Simulated results of* $\sqrt{ISE}$ *by GCV, MML, TSE for the test function* f_{10}

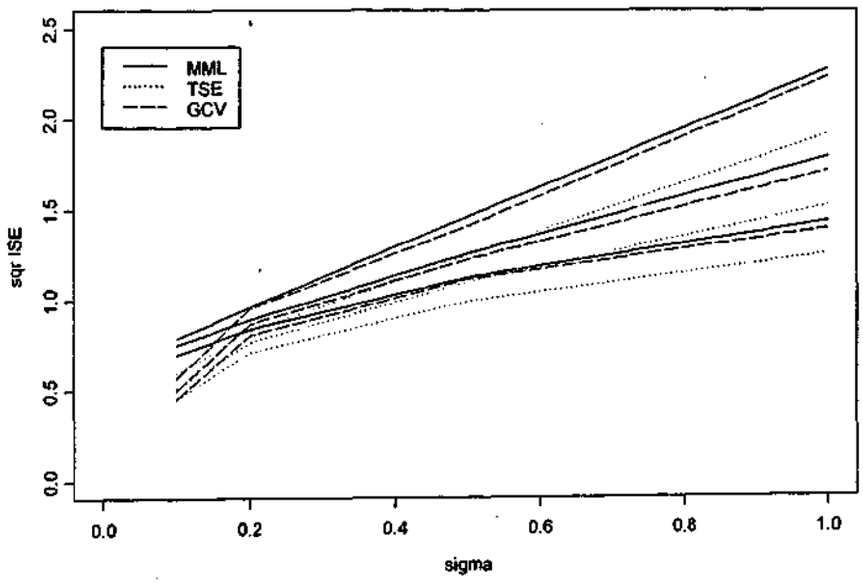

Figure 6.4 (j): *Quartiles of* $\sqrt{ISE}$ *for GCV, MML, TSE with the test function* f_{10}

Concluding from *Tables 6.3 (a) – (j)* and *Figures 6.4 (a) – (j)*, the MML method outperformed the GCV method for functions f_6, f_8. Both methods performed about equally well for functions f_3, f_4, f_5, f_9. The latter method outperformed the former for functions f_1, f_2, f_7, f_{10}.

We constructed boxplots of the ratio of medians and boxplots of the ratio of IQRs of $\sqrt{ISE}$ to compare simulated results from the three methods. We use the following notation. For a given combination of σ and test function let $M_{GCV,MML}$ be the ratio of medians

$$M_{GCV,MML} = \frac{median\sqrt{ISE\left(\hat{\lambda}_{GCV}\right)}}{median\sqrt{ISE\left(\hat{\lambda}_{MML}\right)}}.$$

Define the ratio of medians $M_{GCV,TSE}$ and $M_{MML,TSE}$ similarly. *Figure 6.5* displays the results for sample size $n = 50$ using all 40 combinations of σ and test function.

In *Figure 6.5*, the first plot on the left is a boxplot of $M_{GCV,MML}$, the second plot is a boxplot of $M_{GCV,TSE}$ and the third plot is a boxplot of $M_{MML,TSE}$. Comparing the second and the third boxplot, the average performance of GCV and MML relative to TSE is about the same. The third boxplot has a larger variance compared to the second one indicating that the MML method is less reliable in estimating the unknown function than the GCV method relative to the TSE method. In particular, there are two outlying observations in the third boxplot.

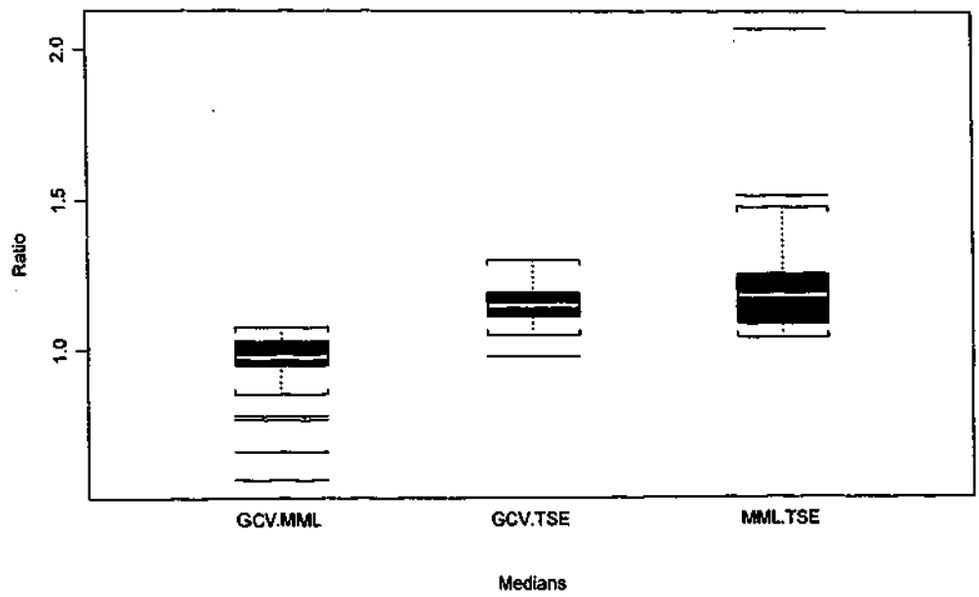

Figure 6.5: *Boxplots comparing the median value of* $\sqrt{ISE}$ *by GCV, MML, TSE*

The first boxpot shows the median value being around unity. This means that on average GCV and MML performed equally well. However the upper portion of the box is larger than the lower which indicates the smaller variance of MML estimates. There are few outliers in the lower portion of the first boxplot indicating that some medians of $ISE\left(\hat{\lambda}_{MML}\right)^{1/2} > ISE\left(\hat{\lambda}_{GCV}\right)^{1/2}$.

We used a similar notation for the analysis of a variance expressed by IQR of $\sqrt{ISE}$. For a given combination of σ and test function, let $IQR_{GCV,MML}$ be the ratio of IQRs

$$IQR_{GCV,MML} = \frac{IQR\sqrt{ISE\left(\hat{\lambda}_{GCV}\right)}}{IQR\sqrt{ISE\left(\hat{\lambda}_{MML}\right)}}.$$

Define the ratios of $IQR_{GCV,TSE}$ and $IQR_{MML,TSE}$ similarly. *Figure 6.6* displays the results for sample size $n = 50$ using all 40 combinations of σ and test function.

