CLUSTER-BASED LACK OF FIT TESTS FOR NONLINEAR REGRESSION MODELS

by

WIJITH PRASANTHA MUNASINGHE

B.Sc., University of Colombo, Sri Lanka, 1999
M.S., Kansas State University, 2005

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2010

Abstract

Checking the adequacy of a proposed parametric nonlinear regression model is important in order to obtain useful predictions and reliable parameter inferences. Lack of fit is said to exist when the regression function does not adequately describe the mean of the response vector. This dissertation considers asymptotics, implementation and a comparative performance for the likelihood ratio tests suggested by Neill and Miller (2003). These tests use constructed alternative models determined by decomposing the lack of fit space according to clusterings of the observations. Clusterings are selected by a maximin power strategy and a sequence of statistical experiments is developed in the sense of Le Cam. L_2 differentiability of the parametric array of probability measures associated with the sequence of experiments is established in this dissertation, leading to local asymptotic normality. Utilizing contiguity, the limit noncentral chi-square distribution under local parameter alternatives is then derived. For implementation purposes, standard linear model projection algorithms are used to approximate the likelihood ratio tests, after using the convexity of a class of fuzzy clusterings to form a smooth alternative model which is necessarily used to approximate the corresponding maximin optimal statistical experiment. It is demonstrated empirically that good power can result by allowing cluster selection to vary according to different points along the expectation surface of the proposed nonlinear regression model. However, in some cases, a single maximin clustering suffices, leading to the development of a Bonferroni adjusted multiple testing procedure. In addition, the maximin clustering based likelihood ratio tests were observed to possess markedly better simulated power than the generalized likelihood ratio test with semiparametric alternative model presented by Ciprian and Ruppert (2004).

CLUSTER-BASED LACK OF FIT TESTS FOR NONLINEAR REGRESSION MODELS

by

Wijith Prasantha Munasinghe

B.Sc., University of Colombo, Sri Lanka, 1999
M.S., Kansas State University, 2005

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2010

Approved by:

Major Professor
James W. Neill

Abstract

Checking the adequacy of a proposed parametric nonlinear regression model is important in order to obtain useful predictions and reliable parameter inferences. Lack of fit is said to exist when the regression function does not adequately describe the mean of the response vector. This dissertation considers asymptotics, implementation and a comparative performance for the likelihood ratio tests suggested by Neill and Miller (2003). These tests use constructed alternative models determined by decomposing the lack of fit space according to clusterings of the observations. Clusterings are selected by a maximin power strategy and a sequence of statistical experiments is developed in the sense of Le Cam. L_2 differentiability of the parametric array of probability measures associated with the sequence of experiments is established in this dissertation, leading to local asymptotic normality. Utilizing contiguity, the limit noncentral chi-square distribution under local parameter alternatives is then derived. For implementation purposes, standard linear model projection algorithms are used to approximate the likelihood ratio tests, after using the convexity of a class of fuzzy clusterings to form a smooth alternative model which is necessarily used to approximate the corresponding maximin optimal statistical experiment. It is demonstrated empirically that good power can result by allowing cluster selection to vary according to different points along the expectation surface of the proposed nonlinear regression model. However, in some cases, a single maximin clustering suffices, leading to the development of a Bonferroni adjusted multiple testing procedure. In addition, the maximin clustering based likelihood ratio tests were observed to possess markedly better simulated power than the generalized likelihood ratio test with semiparametric alternative model presented by Ciprian and Ruppert (2004).

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank the faculty and secretaries of the Department of Statistics for providing an excellent environment for professional learning and growth. Especially, I would like to express my hearty appreciation to my major professor, Dr. James Neill, for his invaluable guidance, thoughtful caring and considerate patience. At the beginning of this research work he spent hours and hours to teach important concepts, in the execution stage he carefully studied my work and guided me towards correct direction, and finally during write up of the dissertation he promptly suggested all necessary changes. All those things definitely helped me to make this achievement possible. Very special thanks goes to Dr. Forrest Miller for all his bright ideas and tips given to me during our group discussions. Drs. Paul Nelson and Haiyan Wang deserve special thanks for their helpful discussions and suggestions that enhanced this research.

I would like to acknowledge the financial support from the Department of Statistics during my graduate studies. My sincere thanks goes to Vice Chancellor, Dean of the Faculty of Science and Head of the Department of Statistics all at University of Colombo, Sri Lanka for supporting me during the last six years.

My deepest gratitude goes to my wife Chamila, my little daughter Iruni, and my parents for their endless love and support throughout the entirety of my education. This dissertation is dedicated to them. Their love and encouragement for me made this achievement possible.

Dedication

To my family..., and my parents...

Chapter 1

Introduction

Checking model adequacy in parametric nonlinear regression models is important in order to obtain useful predictions and reliable parameter inferences. In addition, although non-parametric and semiparametric regression techniques offer flexible approaches to modelling, experimenters often prefer a parametric model for ease of interpretation. Also, provided that a specified model is correct, more efficient inference procedures are available for the parametric case. Consequently, a fundamental problem is that of testing the adequacy of a proposed parametric model. Such tests are refered to as lack of fit or goodness of fit tests by statisticians, and as model specification tests by econometricians. Lack of fit is said to exist when the regression function does not adequately describe the mean of the response vector. That is, expectation of the response vector is not an element of the expectation surface corresponding to the proposed model.

More specifically, consider a parametric nonlinear regression model with additive error given by

$$y_{ni} = f\left(x_{ni}, \beta\right) + \epsilon_{ni} \tag{1.1}$$

where $i = 1, 2, ..., n,\ n \geq 1$, and $f\left(x, \beta\right)$ is a specified regression function. The x_{ni} are deterministic predictor vectors contained in R^q and the observable random variables,

$\{y_{ni} : 1 \leq i \leq n, n \geq 1\}$, form a triangular array of rowwise independent variables. The random errors ϵ_{ni} are assumed to be identically distributed with zero mean and unknown positive finite variance, σ^2. Finally, the unknown parameter β ranges over a subset $B \subseteq R^p$. Let $y_n = (y_{n1}, ..., y_{nn})^T$, $\epsilon_n = (\epsilon_{n1}, ..., \epsilon_{nn})^T$, and $f_n(\beta) = (f(x_{n1}, \beta), ..., f(x_{nn}, \beta))^T$.

The expectation surface generated by β is the curved surface in R^n given by

$$M_n = \{m \in R^n : m = f_n(\beta), \beta \in B\}.$$

Lack of fit is said to exist when $f_n(\beta)$ does not adequately describe the mean of y_n i.e. $E(y_n)$ is not an element of the expectation surface M_n corresponding to the proposed model.

In case lack of fit exists in model (1.1), suppose that

$$E(y_n) = f_n(\beta) + \xi_n(\beta) \tag{1.2}$$

where $\xi_n(\beta)$ denotes a lack of fit vector known only to be in some subspace of the lack of fit space given by $T_\beta M_n^\perp$. $T_\beta M_n$ represents the tangent vector space to the expectation surface M_n at $\beta \in B$, and the superscript $\perp$ represents orthogonal complementation. Furthermore, $T_\beta M_n$ is spanned by the vectors $\partial f_n(\beta)/\partial \beta_j$ for $1 \leq j \leq p$.

In analogy with Christensen(1989, 1991), general nonlinear alternative models can be constructed by decomposing the lack of fit space according to clusterings of the observations. In particular, a decomposition of $T_\beta M_n^\perp$ into orthogonal subspaces is given by

$$C(Z_n(\beta)) \cap T_\beta M_n^\perp \oplus C(Z_n(\beta))^\perp \cap T_\beta M_n^\perp \oplus S$$

where the notation $C(A)$ denotes the column space of a matrix A, and $Z_n(\beta)$ denotes a clustering corresponding to the point $f_n(\beta) \in M_n$. In particular, $C(Z_n(\beta)) \cap T_\beta M_n^\perp$ represents the nonlinear analogue to so-called between-cluster lack of fit, while $C(Z_n(\beta))^\perp \cap T_\beta M_n^\perp$

represents the nonlinear analogue to so-called within-cluster lack of fit. Finally, S represents the orthogonal complement of the direct sum of the nonlinear between- and within-cluster lack of fit subspaces with respect to $T_\beta M_n^\perp$, a relatively low dimensional subspace.

Neill and Miller (2003) presented extensions of Christensen's methods to the nonlinear case. In particular, statistical experiments were constructed which generalized Christensen's orthogonal between- and within-cluster lack of fit tests to the nonlinear case, permitting nonnormal errors as well. Each experiment determine a likelihood ratio test (LRT) and a decision, based on y_n, as to whether or not lack of fit exists in model (1.1). Extensions of the methods of Miller et al. (1998, 1999) were then presented by Neill and Miller in order to choose an optimal such experiment from a large class of such. The term statistical experiment is used in the sense developed by Le Cam (1972), and described by Shiryaev and Spokoiny (2000) and Le Cam and Yang (2001). The choice of an optimal alternative model, against which to test the adequacy of a proposed model, involves comparing experiments and taking limits of experiments using the local asymptotic normality (LAN) of Le Cam.

Chapter two provides a review of the literature which addresses the problem of testing for lack of fit in parametric nonlinear models. Categories of such tests include those based on measures of discrepancy between the parametric fits and nonparametric smooths of regression functions, as well as tests which use empirical processes based on residuals. In addition, generalized LRTs based on semiparametric alternatives and various generalizations of the classical regression lack of fit LRT with replication are noted.

Chapter three reviews the construction of statistical experiments as given by Neill and Miller (2003), which represent nonlinear generalizations of Christensen's (1989, 1991) linear models for orthogonal between- and within-cluster lack of fit. In contrast to the linear case, constructed full models use the convexity of a class of fuzzy clusterings to form smooth alter-

native models. In particular, cluster selection is allowed to vary according to different points along the expectation surface in the nonlinear case. The choice of clustering at such points is based on a maximin power clustering strategy presented by Miller et al. (1998, 1999), and extended by Neill and Miller (2003) to the nonlinear case. The resulting LRTs for assessing nonlinear lack of fit are given, and an implementation algorithm is presented in this chapter.

Indeed, a major objective of this dissertation is the implementation of such LRTs for nonlinear lack of fit, and moreover, to empirically investigate the effectiveness of using maximin power clusterings at different points along the expectation surface, or variations thereof.

Chapter four presents the results of such simulation studies using various data generators to assess the effectiveness of the maximin cluster based LRTs presented by Neill and Miller. Included in this chapter is a comparison study with a generalized LRT based on semiparametric alternatives, which has been previously discussed in the literature.