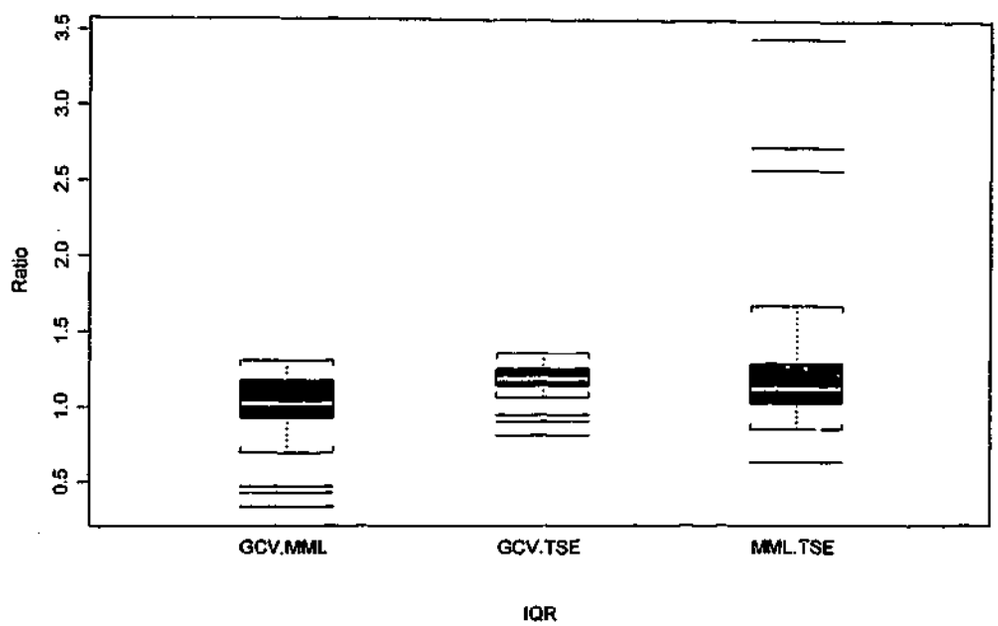

Figure 6.6: *Boxplots comparing the interquartile range of $\sqrt{ISE}$ by GCV, MML, TSE*

The first plot on the left is a boxplot of $IQR_{GCV,MML}$, the second plot is a boxplot of $IQR_{GCV,TSE}$ and the third plot is a boxplot of $IQR_{MML,TSE}$. Comparing the second and the third boxplot, the average performance MML relative to TSE is better than GCV relative to TSE because the median value in the third boxplot is closer to unity than on

the second one. The third boxplot has a larger variance compared to the second one indicating that the MML method is less reliable in estimating the unknown function than the GCV method relative to the TSE method. The first boxpot is showing the mean value is around unity. This means that on average, GCV and MML performed equally well. However the upper portion of the box is larger than the lower one which indicates the smaller variance of MML estimates. There are few outliers in the lower portion of the first boxplot indicating that there are some medians of $ISE\left(\hat{\lambda}_{MML}\right)^{1/2} > ISE\left(\hat{\lambda}_{GCV}\right)^{1/2}$.

6.5 Marginal likelihood methods comparison

We also need to compare our MML method of estimating the smoothing parameter λ given by expression (6.11) with the marginal likelihood estimator used by Kohn et al. (1991, hereafter referred to as the KAT method), presented by expressions (6.10) in Section 6.2.2.

To get a valid comparison between their finding and our method, we partly replicated their simulation study and estimated λ by GCV and MML. We considered their set of ten nonlinear test functions (6.13), which is different to our set (4.11) that was used for all simulation studies in this thesis so far. We assumed only splines of order $m = 2$, sample size $n = 40$, and five values of σ $(\sigma = 0.01, 0.05, 0.1, 0.2, 0.5)$. One hundred replications were performed for each combination of test function and value of σ. The residuals were normally distributed. We collected medians and third

quartiles of $\sqrt{ISE}$ and the results are presented in *Table 6.4*. The new set of test functions (hereafter referred as KAT functions) is

$$
\begin{aligned}
f_{01}(t) &= \left\{\beta_{10,5}(t) + \beta_{5,10}(t) + \beta_{7,7}(t)\right\}/3 \\
f_{02}(t) &= 0.6\beta_{30,17}(t) + 0.4\beta_{3,11}(t) \\
f_{03}(t) &= \left\{\beta_{20,5}(t) + \beta_{12,12}(t) + \beta_{7,30}(t)\right\}/3 \\
f_{04}(t) &= 4.26e^{-3.25t} + 4.0e^{-6.5t} + 3.0e^{-9.75t} \\
f_{05}(t) &= \left\{2.0 - 5t + e^{-(t-0.5)^2/0.04}\right\}/3.77 \\
f_{06}(t) &= 5e^{-5t} \\
f_{07}(t) &= \left[1 + e^{-7(t-0.5)}\right]^{-1} \\
f_{08}(t) &= e^{-3e^{-3t}} \\
f_{09}(t) &= t \\
f_{010}(t) &= t^2,
\end{aligned}
\tag{6.13}
$$

where $\beta_{p,q}(t) = \dfrac{\Gamma(p+q)}{\Gamma(p)+\Gamma(q)} t^{p-1}(1-t)^{q-1}$ is the beta function.

We get the smoothing parameter λ estimate and the square root of integrated squared error given by (6.12) for both the GCV and MML estimation approaches. We refer to Appendix 6B for *S-Plus* codes for both GCV and MML estimation procedures.

Function	Method	Sigma	0.01	0.05	0.1	0.2	0.5
		median	0.0389	0.1596	0.2957	0.5104	1.0540
	GCV	Q3	0.0452	0.1936	0.3640	0.6152	1.3560
		median	0.0612	0.2753	0.3029	0.4969	1.1376
f01	MML	Q3	0.0670	0.3144	0.3680	0.5727	1.3480
		median	0.0456	0.1823	0.3333	0.6118	1.4794
	GCV	Q3	0.0503	0.2153	0.3886	0.7318	1.7906
		median	0.0623	0.2481	0.3591	0.6143	1.4558
f02	MML	Q3	0.0672	0.3094	0.4079	0.6863	1.6556
		median	0.0535	0.2159	0.3964	0.6796	1.6158
	GCV	Q3	0.0588	0.2421	0.4471	0.7881	1.7777
		median	0.0551	0.2936	0.5549	0.6868	1.5545
f03	MML	Q3	0.0608	0.3237	0.6392	0.7776	1.6817
		median	0.0373	0.1376	0.2466	0.4609	0.9899
	GCV	Q3	0.0440	0.1657	0.3118	0.5706	1.2515
		median	0.0412	0.1287	0.2529	0.8205	1.0037
f04	MML	Q3	0.0490	0.1555	0.4838	0.8498	1.1647
		median	0.0341	0.1470	0.2727	0.5029	0.1886
	GCV	Q3	0.0397	0.1697	0.3249	0.6159	1.4690
		median	0.0345	0.1798	0.2929	0.4756	1.0471
f05	MML	Q3	0.0421	0.3000	0.3409	0.5569	1.2163
		median	0.0339	0.1305	0.2438	0.4455	0.9930
	GCV	Q3	0.0404	0.1657	0.2974	0.5532	1.2459
		median	0.0362	0.1257	0.2694	0.8367	0.9798
f06	MML	Q3	0.0416	0.1512	0.8338	0.8608	1.1156
		median	0.0269	0.1161	0.2258	0.4467	0.8254
	GCV	Q3	0.0328	0.1475	0.2956	0.5525	1.2031
		median	0.0614	0.2983	0.2852	0.4623	0.9443
f07	MML	Q3	0.0670	0.3254	0.3470	0.5420	1.1765
		median	0.0260	0.1147	0.2082	0.3355	0.7511
	GCV	Q3	0.0316	0.1404	0.2538	0.4745	0.1284
		median	0.0311	0.1435	0.2747	0.4480	0.9016
f08	MML	Q3	0.0359	0.1621	0.3192	0.5355	1.1219
		median	0.0141	0.0705	0.1410	0.2821	0.7052
	GCV	Q3	0.0218	0.1092	0.2186	0.4372	1.0929
		median	0.0313	0.1509	0.2675	0.4635	0.9427
f09	MML	Q3	0.0355	0.1736	0.3254	0.5827	1.2023
		median	0.0258	0.1063	0.1952	0.3838	0.8294
	GCV	Q3	0.0306	0.1328	0.2519	0.4935	1.1856
		median	0.0623	0.1748	0.2624	0.4481	0.9062
f10	MML	Q3	0.0672	0.2926	0.3236	0.5791	1.1483

Table 6.4 : *Medians and third quartiles of $\sqrt{ISE}$ by GCV and MML for KAT test functions*

Figure 6.7 displays the results for the above-mentioned sampling experiment. The first plot on the left is a boxplot of the ratio of medians

$$M_{GCV,MML} = \frac{medain\sqrt{ISE\left(\hat{\lambda}_{GCV}\right)}}{medain\sqrt{ISE\left(\hat{\lambda}_{MML}\right)}}.$$

The second plot on the right is the boxplot of the ratio of the third quartiles

$$Q3_{GCV,MML} = \frac{Q3\sqrt{ISE\left(\hat{\lambda}_{GCV}\right)}}{Q3\sqrt{ISE\left(\hat{\lambda}_{MML}\right)}}.$$

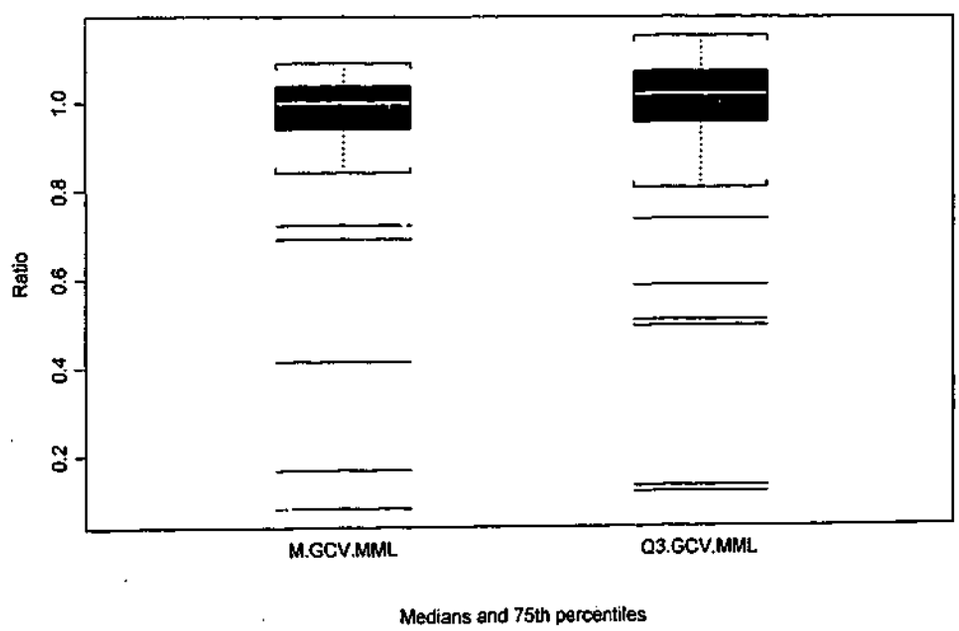

Figure 6.7: *Boxplots comparing the medians and third quartiles of* $\sqrt{ISE\left(\hat{\lambda}_{GCV}\right)}\Big/\sqrt{ISE\left(\hat{\lambda}_{MML}\right)}$ *for KAT test functions*

Now we compare a relevant part of Kohn, Ansley and Tharm (1991) *Figure 2(a)* on page 1046, with our results presented graphically by *Figure 6.7*. The median value on both of our boxplots is around one, which indicate that $ISE\left(\hat{\lambda}_{MML}\right)^{1/2}$ is about the same compared to $ISE\left(\hat{\lambda}_{GCV}\right)^{1/2}$. The upper and the lower portion of both boxplots are about the same. There are few outliers in the lower portion indicating that some $ISE\left(\hat{\lambda}_{MML}\right)^{1/2} > ISE\left(\hat{\lambda}_{GCV}\right)^{1/2}$.

To compare the average performance of the GCV, MML and ML (KAT) methods, the last one is showing the better results as the average is slightly above unity (Kohn, Ansley and Tharm (1991) *Figure 2(a)* on page 1046). This indicates that the KAT marginal likelihood estimate of the smoothing parameter λ slightly outperformed GCV. Our MML method did not outperform the GCV method for the set of KAT test functions. However our MML method is more reliable than the ML (KAT) method, because there are fewer outliers and the variance is smaller in our MML, refer to *Figure 6.7* and the KAT boxplots. Our method is also computational more friendly and easier to implement using conventional statistical software. The KAT marginal likelihood method requires the use of a modified Kalman filter technique, which is computationally more intensive than our method. (Note that a small discrepancy in a comparison of our approach with those by Kohn, et al. (1991) might be expected due to sampling error.)

6.6 Results summary

This chapter reports on an intensive simulation study investigating the statistical properties of spline smoothing estimators when the smoothing parameter λ is estimated by generalised cross-validation, marginal likelihood and true squared error estimation using splines of order 2. The performance criterion for the function estimate was the square root of integrated squared error. Our conclusions apply to equally spaced data and can be summarised as follows:

- Our marginal likelihood method is a competitive alternative to GCV for estimating the smoothing parameter λ.

- Our MML method is computationally more friendly to those being used by Kohn, Ansley and Tharn (1991).

- Our MML method is generally performing well in estimating λ and consequently returns reasonable estimates of the unknown non-linear function $s(t)$.

- Both the MML and GCV estimators performed well relative to the TSE benchmark method.