Chapter five addresses the necessary asymptotic distributional results associated with the LRT statistics described in chapter three, and for carrying out such tests in the simulation studies in chapter four. The derivation of the asymptotic noncentral chi-square distribution under local parameter alternatives is based in part on the corresponding sequence of statistical experiments being LAN. A key condition to ensure LAN for regression problems is that the sequence of experimetns be uniformly differentiable in quadratic mean, as defined in chapter five.

Another objective of this dissertation is to establish the L_2 differentiability of the parametric array of probability measures associated with the sequence of experiments, as defined by Rieder (1994). Using uniform q. m. d. and certain regularity conditions, such L_2 differentiability is determined in chapter five. The LAN property then follows from Theorem 2.3.9

of Rieder. The limit chi-square distribution is then determined by utilizing contiguity and Le Cam's third lemma. The method of proof follows van der Vaart (1998), and represents generalization to the case of triangular arrays. The $\sqrt{n}-$consistency of maximum likelihood estimates (MLE) under the restricted and unrestricted models is a requirement for the proof. Such properties of MLEs are established in this chapter as well.

Chapter 2

Literature Review

In the literature, several approaches for assessing the presence of lack of fit have been suggested. More recent development includes the use of nonparametric regression techniques to test the adequacy of parametric regression models. For this approach, a nonparametric method of estimating the underlying regression function is used to provide a smoothed fit to the observations. Such a smooth can be obtained by kernel and local regression methods (Schucany(2004) and Loader(1999)), smoothing splines (Eubank(1999)) and expansions in terms of basis functions using Fourier series or wavelets (Hart(1997), Hastie et al.(2001)). Additionally, Schimek(2000) provides a comprehensive discussion of univariate and multivariate smoothing techniques for regression. A comparison of the fits given by the smooth and the specified parametric model provides a test for lack of fit. Such testing methodology is presented by Hardle and Mammen(1993), Zheng(1996), Hart(1997), Lee and Hart(1998), Dette(1999), Aerts et al.(1999, 2000, 2004), Eubank et al.(2005) and Li(2005), for example. Lopez and Patilea (2009) consider this approach to testing for lack of fit with censored data. In addition, an adaptive goodness-of-fit test based on signed ranks for one dimensional predictors was presented by Rohde(2008). However, nonparametric estimation of a regression with higher dimensional predictors can suffer from the so-called curse of dimensionality. Thus, implementation of lack of fit testing methods which rely on nonparametric estimation for such case can be problematic.

Other approaches to test for lack of fit include tests which use marked empirical processes based on residuals, as discussed by Stute(1997), Stute et al.(1998a) and Diebolt and Zuber(1999). Tests based on residuals may be useful when specification of possible alternative models is not clear. Guo and Koul(2008) suggest a lack of fit test of a parametric regression model with long memory design and errors based on a marked empirical process, as well. In addition, asymptotically distribution free tests for lack of fit via martingale transforms of partial sum residual empirical processes are presented by Stute et al.(1998b) and Khmaladza and Koul(2004). Koul and Song(2008) construct such tests for the Berkson measurement error regression model. Fan and Huang(2001) employ the adaptive Neyman test on the Fourier transform of the residual vector to assess the adequacy of a parametric regression model against nonparametric alternatives. Tests based on certain minimized L_2 distances between a nonparametric regression function estimator and the fitted parametric model are presented by Koul and Ni(2004) for the classical regression model, and by Koul and Song(2009) for the Berkson measurement error model. A data driven smooth specification test using model residuals was presented by Guerre and Lavergne(2005). In addition, a goodness-of-fit testing procedure based on normalized differences of different estimators was given by Caouder and Huet(1997). Finally, graphical methods used for checking model adequacy based on residuals are given by Cook and Weisburg(1997).

In addition to the preceding lack of fit tests, (restricted) likelihood ratio tests (LRTs) with the proposed parametric regression model embedded into a larger semiparametric family using spline functions is discussed by Claeskens (2004) and Ciprian and Ruppert(2004). Fan, Zhang and Zhang(2001) develop generalized LRTs for checking the goodness of fit for a family of parametric regression models against nonparametric alternatives. Presnell and Boos(2004) introduced a likelihood based test for assessing model adequacy which was motivated by cross-validation. When replicate measurements exist at the predictor settings, then the classical lack of fit LRT introduced by Fisher(1922) may be applied to test the

adequacy of a specified parametric model. Matsumoto and Wakaki(2005) give an asymptotic expansion of the null distribution of the classic test under nonnormality. Akritas and Papadatos(2004) extend the classic test to accommodate the heteroscedastic case, and investigate the limiting distributions of the statistics under local alternatives. Using nearest neighbor weights, Wang et al. (2008) and Akritas and Wang(2009) presented lack of fit tests based on ANOVA-type test statistics which accommodate heterosecdastic errors. Neill(1988) presented a generalization of the classical test for nonlinear regression based on near replicate clusters, which was shown to be consistent whenever the proposed model is the orthogonal projection of the true model. For normal theory linear models, Christensen(1989, 1991) generalized the classical test by deriving uniformly most powerful invariant LRTs for detecting general types of lack of fit, given a choice of near replicate clusters of the input variables.

The clustering based tests of Christensen(1989, 1991) and Neill(1988) require the selection of near replicate clusters, the choice of which was left open. Miller et al.(1998, 1999) and Neill et al. (2002) presented a statistically principled method for choosing a clustering of the predictors for linear models. Their method is based on using Christensen's tests and involves the development of an alternative model using a large class of nonparametric step functions. The problem of cluster selection in this context parallels the choice of a smoothing parameter in nonparametric function estimation. Neill and Miller(2003) extended these methods to the fully parametric nonlinear case, presenting likelihood based models and associated LRTs for detecting general types of nonlinear lack of fit as derived from large classes of possible clusterings.

The problem of testing model adequacy is historically old and has generated much research. The preceding articles, and references therein, provide some more recent examples of such work. This dissertation considers asymptotics, implementation and a comparative performance for the LRTs suggested by Neill and Miller. Broader comparisons with other types

of lack of fit tests would be of interest in future work.

Chapter 3

Likelihood-Based Tests for Detecting General Types of Nonlinear Lack of Fit

This chapter discusses the construction of general alternative models by decomposing the lack of fit subspace according to clusterings of the observations, leading to between-cluster lack of fit experiments.

A maximin power clustering strategy is described for clustering selection. In addition, an implementation algorithm is presented that can be used to obtain an approximation to the LRT for testing nonlinear between-cluster lack of fit.

3.1 Lack of Fit Experiments and Decisions

As indicated in Chapter 1, in analogy with Christensen(1989, 1991), general nonlinear alternative models can be constructed by decomposing the lack of fit space according to clusterings of the observations. A method of constructing such alternative models is given in this subsection.

To specify a clustering of the n observations into c clusters, let Z_n denote a $n \times c$ matrix with $Z_n = [z_{ij}]$ for $1 \leq i \leq n$, $1 \leq j \leq c$, where $z_{ij} \in [0, 1]$ and $\sum_{j=1}^{c} z_{ij} = 1$. Thus, the ij^{th} element of Z_n assigns a measure of membership for the i^{th} observation to the j^{th} cluster. Such a

clustering is called a fuzzy clustering. A special case of a fuzzy clustering is obtained when each observation belongs to exactly one cluster. The matrix Z_n then contains only zeros and ones, and the nonzero values in the j^{th} column of Z_n correspond to the observations in the j^{th} cluster, $1 \leq j \leq c$. Such a clustering is called a crisp clustering. Fuzzy clusterings are of interest in the nonlinear case since they can be used to obtain smooth alternative models for general types of lack of fit.

An alternative model representing nonlinear between-cluster lack of fit is defined by

$$y_n = f_n(\beta) + B_{Z_n(\beta)}\gamma + \epsilon_n \tag{3.1}$$

where $B_{Z_n(\beta)}$ is a matrix whose columns form a basis for $C(Z_n(\beta)) \cap T_\beta M_n^\perp$. An experiment based on model 3.1 is determined by n, the x_{ni} for $1 \leq i \leq n$, and a choice of clustering matrix $Z_n(\beta)$ for each $\beta \in B$. The experiment is specified by the parametric family $P_{n,\theta}$ of probability measures on the Borel sets A^n in R^n where $\theta = (\beta, \gamma, \eta)$ and

$$dP_{n,\theta}(y_n) = \prod_{i=1}^{n} p_{ni,\theta}(y_{ni}) = \prod_{i=1}^{n} p\left(y_{ni} - (f_n(\beta))_i - \left(B_{Z_n(\beta)}\gamma\right)_i, \eta\right). \tag{3.2}$$

Here $p(\epsilon, \eta)$ is the common probability density of the errors ϵ_{ni} for $1 \leq i \leq n$ with respect to some measure μ, say, and η is a vector of nuisance parameters in R^s. Hence, testing for nonlinear between-cluster lack of fit can be carried out by using a LRT to test $H_o : \gamma = 0$ versus $H_a : \gamma \neq 0$. By symmetry, a test for nonlinear within-cluster lack of fit can be obtained by replacing $C(Z_n(\beta))$ with $C(Z_n(\beta))^\perp$ in the previous discussion. However, the focus of this dissertation concerns the detection of nonlinear between-cluster lack of fit.

A class of clusterings from which $Z_n(\beta), \beta \in B$ can be selected is discussed next. The construction is based on a method given by Miller et al.(1998, 1999) for testing cluster-based

lack of fit in the linear case. Let the predictor settings be denoted by $V_n = \{x_{ni} : 1 \leq i \leq n\}$. To summarize the construction, the set of possible clusterings is given by the collection of clusterings consistent with a specified cover $C_n = \{C_{n1}, ..., C_{nm}\}$ of the predictors in V_n. That is (identifying the elements of each C_{nj} with the indices of the corresponding observations), this collection consists of all partitions P of the set of observation indices $\{1, 2, ..., n\}$ which satisfy the condition: if $P = \{A_1, ..., A_a\}$ then, for each $1 \leq i \leq a$, there exists a $1 \leq j \leq m$ such that $A_i \subseteq C_{nj}$. Further, atoms were defined to be those partitions consistent with the cover C_n that group as many observations together as possible.

In addition, a method to determine a cover can be based on a family $F_n = \{F_{n1}, ..., F_{nm}\}$ of overlapping subsets in R^q whose union includes the predictors in V_n. The cover elements are then given by $C_{nj} = F_{nj} \cap V_n$ for $1 \leq j \leq m$. Also, a method of selecting the overlapping subsets F_n consists of dividing the predictor variable space into cells, where any cell indexed by a q-tuple with all odd components is referred to as an odd cell. The overlapping subsets in F_n are then determined by taking the union of each nonempty odd cell with all contiguous cells. Finally, the set of atoms consistent with the cover can be usefully represented as the product space

$$\kappa_{n,o} = K_1 \times ... \times K_n$$

where K_i, $1 \leq i \leq n$, is the set of indices of the odd cells that contain observations with which the i^{th} observation can or must be clustered. Note that like coordinates within each particular n-tuple indicate that the corresponding observations are to be clustered together.

In high dimensional problems, it may be that many of the odd cells are empty. In such case, all that is required to specify a cover is a minimal set of disjoint cells $S_{n1}, \ldots, S_{nc}$ where each S_{nj} contains at least one predictor setting and, moreover, every other cell that contains a predictor setting must be contiguous to one of the S_{nj}, $1 \leq j \leq c$. A cover can then be based on the cells $S_{n1}, \ldots, S_{nc}$ analogous to the construction above.

In the nonlinear case, in order to obtain smooth alternative models representing nonlinear between-cluster lack of fit, $Z_n(\beta)$, $\beta \in B$, is chosen from the class of fuzzy clusters given by the convex hull of the set of atoms $\kappa_{n,o}$ as described below. This class will be denoted as κ_n and the elements will be referred to as fuzzy atoms. Similar to the linear case, alternative models for nonlinear between-cluster lack of fit constructed from the fuzzy atoms lead to (asymptotically) χ^2-distributed test statistics with degrees of freedom parameters that are in concordance with the objective of better power (cf Chapter 5). Under very general circumstances, each such $Z_n(\beta)$ is $n \times c$ with c constant, corresponding to the number of nonempty odd cells. For example, if the occupancy pattern of the odd cells and their neighbors (as far as being empty or not) is the same for each n, then the input data V_n has the same shape for each n and $c = \dim C(Z_n(\beta))$ for each fuzzy atom $Z_n(\beta) \in \kappa_n$. It is henceforth assumed that the input data has the same shape for each n. Note that a simplifying assumption is that $x_{mi} = x_{ni}$ for $1 \leq i \leq m \leq n$ but this is not necessary.

A useful representaion of the class of fuzzy atoms is given next. First, the nonempty odd cells are arbitrarily ordered and labeled $1, ..., c$. Thus, each set K_i is a subset of these labels. In particular, suppose $K_i = \{o_{i_1}, ..., o_{i_{n_i}}\}$ where n_i is the cardinality of K_i and $1 \leq o_{i_j} \leq c$ for $1 \leq i \leq n$. Next, let $co(E)$ denote the convex hull of a set E in R^d. Then $co(K_i)$ is formed by identifying $o_{i_1}, ..., o_{i_{n_i}}$ with the standard basis elements $e_1, ..., e_{n_i}$ of R^{n_i}, and taking convex combinations of the e_j, $1 \leq j \leq n_i$, in R^{n_i}. With the class of fuzzy atoms defined above as $\kappa_n = co(\kappa_{n,o})$, Neill and Miller(2003) noted that

$$\kappa_n = co(K_1) \times ... \times co(K_n).$$

This follows by first noting that the extreme points of the set on the right hand side are the atoms $\kappa_{n,0}$. The equality then follows by the Krein-Milman theorem (e.g. Rudin (1991)).

Next let $\sum_{j=1}^{n_i} \lambda_{ij} e_j \in co(K_i)$, $1 \leq i \leq n$. By the preceding representation of κ_n and the identifications made above, the i^{th} row of the corresponding $Z_n(\beta) \in \kappa_n$ consists of λ_{ij} as the entries corresponding to the odd cells o_{i_j}, $1 \leq j \leq n_i$, and zeros elsewhere. Since $\sum_{j=1}^{n_i} \lambda_{ij} = 1$ for $1 \leq i \leq n$, such a $Z_n(\beta)$ does in fact represent a fuzzy clustering as defined above.

For the reasons given above, consideration is given to experiments that assign a fuzzy clustering $Z_n(\beta)$ to each $\beta \in B$. As indicated above, each such $Z_n(\beta)$ is an $n \times c$ matrix with c constant, so that the θ parameter space is the same for all experiments

$$(R^n, A^n, (P_{n,\theta} : \theta \in \Theta))$$

where $\Theta = B \times R^{c-p} \times R^s$. Recall that the probability measures $P_{n,\theta}$ depend on x_{ni} for $1 \leq i \leq n$ and $Z_n(\beta)$, $\beta \in B$. Therefore, the experiment $(R^n, A^n, P_{n,\theta})$ will be denoted by

$$E_n(x_{n1}, ..., x_{nn}; Z_n(\beta) \in \kappa_n, \beta \in B)$$

and referred to as a *between-cluster lack of fit experiment* (BCLFE).

3.2 Comparing Between-Cluster Lack of Fit Experiments

For comparison of BCLFEs, the strategy employed by Miller et al.(1998,1999) for the linear case was extended by Neill and Miller(2003) for use with nonlinear models. The question to be addressed is, given two experiments which provide a good approximation for a 'true model ', which experiment would give better power in detecting lack of fit? To make this comparison in the linear case, a positive definite quadratic form τ on $C(G_n)^{\perp}$ was introduced and

$$l_{Z_n} = \inf \left\{ \frac{\|\nu\|^2}{\tau(\nu)} : \nu \in C(Z_n) \cap C(G_n)^{\perp}, \nu \neq 0 \right\} \tag{3.3}$$

was maximized over the atoms $Z_n \in \kappa_{n,0}$. For this case, $f(x_{ni}, \beta) = g(x_{ni})^T \beta$ where g is a continuous vector valued function of the q predictor variables, and G_n is the $n \times p$ matrix with i^{th} row given by $g(x_{ni})^T$. Also, $C(Z_n) \cap C(G_n)^{\perp}$ is the linear between-cluster lack of fit subspace corresponding to Z_n as introduced by Christensen (1991). An atom Z_n^M, say, which maximizes l_{Z_n} is called a maximin power clustering, and determines an optimum alternative model for testing linear between-cluster lack of fit. To make the comparison in the nonlinear case let

$$l_{Z_n(\beta)} = \inf \left\{ \frac{\|\nu\|^2}{\tau(\beta)(\nu)} : \nu \in C\left(B_{Z_n(\beta)}\right), \nu \neq 0 \right\} \tag{3.4}$$

for each BCLFE where $\tau(\beta)$ is a positive definite quadratic form on $T_\beta M_n^{\perp}$, and the columns of $B_{Z_n(\beta)}$ form a basis for $C(Z_n(\beta)) \cap T_\beta M_n^{\perp}$. Similar to the linear case, $\tau(\beta)$ is constructed as

$$\tau(\beta)(\nu) = \sum_{Z_n \in \Xi_{E_n}} w_{Z_n} \left\| P_{B_{Z_n(\beta)}} \nu \right\|^2 \tag{3.5}$$

where $\nu \in T_\beta M_n^{\perp}$ with $w_{Z_n} \geq 0$ and $\sum_{Z_n \in \Xi_{E_n}} w_{Z_n} = 1$. Here Ξ_{E_n} denotes the crisp clusterings consistent with a given cover C_n that cluster only two observations, with all other observations being singleton clusters. In addition, $P_{B_{Z_n(\beta)}}$ represents the orthogonal projection operator onto the subspace $C\left(B_{Z_n(\beta)}\right)$. Next, let X_n denote the $n \times q$ matrix of predictor settings with the i^{th} row given by x_{ni}^T. To incorporate nearness as measured by $\|X_n - X_{n,o}\|^2$ into the weight w_{Z_n}, let

$$w_{Z_n} = \|X_n - X_{n,o}\|^2 / \sum_{Z_n^* \in \Xi_{E_n}} \left\|X_n - X_{n,o}^*\right\|^2 \tag{3.6}$$

where $X_{n,o} = P_{Z_n} X_n$ and $X^*_{n,o} = P_{Z^*_n} X_n$ for $Z^*_n \in \Xi_{E_n}$. The notation $\|A\|^2$ denotes the squared matrix norm defined by $\sum_{i=1}^{n} \sum_{j=1}^{q} a_{ij}^2$ where a_{ij} is the ij^{th} element of an $n \times q$ matrix A.

For the nonlinear case, let $Z_n^M(\beta)$, say, maximize $l_{Z_n(\beta)}$ over the fuzzy clusterings κ_n for each $\beta \in B$. Note that $Z_n^M(\beta_0)$ can be considered to be an optimal power selection clustering for testing the approximating linear model around $\beta = \beta_0$. An experiment

$$E_n\left(x_{n1}, ..., x_{nn}; Z_n^M(\beta) \in \kappa_n, \beta \in B\right)$$

is called a *maximin between-cluster lack of fit experiment* (MMBCLFE). As indicated in Chapter 5, using $Z_n^M(\beta)$ for each $\beta \in B$ provides (asymptotically) an optimum BCLFE for testing lack of fit in the model $y_n = f_n(\beta) + \epsilon_n$.

3.3 Asymptotics and the Parameter Space

To determine the role of $l_{Z_n(\beta)}$ in the asymptotic limit, $l_{Z_n(\beta)}$ is next described in terms of the parameter space rather than the observed data space R^n. Note that a choice of the basis vectors $B_{Z_n(\beta)}$ for $C(Z_n(\beta)) \cap T_\beta M_n^\perp$ determines a mapping $\Phi_n : B \times R^{c-p} \to R^n$ by

$$\Phi_n(\beta, \gamma) = f_n(\beta) + B_{Z_n(\beta)}\gamma.$$

The Fisher information metric for the expectation surface parameters is

$$I_n^E(\vartheta)_{jl} = \kappa\left(\sum_{i=1}^{n} \frac{\partial \Phi_{ni}}{\partial \vartheta_j} \frac{\partial \Phi_{ni}}{\partial \vartheta_l}\right)$$

where $\vartheta = (\beta, \gamma)$ and

$$\kappa = \kappa(\eta) = \int_R \frac{\left(p'\right)^2}{p} d\mu,$$

which is $\kappa\Phi_n^*$ (Euclidean metric) with Φ_n^* denoting the pull-back operation under Φ_n. Note κ is independent of n, the x_{ni} for $1 \leq i \leq n$, and $Z_n(\beta)$. Also, I_n^E depends on the nuisance

parameter η only through κ.