- The variance of $\hat{\lambda}_{MML}$ is smaller compared to those by GCV, which makes $\hat{\lambda}_{MML}$ less variable.

- Our MML is slightly less accurate compared to the KAT method.

- The variance of $\hat{\lambda}_{MML}$ is also smaller compared to the KAT marginal likelihood method, which makes our estimates less variable.

We recommend using our MML method for estimating the smoothing parameter λ and consequently the unknown function $s(t)$ as an additional method to GCV and others. Perhaps the average of $\hat{\lambda}_{MML}, \hat{\lambda}_{GCV}, \hat{\lambda}_{ML}$ would be a useful combination estimator in non-parametric smoothing applications.

Appendix 6A.1 *S-Plus* code

```
#######################################################
#Smoothing parameter estimation by GCV and square root of ISE#
#######################################################

n<-50
tt<-(1:n)/n

f0 <- tt
f1 <- exp(-3*exp(-3*tt))
f2 <- tt^2
f3 <- exp(-5*(1-tt))
f4 <- 0.71*(sin((4*tt)^2)+3*tt^8)/3+.25
f5 <- (0.909*(1/sin(3*tt))/15)-0.04
f6 <- 1.4*(exp(sin(3*tt)^2)/2.5)-0.55
f7 <-0.5*sin(4*pi*tt)+0.5
f8 <-((5*tt+5*exp(-((tt-0.5)^2)/0.04))/7)*0.909
f9 <- 2.5*((exp(sin(3*tt)^3)+exp(cos(3*tt)^2)+tt^4)/4.5)-1.7
f10 <-((exp(sin((3*tt)^3))+4*tt^4)/6)-0.1

sigma<-0.1  # 0.1,0.2,0.5 ,1.0#
set.seed (1)
nsim<-600
lambda<-numeric(nsim)
rise<-numeric(nsim)
for(i in 1:nsim)
{
        fmean<-f10
        y<-fmean + rnorm (n,0,sigma)
        fit<-smooth.spline(tt,y) #returns lambda hat as "fit$spar"#
        fit2<-predict(fit,tt)  #returns s(t) hat as "fit2$y"#
        rise[i]<-sqrt(sum ((fmean-fit2$y)^2)) #root of integrated squared error#
        lambda[i]<-fit$spar
}

#exploratory data analysis - four plots#
eda.shape<-function(x)
{
        par(mfrow=c(2,2))
        hist(x)
        boxplot(x)
        iqd<-summary(x)[5] - summary(x)[2]
        plot(density(x,width=2*iqd), xlab="x", ylab="", type= 'l')
        qqnorm(x)
        qqline(x)
}
summary(lambda)
summary(rise)
eda.shape(lambda)
eda.shape(rise)
```

Appendix 6A.2 *S-Plus* code

```
#####################################################
#Smoothing parameter estimation by TSE and square root of ISE#
#####################################################

n<- 50
tt<-(1:n)/n

f0 <- tt
f1 <- exp(-3*exp(-3*tt))
f2 <- tt^2
f3 <- exp(-5*(1-tt))
f4 <- 0.71*(sin((4*tt)^2)+3*tt^8)/3+.25
f5 <- (0.909*(1/sin(3*tt))/15)-0.04
f6 <- 1.4*(exp(sin(3*tt)^2)/2.5)-0.55
f7 <-0.5*sin(4*pi*tt)+0.5
f8 <-((5*tt+5*exp(-((tt-0.5)^2)/0.04))/7)*0.909
f9 <- 2.5*((exp(sin(3*tt)^3)+exp(cos(3*tt)^2)+tt^4)/4.5)-1.7
f10 <-((exp(sin((3*tt)^3))+4*tt^4)/6)-0.1

sigma<-1.0 # 0.1, 0.2, 0.5, 1.0 #
set.seed(1)

ise<-function(ll)
{
        lambda<-exp(ll)
        fit1<-smooth.spline(tt,y,spar=lambda)
        fit11<-predict(fit1,tt)
        res<-fmean-fit11$y
        return(sum(res^2))
}

minfl<- function()
{
        cat(".")
        y<- fmean+rnorm(n,0,sigma)
        assign("y",y,frame=1)
        fred<-nlmin(ise,0)
        valise<-ise(fred$x)
        return(c(exp(fred$x),valise))
}

##Simulation
fmean<-f10
nn<- 600
lmat<- matrix(0,nrow=nn,ncol=2)
for (k in 1:10)     lmat[k,] <- minfl()
.
.
.
for (k in 591:600)lmat[k,] <- minfl()
lambda<-lmat[,1]
rise<-sqrt(lmat[,2])

#exploratory data analysis - four plots#
eda.shape<-function(x)
{
        par(mfrow=c(2,2))
        hist(x)
```

```
        boxplot(x)
        iqd<-summary(x)[5] - summary(x)[2]
        plot(density(x,width=2*iqd), xlab="x", ylab="", type= 'l')
        qqnorm(x)
        qqline(x)
}
summary(lambda)
summary(rise)
eda.shape(lambda)
eda.shape(rise)
```

Appendix 6B.1 *S-Plus* code

```
##########################################################################
#Smoothing parameter estimate by GCV and square root of ISE for 10 test functions #
#by  K.A.T. (1991)                                                                #
##########################################################################

n<-40
tt<-(1:n)/n

### Set up test deterministic functions
beta<-function(t,p,q)
{
        gamma(p+q)/gamma(p)/gamma(q)*t^(p-1)*(1-t)^(q-1)
}
f01<-((beta(tt,10,5)+beta(tt,5,10)+beta(tt,7,7))/3)/1.7921549479
f02<-(0.6*(beta(tt,30,17))+0.4*(beta(tt,3,11)))/3.406258
f03<-((beta(tt,20,5)+beta(tt,12,12)+beta(tt,7,30))/3)/2.101791
f04<-((4.2*exp(-3.25*tt)+4.0*exp(-6.5*tt)+3*exp(-9.75*tt))-0.1690403)/9.747529
f05<- (((2-0.5*tt+exp(-(tt-.5)^2/0.04))/3.77)-0.39839)/0.3310662
f06<-(5*exp(-5*tt)-0.03368973)/4.490497
f07<-((1+exp(-7*(tt-0.5)))^(-1)-0.03356922)/0.9371185
f08<-exp(-3*exp(-3*tt))
f09<-tt
f010<-tt^2

sigma<-0.01  # 0.01,0.02,0.1,0.2,0.5 #
set.seed (1)

nsim<-100
lambda<-numeric(nsim)
rise<-numeric(nsim)
for(i in 1:nsim)
{

        fmean<-f01
        y<-fmean + rnorm (n,0,sigma)

        fit<-smooth.spline(tt,y) #returns lambda hat as "fit$spar"#
        fit2<-predict(fit,tt)  #returns s(t) hat as "fit2$y"#
        rise[i]<-sqrt(sum ((fmean-fit2$y)^2)) #root of integrated squared error#
        lambda[i]<-fit$spar
}

median(rise)
quantile(rise,c(.75))
```