Next choose the columns of $B_{Z_n(\beta)}$ to be an orthogonal basis of $C\left(Z_n\left(\beta\right)\right) \cap T_\beta M_n^\perp$ with respect to the Euclidean metric of length $\sqrt{n}$. It then follows that along the space $\gamma = 0$ in the $\gamma-$direction, Φ_n^* is the Euclidean metric with respect to the parameter γ. This is the direction in which $l_{Z_n(\beta)}$ is calculated. In fact, $l_{Z_n(\beta)}$ is the minimum eignvalue of $\frac{1}{\kappa} I_n^E\left(\beta, 0\right)$ in the $\gamma-$direction. Scaling does not change the comparison so that the minimum eignvalue of $\frac{1}{n\kappa} I_n^E\left(\beta, 0\right)$ in the $\gamma-$direction, $\hat{l}_{Z_n(\beta)}$, say, can be used for comparison purposes. The multiplier $\frac{1}{n}$ is included since the convergence $\frac{1}{n} I_n^E\left(\beta, 0\right) \to I^E\left(\beta, 0\right)$, say, is required for the local type of LAN convergence of experiments (cf Chapter 5).

3.4 Implementation Algorithm

In this subsection an algorithm is presented to obtain an approximating sequence of BCLFEs which will be used in place of the corresponding MMBCLFE sequence for implementation purposes. In particular, for a given data size n, a reasonable smooth approximation to the MMBCLFE can be obtained as follows.

1. First, choose a finite set $\beta_1, ..., \beta_N$ from B and at these points take $Z_n^*\left(\beta_i\right)$ to be the crisp clustering in $\kappa_{n,o}$ which maximizes $l_{Z_n(\beta_i)}$ for $1 \leq i \leq N$.

2. Next, using the fact that the set of fuzzy clusterings κ_n is convex, an algorithm for calculating $Z_n^*\left(\beta\right)$, say, for the remaining $\beta \in B$ can be determined. Specifically, let $g : R \to R$ be 3-times continuously differentiable with $g\left(x\right) = 0$ for $x \leq 0$, $g\left(x\right) = 1$ for $x \geq 1$ and $g^{'}\left(x\right) > 0$ for $0 < x < 1$. Consider first the case where the β parameter space B is one-dimensional and suppose $\beta_1 < \beta_2 < ... < \beta_N$ with $Z_n^*\left(\beta_i\right)$ specified according to step 1. Define

   $Z_n^*\left(\beta\right) = Z_n^*\left(\beta_1\right)$ for $\beta \leq \beta_1$,

$Z_n^*(\beta) = \left[1 - g\left(\frac{\beta-\beta_i}{\beta_{i+1}-\beta_i}\right)\right] Z_n^*(\beta_i) + g\left(\frac{\beta-\beta_i}{\beta_{i+1}-\beta_i}\right) Z_n^*(\beta_{i+1})$ for $\beta_i \leq \beta \leq \beta_{i+1}$ with $i = 1, ..., N-1$,

$Z_n^*(\beta) = Z_n^*(\beta_N)$ for $\beta \geq \beta_N$.

It follows that $Z_n^*(\beta)$ is 3-times continuously differentiable. Note that if $Z_n^*(\beta_i)$ for $i = 1, ..., N$ are differentiable functions of some other parameter then $Z_n^*(\beta)$ would be jointly differentiable, as required in the following generalization.

Next suppose the β parameter space B is two-dimensional and a grid of points $\left(\beta_i^{(1)}, \beta_j^{(2)}\right)$, $1 \leq i, j \leq N$, has been selected with corresponding $Z_n^*\left(\beta_i^{(1)}, \beta_j^{(2)}\right)$ specified according to step 1, and where $\beta_1^{(1)} < \beta_2^{(1)} < ... < \beta_N^{(1)}$ and $\beta_1^{(2)} < \beta_2^{(2)} < ... < \beta_N^{(2)}$. $Z_n^*\left(\beta^{(1)}, \beta^{(2)}\right)$ can then be determined as follows.

First, apply the one-dimensional case with $\beta^{(1)}$ as the parameter to the lines $\beta^{(2)} = \beta_1^{(2)}$, $\beta^{(2)} = \beta_2^{(2)}$, ... , $\beta^{(2)} = \beta_N^{(2)}$ to obtain $Z_n^*\left(\beta^{(1)}, \beta_1^{(2)}\right)$, $Z_n^*\left(\beta^{(1)}, \beta_2^{(2)}\right)$, ... , $Z_n^*\left(\beta^{(1)}, \beta_N^{(2)}\right)$ for $\left(\beta^{(1)}, \beta_j^{(2)}\right) \in B$.

Next, for each $\beta_*^{(1)}$ apply the one-dimensional case to the line $\beta^{(1)} = \beta_*^{(1)}$ with $Z_n^*\left(\beta_*^{(1)}, \beta_1^{(2)}\right)$, $Z_n^*\left(\beta_*^{(1)}, \beta_2^{(2)}\right)$, ... , $Z_n^*\left(\beta_*^{(1)}, \beta_N^{(2)}\right)$ specified to obtain $Z_n^*\left(\beta_*^{(1)}, \beta^{(2)}\right)$ for $\left(\beta_*^{(1)}, \beta^{(2)}\right) \in B$. The preceding construction extends to the case in which the β parameter space B is p-dimensional as well.

Thus, the preceding algorithm provides an approximating sequence of BCLFEs

$$E_n\left(x_{n1}, ..., x_{nn}; Z_n^*(\beta) \in \kappa_n, \beta \in B\right)$$

which will be used in place of the corresponding sequence of MMBCLFEs

$$E_n\left(x_{n1}, ..., x_{nn}; Z_n^M(\beta) \in \kappa_n, \beta \in B\right)$$

for implementation purposes.

Note in the linear case, the choices in step 1 always give the same crisp atom so that this method provides the crisp atom maximin power clustering.

The calculation of the LRT statistic can now be discussed using the approximating sequence of BCLFEs given above. In particular, computation of least squares estimates for β and γ in model 3.1 is considered next. Thus, it is a matter of finding the nearest point to the observed data y_n on the expectation surface M_n, and also the nearest point to y_n on the expectation surface

$$M_n^* = \left\{ m \in R^n : m = f_n(\beta) + B_{Z_n^*(\beta)}\gamma : (\beta, \gamma) \in B \times R^{c-p} \right\}$$

for the alternative (full) model provided by $E_n(x_{n1}, ..., x_{nn}; Z_n^*(\beta) \in \kappa_n, \beta \in B)$. Note that $M_n \subset M_n^*$. Since $f_n(\beta)$ is a specified algebraic expression, a Gauss-Newton least squares estimation method can be used to estimate β, and hence solve the nearest point calculation for M_n (e.g. Seber and Wild(1989)).

The nearest point calculation for M_n^* can be reduced to a β-space calculation since M_n^* is fibered across M_n by affine varieties. Specifically, for each $\beta \in B$, let

$$V_\beta = \left\{ m \in R^n : m = f_n(\beta) + B_{Z_n^*(\beta)}\gamma : \gamma \in R^{c-p} \right\}.$$

Now suppose (β_0, γ_0) gives the nearest point $\hat{y}_n$ to y_n on M_n^*. Then $\hat{y}_n \in V_{\beta_0}$ and $\hat{y}_n$ is the nearest point to y_n on V_{β_0}. In fact, for each β let $\hat{\gamma}(\beta)$ give the nearest point to y_n on V_β and let

$$M_n^{**} = \left\{ m \in R^n : m = f_n(\beta) + B_{Z_n^*(\beta)}\hat{\gamma}(\beta) : \beta \in B \right\},$$

consisting of the nearest points to y_n on the fibers V_β. Since each V_β is an affine variety, standard linear model projection algorithms can be used to determine the parameter values $\hat{\gamma}(\beta)$ for each β (e.g. Christensen(2002)). The preceding provides a computational

procedure for obtaining any point on M_n^{**} coming from a particular value of β. Finally, application of a Gauss-Newton method would allow one to find the nearest point to y_n on M_n^{**}, and hence on M_n^*.

The preceding computation provides MLEs for β and γ in the case of Normally distributed random errors. The more general case of exponential family regression is the subject of future work.

Chapter 4

Simulation Results

This chapter presents empirical power results for the cluster based LRTs proposed by Neill and Miller (2003). Alternative models were constructed using a maximin power clustering strategy. In particular, the adequacy of proposed exponential, cosine and Michaelis-Menten nonlinear models was assessed when lack of fit was present due to data generators incorporating nonlinear between-cluster lack of fit, and also in cases when functionally different data generators were used.

The proposed test procedures were also compared empirically with a generalized LRT presented by Ciprian and Ruppert (2004). This approach uses penalized splines to construct a semiparametric alternative model.

4.1 Example 1: Exponential Model

The nonlinear regression function considered is given by $f(x, \beta) = \beta_1 e^{-x\beta_2}$ where $\beta_1 > 0$ and $\beta_2 > 0$ and is called an exponential decay model. Three examples from '*curvefit.com*' for which exponential decay applies are:

1. When ligands dissociate from receptors, the number of molecules that dissociate in any short time interval is proportional to the number that were bound at the beginning

of that interval. Equivalently, each individual molecule of ligand bound to a receptor has a certain probability of dissociating from the receptor in any small time interval. That probability does not get higher as the ligand stays on the receptor longer.

2. When radioactive isotopes decay, the number of atoms that decay in any short interval is proportional to the number of undecayed atoms that were present at the beginning of the interval. This means that each individual atom has a certain probability of decaying in a small time interval, and that probability is constant. The probability that any particular atom will decay does not change over time. The total decay of the sample decreases with time because there are fewer and fewer undecayed atoms.

3. When drugs are metabolized by the liver or excreted by the kidney, the rate of metabolism or excretion is often proportional to the concentration of drug in the blood plasma. Each drug molecule has a certain probability of being metabolized or secreted in a small time interval. As the drug concentration decreases, the rate of its metabolism or excretion decreases as well.

In the function $f(x, \beta) = \beta_1 e^{-x\beta_2}$, β_2 and β_1 are defined as the rate constant and value of the function when x is set to zero, respectively.

The power of the proposed LRT for detecting nonlinear between-cluster lack of fit is investigated in the case of Normal errors. Specifically, Uniformly distributed predictor values for x on the interval $[0, 10]$ were used to analyze data sizes of 30 and 100. One run of each such data size was generated in R (cf Appendices A and B). As discussed in Chapter 3, a cover C_n was specified by dividing $[0, 10]$ into the cells $\{[0, 2), [2, 4), [4, 6), [6, 8), [8, 10]\}$, with associated overlapping subsets given by $F_n = \{F_{n1}, F_{n2}, F_{n3}\} = \{[0, 4), [2, 8), [6, 10]\}$. Since the cardinality of the set of crisp clusterings $\kappa_{n,o}$ becomes very large with increasing data size, a subset comprised of *ordered* partitions from the full collection $\kappa_{n,o}$ was utilized to facilitate computational feasibility for the simulation. In particular, for the specified cover

C_n, suppose in the even cells $[2, 4)$ and $[6, 8)$ there are k and l predictor values and denote their ordered values by $\{x_{11}, x_{12}, ..., x_{1k}\}$ and $\{x_{21}, x_{22}, ..., x_{2l}\}$, respectively. Now group all predictor values $< x_{1m}$ in even cell $[2, 4)$ with odd cell $[0, 2)$, while all predictor values $\geq x_{1m}$ in even cell $[2, 4)$ are grouped with odd cell $[4, 6)$ for $m = 1, 2, ..., k$. In addition, if $m = k$ group all predictor values $\leq x_{1m}$ in even cell $[2, 4)$ with odd cell $[0, 2)$. Similarly, group all predictor values $< x_{2n}$ in even cell $[6, 8)$ with odd cell $[4, 6)$, while all predictor values $\geq x_{2n}$ in even cell $[6, 8)$ are grouped with odd cell $[8, 10]$ for $n = 1, 2, ..., l$. In addition, if $n = l$ group all predictor values $\leq x_{2n}$ in even cell $[6, 8)$ with odd cell $[4, 6)$. This leads to $(k + 1)(l + 1)$ different groupings (crisp clusterings) of the predictor values based on the 5 cells chosen. Based on the specified cover C_n, 49 and 459 such crisp clusterings were produced for the data sizes 30 and 100, respectively. The maximin power clustering criterion, as described in Section 3.2, was then used to find the best such crisp atom $Z_n^*(\beta_i)$ at each of $N = 5$ chosen β values. For illustration purposes, the values of $\beta = (\beta_1, \beta_2)$ were selected as $\{(1, 1), (1, 3), (1, 5), (1, 7), (1, 9)\}$. This is the first step in the implementation algorithm as discussed in Section 3.4. The plots for the regression function $f(x, \beta) = \beta_1 e^{-x\beta_2}$ for the selected $(\beta_1, \beta_2, \sigma)$ values are depicted in Figures 4.1 and 4.2.

Note that the cover C_n specified above is based on $c = 3$ nonempty odd cells, and thus $C(Z_n(\beta))$ has dimension three for all fuzzy clusterings. Further, the dimension of $T_\beta M_n$ is equal to two so that the dimension of $C(B_{Z_n(\beta)})$ is equal to one for all corresponding nonlinear between-cluster lack of fit subspaces.

Tables 4.1 and 4.2 give the crisp clusterings $Z_n^*(\beta_i)$ for each of the $N = 5$ selected β values and for each data size, respectively. These tables indicate that the maximin clusterings do indeed vary according to different points along the expectation surface as indexed by the parameters (β_1, β_2). Considerable difference in selected clusters can be noted especially with increasing data size.

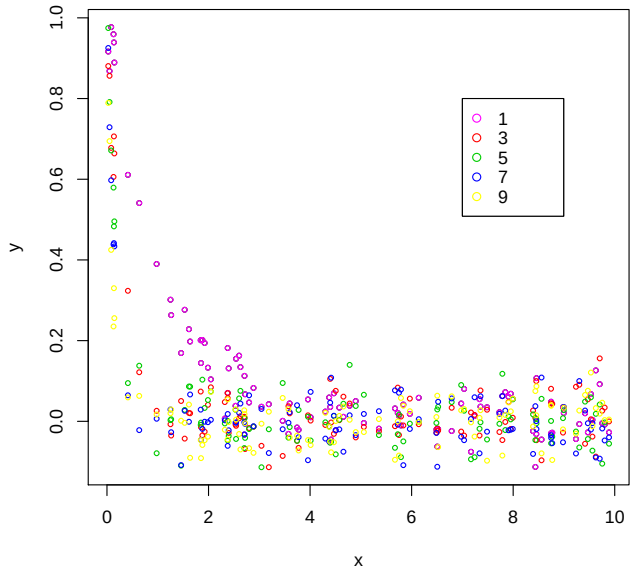

Figure 4.1: *Plots of* $y_i = \beta_1 e^{-x_i \beta_2} + \epsilon_i$ *with* $n = 100$, $\sigma = .05$ *and the selected* β_2 *values*

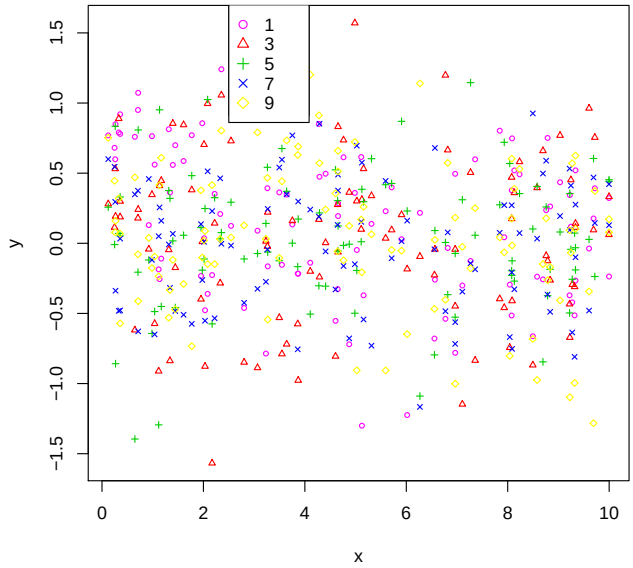

Figure 4.2: *Plots of* $y_i = \beta_1 e^{-x_i \beta_2} + \epsilon_i$ *with* $n = 100$, $\sigma = .5$ *and the selected* β_2 *values*

values of β	$(\beta_1, \beta_2) = (1, 1)$	$(\beta_1, \beta_2) = (1, 3)$ and $(\beta_1, \beta_2) = (1, 5)$	$(\beta_1, \beta_2) = (1, 7)$ and $(\beta_1, \beta_2) = (1, 9)$
$Z_n^*(\beta)$	$\{x_1 : x_9\}$ $\{x_{10} : x_{17}\}$ $\{x_{18} : x_{30}\}$	$\{x_1 : x_9\}$ $\{x_{10} : x_{19}\}$ $\{x_{20} : x_{30}\}$	$\{x_1 : x_9\}$ $\{x_{10} : x_{20}\}$ $\{x_{21} : x_{30}\}$

Table 4.1: *Maximin clustering results for exponential model with* $n = 30$ *and the selected* (β_1, β_2) *values*

values of β	$(\beta_1, \beta_2) = (1, 1)$ $(\beta_1, \beta_2) = (1, 3)$	$(\beta_1, \beta_2) = (1, 5)$	$(\beta_1, \beta_2) = (1, 7)$	$(\beta_1, \beta_2) = (1, 9)$
$Z_n^*(\beta)$	$\{x_1 : x_{19}\}$ $\{x_{20} : x_{72}\}$ $\{x_{73} : x_{100}\}$	$\{x_1 : x_{19}\}$ $\{x_{20} : x_{75}\}$ $\{x_{76} : x_{100}\}$	$\{x_1 : x_{36}\}$ $\{x_{37} : x_{75}\}$ $\{x_{76} : x_{100}\}$	$\{x_1 : x_{43}\}$ $\{x_{44} : x_{77}\}$ $\{x_{78} : x_{100}\}$

Table 4.2: *Maximin clustering results for exponential model with* $n = 100$ *and the selected* (β_1, β_2) *values*

In the implementation algorithm described in Section 3.4, the function

$$g(x) = \begin{cases} 0 & x \leq 0 \\ 1/32\left[-5\left(2x-1\right)^7 + 21\left(2x-1\right)^5 - 35\left(2x-1\right)^3 + 35\left(2x-1\right) + 16\right] & x \in (0,1) \\ 1 & x \geq 1 \end{cases}$$

was used. Using this function g and the maximin crisp clusterings for the $N = 5$ chosen β values, $Z_n^*(\beta)$ can be determined in principle for the remaining $\beta \in B$ according to the formula given in Section 3.4. A grid of 100 β_1 and 100 β_2 values each over the interval $[0, 10]$ was considered for approximation purposes.

As mentioned in Section 3.4, standard linear model projection techniques are used to determine the nearest point to a simulated y_n on M_n^*. Accordingly, 10,000 points on M_n^{**} were considered and the nearest point calculation was approximated by selecting the point

on M_n^{**} with minimal Euclidean distance to y_n. The same approach, rather than a Gauss-Newton method, was also used to solve the nearest point calculation for M_n. Thus, the calculation of the log LRT statistic

$$\Lambda_n = 2 \log \frac{\sup_{\Theta} dP_\theta^n}{\sup_{\Theta_o} dP_\theta^n}(y_n)$$

can be approximated for a simulated y_n.

4.1.1 Data Generation by Perturbing the Proposed Exponential Model

The simulated y_n is obtained using the model $y_n = \beta_1 e^{-x_n \beta_2} + \gamma B_0 + \epsilon_n$, where B_0 is a basis vector for the between cluster lack of fit subspace at selected values of β_2. The power of the test was investigated at particular values of $(\beta_1, \beta_2, \gamma, \sigma)$ with size $\alpha = .05$, and the results for 1000 simulations of each such setting are given in Tables 4.3, 4.4, 4.5 and 4.6.

As noted in Tables 4.1 and 4.2, the maximin clusterings vary according to different points along the expectation surface. However, for computational purposes, it is of interest to determine whether comparable power can be obtained by using a single clustering. To investigate such, the maximin clustering associated with $(\beta_1, \beta_2) = (1, 5)$ was used for all β values. Simulation results corresponding to the same $(\beta_1, \beta_2, \gamma, \sigma)$ settings are summarized in Tables 4.3, 4.4, 4.5 and 4.6. Notably, the power of the test with the single maximin clustering can be less than the power attained when clusterings are allowed to vary according to points on the expectation surface, especially with increasing data size.

In addition, the ML estimates of the parameters $(\beta_1, \beta_2, \gamma, \sigma)$ via standard linear model projection techniques were consistent with parameter settings used to generate the data, especially with the case of multiple maximin clusterings.