Appendix 6B.2 *S-Plus* code

```
##########################################################################
#Smoothing parameter estimation by MML and square root of ISE for 10 test functions #
#by  K.A.T. (1991)                                                                  #
##########################################################################

### Generate x data
n <- 40
set.seed(1)
tt <- (1:n)/n

## Set up equation for y data
sigma <- 0.01 #stdev 0.01,0.05,0.1,0.2,0.5#
ones <- seq(1,1,l=n)
S <- cbind(ones,tt)
halfN <- 0.5*n

### Set up Sigma matrix
h3 <-1/(n^3)
Sigma <- matrix(0,nrow=n,ncol=n)
for (i in 1:n)
        Sigma[i,i:n] <- Sigma[i:n,i] <- (i*i*(i+1.5*(0:(n-i))))/3*h3

### Set up test deterministic functions
beta<-function(t,p,q)
{
        gamma(p+q)/gamma(p)/gamma(q)*t^(p-1)*(1-t)^(q-1)
}
f01<-((beta(tt,10,5)+beta(tt,5,10)+beta(tt,7,7))/3)/1.7921549479
f02<-(0.6*(beta(tt,30,17))+0.4*(beta(tt,3,11)))/3.406258
f03<-((beta(tt,20,5)+beta(tt,12,12)+beta(tt,7,30))/3)/2.101791
f04<-((4.2*exp(-3.25*tt)+4.0*exp(-6.5*tt)+3*exp(-9.75*tt))-0.1690403)/9.747529
f05<- (((2-0.5*tt+exp(-(tt-.5)^2/0.04))/3.77)-0.39839)/0.3310662
f06<-(5*exp(-5*tt)-0.03368973)/4.490497
f07<-((1+exp(-7*(tt-0.5)))^(-1)-0.03356922)/0.9371185
f08<-exp(-3*exp(-3*tt))
f09<-tt
f010<-tt^2


### Functions to compute likelihood

logdet <- function(x)
        sum(log(eigen(x)$values))

fl<-function(ll)
{
        C <- chol(SS + exp(ll)*Sigma)
        Cinv <- t(solve(C))
        Ystar <- Cinv %*% Y
        return(logdet(C) + halfN*log(sum(Ystar*Ystar)))
}

minfl <- function()
{
        cat(".")
        y <- fmean + rnorm(n,0,sigma)
        Y <- matrix(y,ncol=1)
        assign("Y",Y,frame=1)
```

```
        fl0 <- fl(-1e20) # Value at zero
        fredall <- nlmin(fl,1) # Minimize starting near zero
        fl.fred <- fl(fredall$x) # Best fl
        if(fl0 < fl.fred)
                lstar <- 0.000000001
        else
        {

                lstar <- exp(fredall$x)
        }

        fit1<-smooth.spline(tt,y,spar=lstar^(-1))
        fit2<-predict(fit1,tt)
        rise<-sqrt(sum((fmean-fit2$y)^2))

        return(c(lstar,rise))
}

###############
## Simulation. ##
##############

fmean <- f01
cc <- 1000
SS <- cc*(S %*% t(S)) + diag(n)

nn <-100

lmat <- matrix(0,nrow=nn,ncol=2)

for (k in 1:10)     lmat[k,] <- minfl()
for (k in 11:20)    lmat[k,] <- minfl()
for (k in 21:30)    lmat[k,] <- minfl()
for (k in 31:40)    lmat[k,] <- minfl()
for (k in 41:50)    lmat[k,] <- minfl()
for (k in 51:60)    lmat[k,] <- minfl()
for (k in 61:70)    lmat[k,] <- minfl()
for (k in 71:80)    lmat[k,] <- minfl()
for (k in 81:90)    lmat[k,] <- minfl()
for (k in 91:100)   lmat[k,] <- minfl()


lstar<-lmat[,1]
rise<-lmat[,2]
lambda<- lmat[,1]^(-1)

median(rise)
quantile(rise,c(.75))
```

CHAPTER 7

Conclusions

The non-parametric methodology presented in this thesis, namely the cubic smoothing spline approach to nonlinear inference, estimates a semiparametric regression model. Semiparametric regression has attracted researchers' attention for its flexibility in modelling some of the explanatory variables non-parametrically and the rest of the variables by parametric regression. This is very useful in economic applications because the variable of interest is evolving over time and often depends on other explanatory variables. Semiparametric regression is extremely helpful when the time component of the model is nonlinear and difficult to model by the parametric approach. A wide variety of nonlinear trend functions were considered in our simulation studies.

The regression model is central to econometrics and there is a need to be able to test regression coefficients and regression disturbances in an efficient and accurate manner. When working with the semiparametric model, testing for the inclusion of a possibly nonlinear component is an interesting issue. A related issue is the ability to test for linearity of a possibly nonlinear component, because it is an important to know whether and in what form to include the time component in the semiparametric regression. The review of the literature suggests the use of the regression specification error F test which is a mean based test or variance based tests, for example the locally most powerful invariant test, the point optimal invariant test etc. One problem with variance based tests is the cumbersome computation of the critical values.

Many econometric tests are based either on the likelihood ratio, Wald or Lagrange multiplier test principles. The literature review on linking the semiparametric regression with any of these testing principles revealed a need to develop new tests. We have focussed on deriving new likelihood ratio tests based on the marginal likelihood function of the semiparametric model. Both new tests proposed in this thesis are one-sided tests and each of these test statistics follows a weighted mixture of chi-squared distributions asymptotically under the null hypothesis. To be able to obtain these tests, the exact variance-covariance structure of the disturbance vector of the semiparametric model has to be known. One contribution of this thesis was to derive the variance-covariance matrix of the error term and to implement it in further nonlinear inference for time series. Apart from the application of our finding in hypothesis testing, we derived the best linear unbiased predictor and the variance-covariance matrix of this predictor applying the

stochastic model with a non-parametric component. We used our derivation to construct prediction intervals.

Another aim of this thesis was to look at the problem of estimating the smoothing constant in smoothing splines. This is a vital problem in all smoothing approaches in order to get the smoothing procedure robust to smoothing bias. The literature search in this area revealed that there exists two main approaches, namely methods based on penalised regression and methods based on maximum likelihood. The smoothing constant is a by-product of our maximising problem in the new proposed likelihood ratio based tests.

In Chapter 2, we reviewed the literature to understand the issues and problems concerned with spline functions. We focussed on univariate smoothing splines.