(β_1, β_2)	γ	with multiple maximin clusters	with a single maximin cluster
$(1, 1)$	0	.036	.028
	.1	.06	.05
	.2	.19	.13
	.5	.81	.56
	1	1	1
$(1, 3)$	0	.051	.055
	.025	.1	.08
	.05	.22	.17
	.1	.59	.57
	.2	.99	.99
	.5	1	1
$(1, 5)$	0	.056	.052
	.025	.07	.08
	.05	.21	.16
	.1	.63	.59
	.2	.98	.97
	.5	1	1
$(1, 7)$	0	.053	.049
	.025	.07	.08
	.05	.21	.14
	.1	.63	.55
	.2	.98	.97
	.5	1	1
$(1, 9)$	0	.056	.051
	.025	.08	.07
	.05	.21	.14
	.1	.59	.51
	.2	.96	.95
	.5	1	1

Table 4.3: *Power of the test for exponential model with* $n = 30$, $\sigma = .05$, *selected* (β_1, β_2), *and* γ *values*

(β_1, β_2)	γ	with multiple maximin clusters	with a single maximin cluster
$(1, 1)$	0	.043	.044
	.025	.07	.1
	.05	.2	.15
	.1	.48	.36
	.2	.96	.83
	.5	1	1
$(1, 3)$	0	.058	.057
	.025	.08	.08
	.05	.17	.19
	.1	.49	.43
	.2	.96	.93
	.5	1	1
$(1, 5)$	0	.05	.054
	.025	.08	.09
	.05	.18	.16
	.1	.54	.48
	.2	.98	.98
	.5	1	1
$(1, 7)$	0	.052	.059
	.025	.19	.16
	.05	.22	.13
	.1	.47	.43
	.2	.96	.92
	.5	1	1
$(1, 9)$	0	.052	.05
	.025	.13	.12
	.05	.23	.08
	.1	.60	.24
	.2	.96	.71
	.5	1	1

Table 4.4: *Power of the test for exponential model with* $n = 100$, $\sigma = .05$, *selected* (β_1, β_2), *and* γ *values*

(β_1, β_2)	γ	with multiple maximin clusters	with a single maximin cluster
$(1, 1)$	0	.041	.039
	.5	.07	.11
	1	.33	.33
	2	.9	.9
	3	1	1
$(1, 3)$	0	.058	.053
	.2	.06	.07
	.5	.23	.19
	1	.64	.55
	2	.98	.97
	3	1	1
$(1, 5)$	0	.053	.047
	.2	.06	.08
	.5	.2	.18
	1	.61	.52
	2	.96	.97
	3	1	1
$(1, 7)$	0	.055	.051
	.2	.06	.09
	.5	.2	.18
	1	.59	.5
	2	.96	.95
	3	.99	1
$(1, 9)$	0	.05	.046
	.2	.06	.08
	.5	.2	.17
	1	.58	.51
	2	.96	.95
	3	.99	1

Table 4.5: *Power of the test for exponential model with* $n = 30$, $\sigma = .5$, *selected* (β_1, β_2), *and* γ *values*

(β_1, β_2)	γ	with multiple maximin clusters	with a single maximin cluster
$(1, 1)$	0	.059	.04
	.2	.06	.08
	.5	.16	.11
	1	.37	.37
	2	.87	.83
	3	.98	.98
$(1, 3)$	0	.057	.05
	.2	.12	.08
	.5	.24	.21
	1	.49	.42
	2	.98	.93
	3	1	1
$(1, 5)$	0	.05	.055
	.2	.14	.11
	.5	.26	.18
	1	.54	.5
	2	.98	.96
	3	1	1
$(1, 7)$	0	.052	.049
	.1	.13	.09
	.2	.25	.19
	1	.6	.53
	2	.98	.96
	3	1	1
$(1, 9)$	0	.059	.047
	.2	.18	.09
	.5	.27	.2
	1	.61	.39
	2	.97	.86
	3	1	1

Table 4.6: *Power of the test for exponential model with* $n = 100$, $\sigma = .5$, *selected* (β_1, β_2), *and* γ *values*

4.1.2 Data Generation using a Functionally Different Model than the Proposed Exponential Model

The model $y_n = \frac{\alpha}{1+x_n} + \epsilon_n$ is used to generate the response vector. This data generator is plotted in Figures 4.4 and 4.3 for selected (α, σ) values when $n = 100$. Note that, other than the simulated y_n, the simulation parameters are the same as the previous case. The power of the test was investigated at particular values of (α, σ) with nominal size .05, and the results for 1000 simulations of each such setting are given in Tables 4.7 and 4.8.

For comparison, the power of the test based on using a single maximin clustering was considered. In particular, the clustering associated with $(\beta_1, \beta_2) = (1, 5)$ was used. Notably, for this case, the power of the test with the single clustering is generally better than the power attained when clusterings are allowed to vary according to points on the expectation surface. Detecting lack of fit of the proposed model based on multiple maximin clusterings or a single maximin clustering is also of general interest. For this purpose, a simple Bonferroni adjusted multiple testing procedure is proposed and simulation results are summarized in the same table. As depicted in the multiple testing column of Tables 4.7 and 4.8, the multiple test is reasonably successful in detecting departures from the proposed model.

For the case of $\sigma = .05$, most of the observations in the above paragraph apply as shown in Table 4.8. However, the empirical power results for the cases of $\alpha = .1$ and $\alpha = .3$ do not follow the same pattern as indicated for the other α values. When $\alpha = .1$, the power of the test with multiple maximin clusterings is better than the corresponding power with a single maximin cluster. This can be justified by comparing Figure 4.1 with the black color line in Figure 4.4. When $\alpha = .3$, the powers of the tests with multiple maximin clusterings and with a single maximin cluster are very low. One can see that the green color line in Figure 4.4 does not deviate substantially with plots in Figure 4.1 for all selected β_2 values.

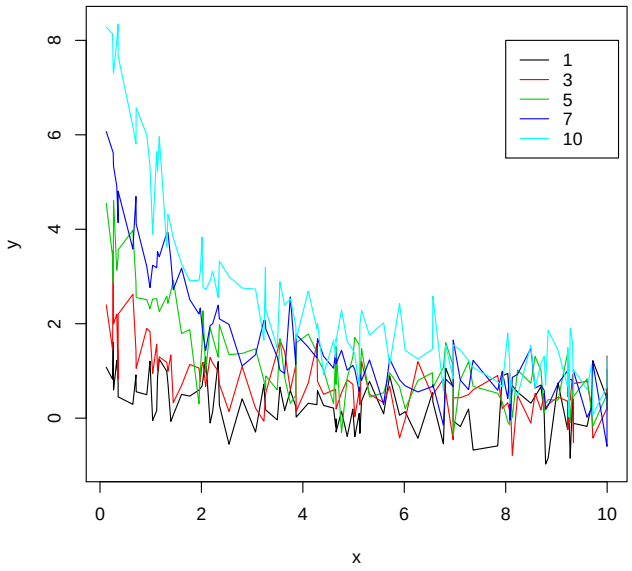

Figure 4.3: *Plots of the data generator* $y_i = \frac{\alpha}{1+x_i} + \epsilon_i$ *with* $n = 100$, $\sigma = .5$ *and selected* α *values*

α	with multiple maximin clusters	with a single maximin cluster	with multiple testing
1	.07	.08	.05
3	.15	.36	.27
5	.22	.77	.7
7	.34	.98	.93
10	.51	1	1

Table 4.7: *Power of the test for exponential model with data generator* $y_i = \frac{\alpha}{1+x_i} + \epsilon_i$, $n = 100$, $\sigma = .5$ *and selected* α *values*

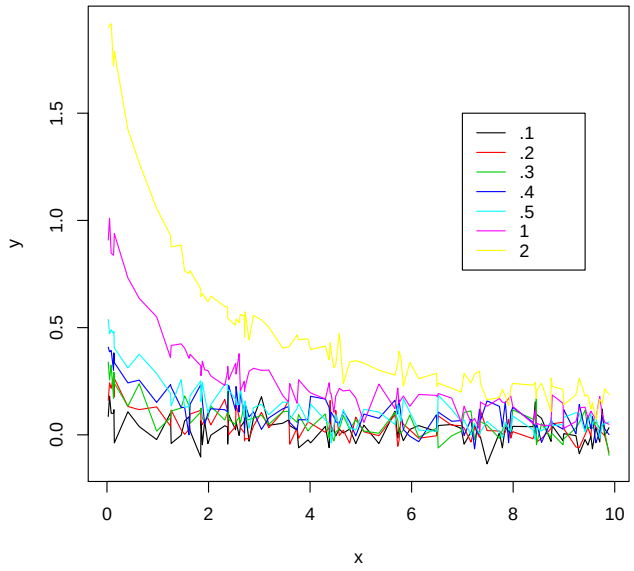

Figure 4.4: *Plots of the data generator* $y_i = \frac{\alpha}{1+x_i} + \epsilon_i$ *with* $n = 100$, $\sigma = .05$ *and selected* α *values*

α	with multiple maximin clusters	with a single maximin cluster	with multiple testing
.1	.33	.19	.31
.2	.07	.43	.33
.3	.04	.09	.07
.4	.29	.63	.73
.5	.21	.56	.43
1	.4	.87	.86
2	.68	1	1

Table 4.8: *Power of the test for exponential model with data generator* $y_i = \frac{\alpha}{1+x_i} + \epsilon_i$, $n = 100$, $\sigma = .05$ *and selected* α *values*

4.1.3 Comparison with a Non-Maximin Cluster and the Ciprian and Ruppert (CR) Test

In this section, the tests based on maximin clusterings are compared to a likelihood based approach suggested by Ciprian and Ruppert(2004). Ciprian and Ruppert proposed a likelihood and restricted likelihood ratio test for lack of fit of a nonlinear regression function. In their approach, the proposed model under the null hypothesis is approximated by using a first order Taylor series around the MLEs of the proposed model parameters. The alternative (full) model is constructed nonparametrically using penalized splines and the linearized version of the proposed model, which leads to a semiparametric model. In addition, the error distribution assumed to be Normal with homogeneous variance.

Testing for an exponential regression function (proposed model), $y_i = \gamma_1 + \delta_1 \exp(\delta_2 x_i) + \epsilon_i$, was one of the examples in the Ciprian and Ruppert(2004) paper. They have studied the empirical power results with parameter settings $\gamma_1 = 1$, $\delta_1 = 1$, $\delta_2 = -1$ and $\sigma = .05$ when $n = 100$. Uniformly distributed predictor values for x on the interval $[0, 1]$ were used. MLEs of the parameters of the proposed model are obtained using the Gauss-Newton procedure in standard software packages. A piecewise constant spline with 15 knots was used when constructing the alternative model. The simulated y_n vector is obtained via the model $y_i = \gamma_1 + \delta_1 \exp(\delta_2 x_i + d x_i^2) + \epsilon_i$ and plotted in Figure 4.5 for selected values of d. The power of the test was investigated with nominal size .05 and empirical power results with 10,000 simulations are given in Table 4.9. The restricted likelihood ratio test statistic as given by Ciprian and Ruppert was used for this purpose.