The purpose of Chapter 3 was to introduce a cubic smoothing spline and to link it with stochastic processes. The cubic smoothing spline was presented in a state-space form and extra covariates were introduced into the model. This semiparametric model was also presented in a state-space form. The variance-covariance matrix of the disturbance vector was derived by Yule-Walker equations. The best unbiased predictor along with the variance-covariance structure of the error term were derived and prediction intervals were obtained.

We applied our finding to two likelihood ratio based tests in Chapters 4 and 5. Both tests aim to detect nonlinearity in the semiparametric model and are based on maximizing the marginal likelihood function. The MLR test is a test for the inclusion of the possibly nonlinear component into regression model. The MLR* test is a test for linearity of the possible nonlinear component in the regression model. The MLR* test is a logical further step in the sense that inclusion of the possibly nonlinear component can be suggested by the MLR test, hence the MLR* test tests whether the additional component is or is not linear. From our simulation study we conclude that both the MLR and the MLR* tests perform well. Though they were not superior to the competitive tests, they are reliable, easy to perform and computationally non-intensive.

The main purpose of Chapter 6 was to investigate some smoothing constant estimation methods. When working with smoothing spline techniques, the smoothing parameter is usually estimated by the Cross-validation method or as a by-product of optimizing the likelihood function. Its value is crucial to the overall smoothness and the accuracy of the estimated nonlinear function. The study on the bandwidth selection method in Chapter 6 showed that our method was competitive with the most popular alternatives and computationally easy to implement.

In conclusion, we can summarise the original contributions of this thesis as follows:

- to the stochastic process formulation of smoothing splines in state-space form by deriving the variance-covariance structure of the error term for the cubic smoothing spline,

- to non-parametric prediction by deriving the best unbiased predictor and the covariance matrix of its prediction error, based on cubic smoothing splines, and subsequently by formulating the associated prediction intervals,

- in the field of hypothesis testing by deriving two new likelihood ratio based tests for detecting nonlinearity in the semiparametric regression models. Also the empirical comparison of the range of available alternative tests makes a useful contribution to the literature,

- to the bandwidth selection methodology for cubic spline smoothers, by finding that our MML method is competitive to the widely used GCV method, being fast and easy to program.

Finally we wish to close by mentioning some related areas for further research.

- The concept of additional generalisation of variance-covariance matrix derivation for higher order smoothing splines and unequally spaced time can be done by the help of a computer algebra package such as Matlab or Mathematica, (refer to Sections 3.6.2 and 3.6.3).

- Also the derived variance-covariance matrix of the disturbance vector of the semiparametric regression model can be used in the derivation of Wald or Lagrange-multiplier based tests for testing for presence and linearity of a possible nonlinear component.

- The mean based F spline and F* spline tests, which were used as a modification to F and F* tests, need more investigation regarding their size and power.

- Perhaps to take the average of estimated smoothing parameter λ by various methods might improve the smoothing results.

References

AGARWAL, G. and W.J. STUDDEN, (1980), Asymptotic integrated mean square error using least squares and bias minimising splines, *The Annals of Statistics*, 8, 1307-1325.

ALLEN, D., (1974), The relationship between variable selection and data augmentation and a method for prediction, *Technometrics*, 16, 125-127.

ANDERSSEN, R.S. and P. BLOOMFIELD, (1974), A time series approach to numerical differentiation, *Technometrics*, 16, 69-75.

ANSLEY, C.F. and R. KOHN, (1985), Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions, *The Annals of Statistics*, 13, 1286-1316.

ARA, I., (1995), Marginal likelihood based tests of regression disturbances, unpublished Ph.D. thesis, Department of Econometrics and Business Statistics, Monash University, Australia.

ATKINSON, K., (1993), *Elementary Numerical Analysis*, Wiley, New York.

AZZALINI, A. and A. BOWMAN, (1990), On the use of nonparametric regression for checking linear relationships, *Journal of the Royal Statistical Society B*, 55, 549-557.

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.N. and H.D. BRUNK, (1972), *Statistical Inference under Order Restrictions*, Wiley, New York.

BARNDOFF-NIELSEN, O.E., (1978), *Information and Exponential Families*, Wiley, Chichester.

BARNDOFF-NIELSEN, O.E., (1983), On a formula for the distribution of the maximum likelihood estimator, *Biometrika*, 70, 343-365.

BASU, D., (1977), On the elimination of nuisance parameters, *Journal of the American Statistical Association,* 72, 355-366.

BHAPKER, V.P., (1989), Conditioning on ancillary statistics and loss of information in the presence of nuisance parameters, *Journal of Statistical Planning and Inference,* 21, 139-160.

BHAPKER, V.P., (1991), Loss of information in the presence of nuisance parameters and partial sufficiency, *Journal of Statistical Planning and Inference,* 28, 185-203.

BREIMAN, L. and J.H. FRIEDMAN, (1985), Estimating optimal transformations for multiple regression and correlation, *Journal of the American Statistical Association,* 80, 580-619.

BUCKLEY, M.J., (1991), Detecting a smooth signal: Optimality of a cusum based procedure, *Biometrika,* 78, 253-262.

BUJA, A., HASTIE, T.J. and R.J. TIBSHIRANI, (1989), Linear smoothers and additive models, *The Annals of Statistics,* 17, 453-555.

BURDEN, R.L. and J.D. FAIRES, (1989), *Numerical Analysis*, PWS-Kent, Boston.

BUSE, A. and L. LIM, (1977), Cubic splines as a special case of restricted least squares, *Journal of the American Statistical Association*, 72, 64-68.

CHAPRA, S.C. and R.P. CANALE, (1990), *Numerical Methods for Engineers*, McGraw-Hill, New York.

CHEN, V., RUPPERT, C.P. and D. SHOEMAKER, (1999), Applying experimental design and regression splines to high-dimensional continuous-state stochastic dynamic programming, *Operations Research,* 47, 38-53.

COULSON, N.E., (1992), Semiparametric estimates of the marginal price of floorspace, *Journal of Real Estate Finance and Economics*, 5, 73-83.

COX, D.D. and E. KOH, (1989), A smoothing spline based test of adequacy in polynomial regression, *Annals of the Institute of Statistical Mathematics*, 41, 383-400.

COX, D.D., KOH E., WAHBA, G. and B.S. YANDELL, (1988), Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models, *The Annals of Statistics*, 16, 113-119.

COX, D.R. and N. REID, (1987), Parameter orthogonality and approximate conditional inference (with discussion), *Journal of the Royal Statistical Society*, B, 49, 1-39.

CRAVEN, P. and G. WAHBA, (1979), Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalised cross-validation, *Numerical Mathematics*, 31, 377-403.

CROWDER, M.J., (1976), Maximum likelihood estimation for dependent observations, *Journal of the Royal Statistical Society*, B, 55, 467-471.

DE-BOOR, C., (1978), *A Practical Guide to Splines*, Springer, New York.