Note that above parameter settings give a shifted version of the exponential decay model, $y_i = \beta_1 \exp(-\beta_2 x_i) + \epsilon_i$, when $\beta_2 = 1$ with corresponding n and σ values. Therefore, the data generator used by Ciprian and Ruppert gives another different simulated y_n vector to test the lack of fit of the proposed exponential decay model with the proposed test. However,

maximin cluster selection depends upon the values of the predictor variable x. Previously we have used Uniformly distributed predictor values for x on the interval $[0, 10]$ but Ciprian and Ruppert used the interval $[0, 1]$. Notably, one can use the same maximin clusters that are given in Table 4.2 for selected points on the expectation surface. However, values of the predictor variable x are downscaled by .1 for this purpose. Empirical power results with 10,000 simulations and .05 nominal size are summarized in Table 4.9. As Table 4.9 indicates, the power of the Ciprian and Ruppert test can be considerably less than the power of the proposed test, especially with smaller d values. This observation applies to any of the cluster based testing procedures (i.e. multiple maximin clusters, single maximin cluster, or even with the Bonferroni adjusted multiple testing approach), noting Figures 4.5 and 4.6.

It is of interest to compare the power of the proposed lack of fit test with the power of the Ciprian and Ruppert test for other parameter settings of the data generator, e.g. $\delta_2 = (-3, -5, -7, -9)$. This additional information was not available in the Ciprian and Ruppert paper. In fact, convergence problems were noted when obtaining MLEs of the parameters of the proposed model with the Gauss-Newton method for the other δ_2 settings. Alternatively, a grid search method via standard linear model projection techniques was used to obtain MLEs of the parameters of the proposed model. Accordingly, power results were obtained for the Ciprian and Ruppert test for the other parameter settings. The empirical power results for this test and the proposed lack of fit tests were obtained with 10,000 simulations and .05 nominal size and are recorded in Tables 4.10 and 4.11. As one can note from these tables, the proposed tests are successful in detecting lack of fit of the proposed model. In fact, the proposed lack of fit tests are superior to the Ciprian and Ruppert test across all parameter settings.

Constructing alternative models based on maximin clusterings of observations in predictor space is the approach taken thus far. Alternatively, one might select a structured but non-

maximin cluster based on values in predictor space and then construct the alternative model for use in the proposed test procedure. Since the x values are Uniformly distributed over the interval $[0, 1]$ it may be reasonable to use the cut off points $\frac{1}{3}$ and $\frac{2}{3}$ to create 3 clusters and carryout the proposed test as in previous cases. Simulation results with such a selected clustering are summarized in Tables 4.9, 4.10 and 4.11. It is evident from these tables that the empirical power of the proposed test can be considerably less with such a non-maximin cluster, compared to cluster selection based on the maximin power clustering criterion to construct alternative models. However, note that the empirical power of the proposed test with such a non-maximin cluster is still better than power results for CR test.

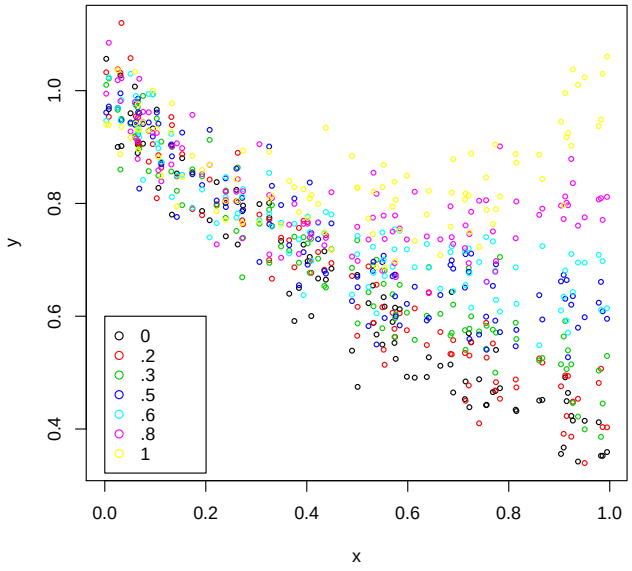

Figure 4.5: *Plots of the data generator* $y_i = \gamma_1 + \delta_1 \exp(\delta_2 x_i + d x_i^2) + \epsilon_i$ *with* $n = 100$, $\sigma = .05$, $(\gamma_1, \delta_1, \delta_2) = (1, 1, -1)$ *and selected* d *values*

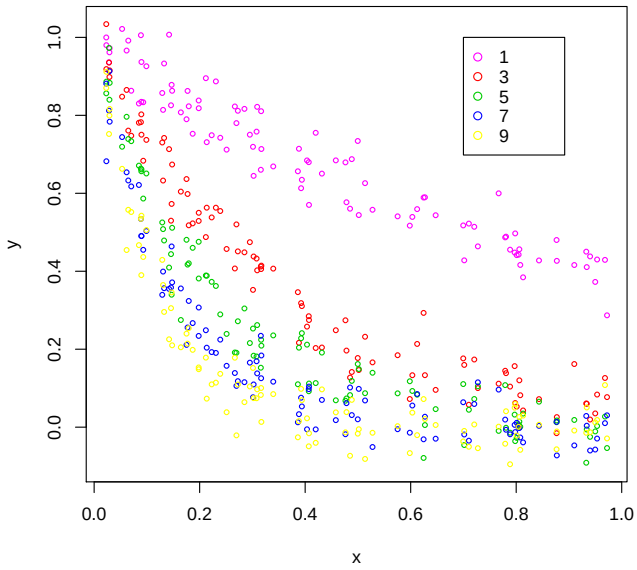

Figure 4.6: *Plots of* $y_i = \beta_1 e^{-x_i \beta_2} + \epsilon_i$ *with* X *from* $U(0,1)$, $n = 100$, $\sigma = .05$ *and selected* β_2 *values to do maximin clusterings*

(δ_2, d)	with multiple maximin clusters	with a single maximin cluster	with multiple testing	with a non-maximin cluster	CR test
$(-1, 0)$	.041	.038	.037	.04	.048
$(-1, .2)$	.29	.21	.25	.2	.067
$(-1, .3)$	.67	.7	.79	.8	.072
$(-1, .5)$	.79	.84	.93	.98	.134
$(-1, .6)$	.99	.86	.99	.99	.223
$(-1, .8)$	.99	.99	.99	.96	.872
$(-1, 1)$	1	1	1	1	1

Table 4.9: *Power of the test for exponential model with data generator* $y_i = \gamma_1 + \delta_1 \exp(\delta_2 x_i + d x_i^2) + \epsilon_i$, $n = 100$, $\sigma = .05$, $\delta_2 = -1$ *and selected* d *values*

(δ_2, d)	with multiple maximin clusters	with a single maximin cluster	with multiple testing	with a non-maximin cluster	CR test
$(-3, 0)$	.04	.05	.03	.05	.047
$(-3, .3)$	.13	.09	.07	.08	.059
$(-3, .5)$	.22	.18	.17	.12	.063
$(-3, .6)$	.32	.22	.23	.15	.09
$(-3, .8)$	.61	.53	.52	.27	.157
$(-3, 1)$	.78	.76	.74	.53	.372
$(-3, 2)$	1	1	1	1	.761
$(-5, 0)$	.05	.05	.03	.07	.049
$(-5, .8)$	.08	.06	.03	.07	.051
$(-5, 1)$	.07	.06	.05	.07	.065
$(-5, 1.5)$	.2	.17	.15	.08	.089
$(-5, 2)$	.54	.39	.43	.2	.22
$(-5, 2.5)$	.95	.79	.89	.44	.356
$(-5, 3)$	1	.99	1	.93	.725

Table 4.10: *Power of the test for exponential model with data generator* $y_i = \gamma_1 + \delta_1 \exp(\delta_2 x_i + d x_i^2) + \epsilon_i$, $n = 100$, $\sigma = .05$, $\delta_2 = (-3, -5)$ *and selected* d *values*

(δ_2, d)	with multiple maximin clusters	with a single maximin cluster	with multiple testing	with a non-maximin cluster	CR test
$(-7, 0)$	.03	.05	.03	.02	.051
$(-7, 2)$	.06	.07	.05	.04	.067
$(-7, 3)$	.21	.2	.16	.13	.098
$(-7, 3.5)$	.37	.37	.27	.29	.124
$(-7, 4)$	.66	.67	.54	.52	.364
$(-7, 4.5)$	.96	.95	.93	.78	.543
$(-7, 5)$	1	1	1	.98	.739
$(-9, 0)$	.04	.03	.03	.02	.046
$(-9, 4)$	.09	.11	.1	.01	.057
$(-9, 5)$	.18	.27	.21	.03	.076
$(-9, 5.5)$	.23	.39	.33	.07	.097
$(-9, 6)$	.49	.64	.53	.29	.222
$(-9, 6.5)$	.78	.87	.83	.65	.557
$(-9, 7)$	.99	.98	.99	.94	.754

Table 4.11: *Power of the test for exponential model with data generator* $y_i = \gamma_1 + \delta_1 \exp(\delta_2 x_i + d x_i^2) + \epsilon_i$, $n = 100$, $\sigma = .05$, $\delta_2 = (-7, -9)$ *and selected* d *values*

4.2 Example 2: Cosine Model

Sine and cosine functions are useful in studies of general periodic patterns. Those are applicable when modeling recurring phenomena such as sound or light waves. Some important applications can be found in physics, electrical engineering and biology. In physics, sine and cosine functions are used to describe the simple harmonic motion, which models many natural phenomena, such as the movement of a mass attached to a spring and, for small angles, the pendular motion of a mass hanging by a string. Some biological processes follows predictable patterns that repeat approximately once every 24 hours which leads to circadian data/rhythms. For example, pressure against the walls of the blood vessels of a patient, doppler signals for carotid and femoral arteries etc.

In particular, a cosine regression function is considered and is given by $f(x, \beta) = \beta_1 \cos \beta_2 x$ with the usual notations, β_1 for amplitude and β_2 for frequency. Uniformly distributed predictor values for x on the interval $[0, 10]$ were used to analyze a data size of 100. The error distribution was assumed to be Normal when investigating power of the proposed LRT for detecting between-cluster lack of fit. The cover, overlapping subsets and ordered crisp clusterings were created as described in Section 4.1. The maximin power clustering criterion, as described in Section 3.2, was then used to find the best crisp atom $Z_n^*(\beta_i)$ at each of $N = 9$ chosen β values. For illustration purposes, the values of $\beta = (\beta_1, \beta_2)$ were selected as $\{(1, .3), (1, .6), (1, .9), (1, 1.2), (1, 1.5), (1, 1.8), (1, 2.1), (1, 2.4), (1, 2.7)\}$. The plots of the nonlinear regression function $f(x, \beta) = \beta_1 \cos \beta_2 x$ for the selected (β_1, β_2) values are depicted in Figure 4.7.