DONOHO, D.L. and I.M. JOHNSTONE, (1994), Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, 81, 425-455.

DONOHO, D.L. and I.M. JOHNSTONE, (1995), Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, 90, 1200-1224.

DONOHO, D.L., JOHNSTONE, I.M., KERKYACHARIN, G. and D. PICARD, (1995), Wavelet shrinkage: Asymptopia? (with discussion), *Journal of the Royal Statistical Society*, B, 57, 301-370.

ENGLE, R.F., GRANGER, C.W.J., RICE, J. and A. WEISS, (1986), Semiparametric estimates of the relationship between weather and electricity sales, *Journal of the American Statistical Association*, 81, 310-320.

EUBANK, R.L., (1988), *Spline Smoothing and Non-parametric Regression*, Marcel Dekker, New York - Basel.

EUBANK, R.L. and C.H. SPIEGELMAN, (1990), Testing for goodness of fit of a linear model using nonparametric regression techniques, *Journal of the American Statistical Association*, 85, 387-392.

FAN, J. and I. GIJBELS, (1996), *Local Polynomial Modelling and its Application*, Chapman and Hall, London.

FAREBROTHER, R.W., (1988), Some recent econometric work on tests for inequality constraints, *Journal of Quantitative Economics*, 4, 227-238.

FISHER, R.A., (1922), On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society of London*, A, 222, 309-368.

FISHER, M., NYCHKA, D. and D. ZERVOS, (1995), Fitting the term structure of interest rates with smoothing splines, *Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series: 95/01.*

FRIEDMAN, J.H. and B.W. SILVERMAN, (1989), Flexible parsimonious smoothing and additive modeling, *Technometrics*, 31, 3-39.

GERALD, C.F. and P.O. WHEATLEY, (1989), *Applied Numerical Analysis*, Addison -Wesley, New York.

GODFREY, L.G., (1988), *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches,* Cambridge University Press, Cambridge.

GOURIEROUX, C., HOLLY, A. and A. MONFORT, (1980), Kuhn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters, *Discussion Paper No. 770*, Harvard Institute of Economic Research, Harvard University.

GRANGER, C.W.J., (1981), Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, 16, 121 – 130.

GREEN, P.J. and B.W. SILVERMAN, (1994), *Nonparametric Regression and Generalised Linear Models,* Chapman and Hall, London.

GU, C., (1990), Adaptive spline smoothing in non-Gaussian regression models, *Journal of the American Statistical Association*, 85, 801-807.

HALL, A.D., ANDERSON, H.M. and C.W.J. GRANGER (1992), A cointegration analysis of Treasury bill yields, *The Review of Economics and Statistics*, 74, 116 – 126.

HARDLE, W., (1990), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

HASTIE, T. and R. TIBSHIRANI, (1990), *Generalized Additive Models*, Chapman and Hall, London.

HUTCHISON, M.F. and F.R. DE HOOG, (1985), Smoothing noisy data with spline functions, *Numerical Mathematics*, 47, 99-106.

IMHOF, P.J., (1961), Computing the distribution of quadratic forms in normal variables, *Biometrika*, 48, 419-426.

JUDGE, G.G., HILL, R.C., GRIFFITHS, W.E., LUTKEPOHL, H. and T. LEE, (1988), *Introduction to the Theory and Practice of Econometrics*, Wiley and Sons, New York.

KADIYALA, K.R., (1970), Testing for the independence of regression disturbances, *Econometrica*, 38, 97-117.

KALBFLEISCH, J.D. and D.A. SPROTT, (1970), Application of likelihood methods to models involving large numbers of parameters (with discussion), *Journal of the Royal Statistical Society*, B, 32, 175-208.

KING, M.L., (1979), Some aspects of statistical inference in the linear regression model, unpublished Ph.D. thesis, University of Canterbury, Christchurch.

KING, M.L., (1980), Robust tests for spherical symmetry and their application to least squares regression, *The Annals of Statistics*, 8, 1265-1271.

KING, M.L., (1981), The Durbin-Watson test for serial correlation: Bounds for regressions with trend and/or seasonal dummy variables, *Econometrica*, 49, 1571-1581.

KING, M.L., (1987), Towards a theory of point optimal testing, *Econometric Reviews*, 6, 169-218.

KING, M.L. and M.A. EVANS, (1984), A joint test for serial correlation and heteroscedasticity, *Economic Letters*, 18, 31-34.

KING, M.L. and G.H. HILLIER, (1985), Locally best invariant tests of the error covariance matrix of the linear regression model, *Journal of Royal Statistical Society B*, 47, 98-102.

KOHN, R. and C.F. ANSLEY, (1987), A new algorithm for spline smoothing based on smoothing a stochastic process, *SIAM Journal on Scientific and Statistical Computing*, 8, 33-48.

KOHN, R., ANSLEY, C.F. and D. THARM, (1991), The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters, *Journal of the American Statistical Association,* B, 86, 1042-1050.

KOHSAKA, H., (1992), Three-dimensional representation and estimation of retail store demand by bicubic splines, *Journal of Retailing*, 68, 221-241.

KUHN, H. and A. TUCKER, (1951), Non-linear programming, *Proceedings of the Second Berkeley Symposium,* University of California Press, 481-492.

LEHMANN, E.L., (1983), *Theory of Point Estimation,* Wiley and Sons, New York.

LEHMANN, E.L., (1986), *Testing Statistical Hypothesis,* Wiley and Sons, New York.

LIN, X. and D. ZHANG, (1999), Inference in generalised additive mixed models by using smoothing splines, *Journal of Royal Statistical Society B*, 61, 381-400.

LLOYD, C.J., (1987), Optimality of marginal likelihood estimating equations, *Communications in Statistics: Theory and Methods*, 16, 1733-1741.

LUO, Z. and G. WAHBA, (1997), Hybrid adaptive splines, *Journal of the American Statistical Association*, 92, 107-116.

MADDALA, G.S., (1992), *Introduction to Econometrics*, Maxwell Macmillan, New York.

MATHEWS, J.H., (1992), *Numerical Methods for Mathematics, Science and Engineering*, Prentice-Hall International, USA.

MAZZI, S. and P. DE JONG, (1998), The smoothing of sequence data using filtering technique, *unpublished paper*, University of Hull and London School of Economics, UK.

MOYEED, R.A., (1995), Spline smoothers as dynamic linear model, *Australian Journal of Statistics*, 37, 193-204.

MUNSON, P.J. and R.W. JERNIGAN, (1989), A cubic spline extension of the Durbin-Watson test, *Biometrika*, 76, 39-47.

NEUBAUER, G.P. and M.G. SCHIMEK, (1994), A note of cross-validation for smoothing splines, in *Statistical Theory and Computational Aspects of Smoothing*, Proceedings of the COMPSTAT '94 Satellite Meeting held in Semmering, Austria, 165-177, Physica-Verlag.

NIU, X., (1996), Nonlinear additive models for environmental time series, *Journal of the American Statistical Association*, 91, 1310-1321.

NYCHKA, D., WAHBA, G., GOLDFARB, S. and T. PUGH, (1984), Cross-validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two-dimensional cross sections, *Journal of the American Statistical Association*, 79, 832-846.

POIRIER, D., (1973), Piecewise regression using cubic spline, *Journal of the American Statistical Association*, 68, 515-524.

PRENTER, P., (1975), *Splines and Variational Methods*, John Wiley, New York.

RAMSEY, J.B., (1969), Tests for errors in classical least-squares regression analysis, *Journal of the Royal Statistical Society, B*, 31, 350-371.

RAMSEY, J.B. and R. GILBERT, (1972), A Monte Carlo study of some small sample properties of tests for specification error, *Journal of the American Statistical Association*, 67, 180-186.

REINCH, C., (1967), Smoothing by spline functions, *Numerical Mathematics*, 10, 177-183.

REINSEL, G.C., (1997), *Elements of Multivariate Time Series Analysis*, Springer – Verlag, New York.

REMON, M., (1984), On a concept of partial sufficiency: L-sufficiency, *International Statistical Review*, 52, 127-136.

ROBERTSON, T., WRIGHT, F.T. and R.L. DYKSTRA, (1988), *Order Restricted Statistical Influence,* John Wiley, New York.

ROBINSON, P.M., (1983), Nonparametric estimators for time series, *Journal of Time Series Analysis*, 4, 185-207.

ROBINSON, T. and MOYEED, R., (1989), Making robust the cross-validity choice of smoothing parameter in spline smoothing regression, *Communication Statistics*, 18, 523–539.

RODRIGUIEZ-POO, J.M, (1992), Estimating the time-of-day electricity demand using the constrained smoothing spline estimator, *working paper*, Universite Catholique de Louvain.

RODRIGUIEZ-POO, J.M, (1999), Constrained smoothing splines, *Econometric Theory*, 15, 114-138.

SCHIMEK, M.G., (2000), *Smoothing and Regression*, John Wiley and Sons, New York.

SCHLITTGEN, R., (1991), Resistant decomposition of economic time series, *Empirica*, 18, 47-63.

SCHUMAKER, L., (1981), *Spline Functions*, John Wiley, New York.

SCHWARZ, G., (1978), Estimating the dimension of a model, *The Annals of Statistics*, 6, 461-464.

SHEPP, L., (1966), Radon-Nikodym derivatives of Gaussian measures, *The Annals of Mathematical Statistics*, 37, 321-354.

SHIVELY, T., KOHN, R. and C.F. ANSLEY, (1994), Testing for linearity in a semiparametric regression model, *Journal of Econometrics*, 64, 77-96.

SHIVELY, T. and T. SAGER, (1997), A semiparametric regression approach to adjusting for meteorological variables in pollution trends, *working paper*, Department of Management Science and Information Systems, University of Texas at Austin.

SHOENBERG, I.J., (1964), Spline functions and the problem of graduation, *Proceedings of the National Academy of Sciences of the United States of America*, 52, 947-950.

SILVERMAN, B.W., (1985), Some aspects of the spline smoothing approach to nonparametric regression curve fitting, *Journal of the Royal Statistical Society*, B, 47, 1-52.

SIMONOFF, S.S., (1996), *Smoothing Methods in Statistics*, Springer – Verlag, New York.

SMITH, M., (1998), Short term forecasting of New South Wales electricity system load, *working paper*, Australian Graduate School of Management, University of New South Wales.

SMITH, M. and R. KOHN, (1996), Nonparametric regression via Bayesian variable selection, *Journal of Econometrics*, 75, 317-344.

SMITH, M. and R. KOHN, (1997), A Bayesian approach to nonparametric bivariate regression, *Journal of the American Statistical Association*, 92, 1522-1535.

SMITH, M., SHEATHER, S. and R. KOHN, (1996), Finite sample performance of robust Bayesian regression, *Computational Statistics*, 11, 269-301.

SMITH, M., WONG, C. and R. KOHN, (1998), Additive nonparametric regression with autocorrelated errors, *Journal of Royal Statistical Society*, 60, 311-331.

SMITH, M., YAU, P., SHIVELY, T. and R. KOHN, (1998), Estimating long-term trends in tropospheric ozone levels, *working paper*, Department of Econometrics and Business Statistics, Monash University.

SMITH, R. and L. HUANG, (1993), Modelling high threshold exceedances of urban ozone, *Technical Report Number 6*, National Institute of Statistical Sciences.

SPEED, T.P., (1991), Comment on that BLUP is a good thing: The estimation of random effects, by G.K. Robinson, *Statistical Science*, 6, 42-44.

STUART, A. and K. ORD, (1991), *Kendall's Advanced Theory of Statistics: Volume 2*, Oxford University Press, New York.

TUNNICLIFFE WILSON, G., (1989), On the use of marginal likelihood in time series model estimation, *Journal of the Royal Statistical Society*, B, 51, 15-27.

VASICEK, O., FONG, A. and H. GIFFORD, (1982), Term structure modelling using exponential splines, *Journal of Finance*, 37, 339-348.

VERBYLA, A.P., (1990), A conditional derivation of residual maximum likelihood, *Australian Journal of Statistics*, 32, 227-230.

WAHBA, G., (1975), Smoothing noisy data with spline functions, *Numeric Mathematics*, 24, 383-393.

WAHBA, G., (1978), Improper priors, spline smoothing and the problem of guarding against model errors in regression, *Journal of the Royal Statistical Society*, B, 40, 364-372.

WAHBA, G., (1985), A comparison of GCV and GML for choosing the smoothing parameter in the generalised spline smoothing problem, *The Annals of Statistics*, 13, 1378-1402.

WAHBA, G., (1990), *Spline Models for Observational Data*, SIAM, Philadelphia.

WAHBA, G. and S. WOLD, (1975), A completely automatic French curve, *Communications in Statistics*, 4, 1-17.

WECKER, W.E. and C.F. ANSLEY, (1983), The signal extraction approach to nonlinear regression and spline smoothing, *Journal of the American Statistical Association*, 78, 81-89.

WHITTAKER, E.T., (1923), On a new method of graduation, *Proceedings of Edinburgh Mathematical Society*, 41, 63-75.

WU, P.X., and M.L. KING, (1994), One-sided testing, *Pakistan Journal of Statistics*, 10, 261-300.

YANAGIMOTO, T. and M. YANAGIMOTO, (1987), The use of marginal likelihood for the diagnostic test for the goodness of fit of the simple linear regression model, *Technometrics*, 29, 95-101.