Similar to the exponential model, the dimension of $C(B_{Z_n(\beta)})$ is equal to one for all corresponding nonlinear between-cluster lack of fit subspaces. Table 4.12 gives the crisp clusterings $Z_n^*(\beta_i)$ for each of the $N = 9$ selected β values. This table indicates that the maximin clusterings do indeed vary according to different points along the expectation surface as in-

dexed by the parameters (β_1, β_2). In fact, we get different clusterings for all selected β values.

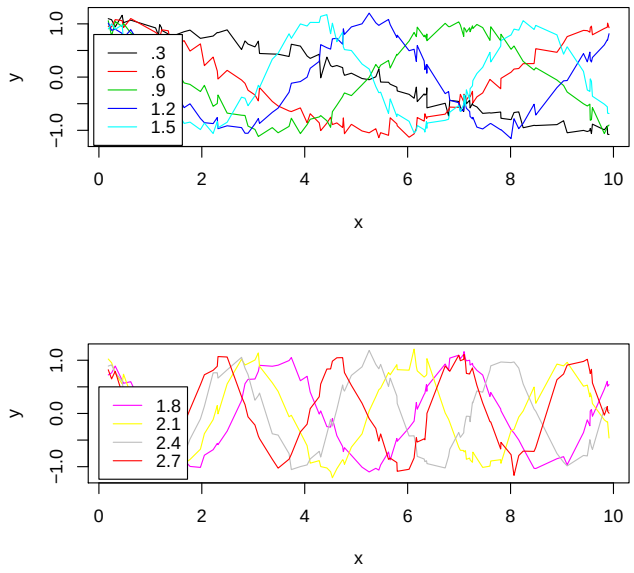

Figure 4.7: *Plots of* $y_i = \beta_1 \cos \beta_2 x_i + \epsilon_i$ *with* $n = 100$, $\sigma = .1$ *and selected* β_2 *values*

values of β	$Z_n^*(\beta)$
$(\beta_1, \beta_2) = (1, .3)$	$\{x_1 : x_{25}\}, \{x_{26} : x_{72}\}, \{x_{73} : x_{100}\}$
$(\beta_1, \beta_2) = (1, .6)$	$\{x_1 : x_{21}\}, \{x_{22} : x_{72}\}, \{x_{73} : x_{100}\}$
$(\beta_1, \beta_2) = (1, .9)$	$\{x_1 : x_{19}\}, \{x_{20} : x_{74}\}, \{x_{75} : x_{100}\}$
$(\beta_1, \beta_2) = (1, 1.2)$	$\{x_1 : x_{44}\}, \{x_{45} : x_{76}\}, \{x_{77} : x_{100}\}$
$(\beta_1, \beta_2) = (1, 1.5)$	$\{x_1 : x_{37}\}, \{x_{38} : x_{80}\}, \{x_{81} : x_{100}\}$
$(\beta_1, \beta_2) = (1, 1.8)$	$\{x_1 : x_{30}\}, \{x_{31} : x_{78}\}, \{x_{79} : x_{100}\}$
$(\beta_1, \beta_2) = (1, 2.1)$	$\{x_1 : x_{37}\}, \{x_{38} : x_{88}\}, \{x_{89} : x_{100}\}$
$(\beta_1, \beta_2) = (1, 2.4)$	$\{x_1 : x_{19}\}, \{x_{20} : x_{76}\}, \{x_{77} : x_{100}\}$
$(\beta_1, \beta_2) = (1, 2.7)$	$\{x_1 : x_{20}\}, \{x_{21} : x_{73}\}, \{x_{74} : x_{100}\}$

Table 4.12: *Maximin clustering results for cosine model with* $n = 100$ *and the selected* (β_1, β_2) *values*

Using the g function given in section 4.1 and the maximin crisp clusterings for the $N = 9$ chosen β values, $Z_n^*(\beta)$ can be obtained for the remaining $\beta \in B$ according to the formula

given in Section 3.4. A grid of 50 β_1 and 50 β_2 values each over the interval $[0, 3]$ was considered for approximation purposes.

The parameter estimation under the proposed model and the constructed full model were obtained via standard linear model projection techniques and the log LRT statistic was computed for a simulated y_n, similar to the exponential model case.

4.2.1 Data Generation by Perturbing the Proposed Cosine Model

The simulated y_n is obtained using the model $y_n = \beta_1 \cos(\beta_2 x_n) + \gamma B_0 + \epsilon_n$, where B_0 is a basis vector for between cluster lack of fit subspace at selected values of β. The power of the test was investigated at particular values of $(\beta_1, \beta_2, \gamma, \sigma)$ with size $\alpha = .05$, and the results for 1000 simulations of each such setting are given in Tables 4.13 and 4.14. The empirical power of the proposed LRT for detecting nonlinear between-cluster lack of fit is presented for both multiple maximin clusterings and a single maximin cluster i.e. the maximin clustering associated with $(\beta_1, \beta_2) = (1, 1.5)$. Notably, the power of the test with the single maximin clustering can be considerably less than the power attained when clusterings are allowed to vary according to points on the expectation surface, especially with the (β_1, β_2) settings $\{(1, .3), (1, .6), (1, 2.1), (1, 2.7)\}$. However, power results for the $(\beta_1, \beta_2) = (1, .9)$ case were very low for both multiple maximin clusterings and a single maximin cluster. Parameter estimates for $(\beta_1, \beta_2, \gamma, \sigma)$ under the proposed cosine model and constructed full model were consistent with parameter settings used to simulate y_n, especially for the multiple maximin clustering case. The proposed cosine model is plotted in Figure 4.8 with the parameter settings $(\beta_1, \beta_2, \sigma) = \{(1, .9, .5), (1, 1.5, .5)\}$. The constructed full model is plotted in Figure 4.9 with the parameter settings $(\beta_1, \beta_2, \sigma) = (1, .9, .5)$ and selected γ values. By comparing Figures 4.8 and 4.9, the low power for the proposed LRT can be explained i.e the proposed model and simulated data agree closely with each other.

(β_1, β_2)	γ	σ	with multiple maximin clusters	with a single maximin cluster
$(1, .3)$	0	.5	.03	0
	.5		.08	0
	1		.33	0
	2		.88	.02
	3		.93	.02
$(1, .6)$	0	.1	.08	.08
	.025		.12	.1
	.25		.58	.24
	.5		1	.47
	1		1	.78
	-.025		.07	.09
	-.25		.66	.23
	-.5		1	.67
	-1		1	.66
$(1, .9)$	0	.5	.04	.07
	1		.3	.08
	2		.46	.19
	3		.25	.4
	-.5		.11	.07
	-2		.18	.07
	-3		.08	.06
$(1, 1.2)$	0	.5	.03	.06
	1		.48	.34
	2		.96	.82
	-.5		.17	.11
	-1		.5	.38
	-2		.99	.84
	-3		1	.99
$(1, 1.5)$	0	.5	.05	.05
	1		.5	.5
	2		.95	.95
	-.5		.13	.13
	-1		.43	.43
	-2		.95	.95

Table 4.13: *Power of the test for cosine model with* $n = 100$ *selected* σ, (β_1, β_2), *and* γ *values*

(β_1, β_2)	γ	σ	with multiple maximin clusters	with a single maximin cluster
$(1, 1.8)$	0	.5	.07	.09
	1		.47	.38
	2		.99	.76
	-.5		.1	.12
	-1		.5	.36
	-2		.96	.8
$(1, 2.1)$	0	.1	.03	.04
	-.025		.03	.05
	-.1		.15	.05
	-.5		.99	.15
	-1		1	.27
$(1, 2.4)$	0	.5	.07	.03
	.5		.22	.14
	1		.44	.37
	2		.98	.91
$(1, 2.7)$	0	.5	.05	.07
	.5		.16	.14
	1		.46	.26
	2		.98	.68

Table 4.14: *Power of the test for cosine model with* $n = 100$ *selected* σ, (β_1, β_2), *and* γ *values*

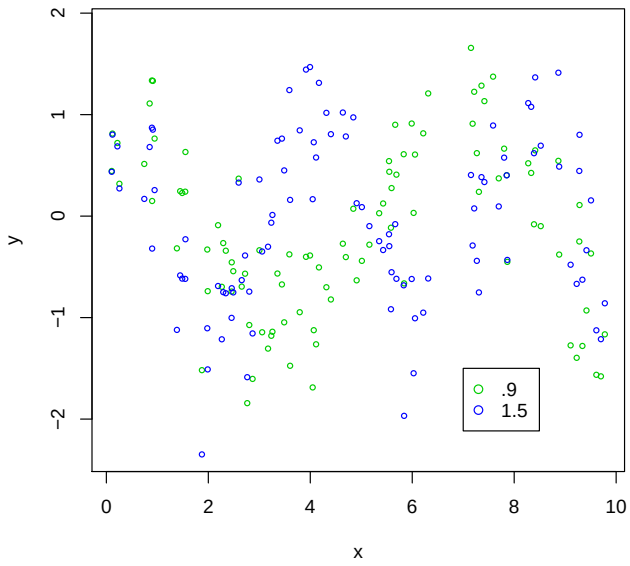

Figure 4.8: *Plots of* $y_i = \beta_1 \cos \beta_2 x_i + \epsilon_i$ *with* $n = 100$, $\sigma = .5$ *and* $\beta_2 = (.9, 1.5)$

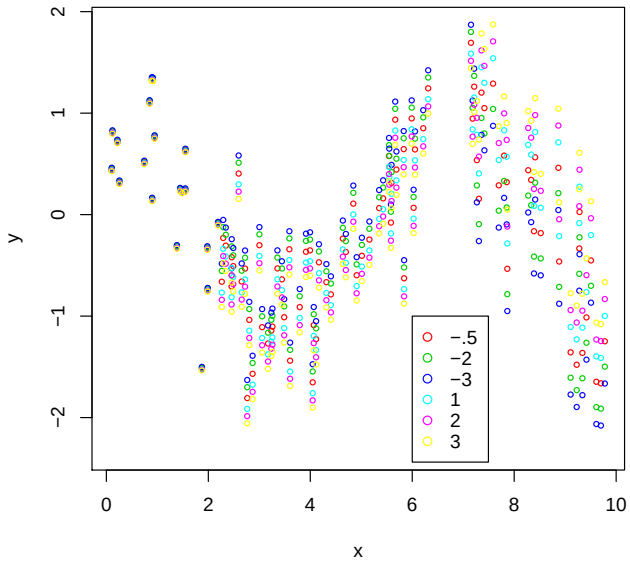

Figure 4.9: *Plots of the data generator* $y_i = \cos(.9x_i) + \gamma B_0 + \epsilon_i$ *with* $n = 100$, $\sigma = .5$ *and selected* γ *values*

4.2.2 Data Generation using a Functionally Different Model than the Proposed Cosine Model

The model $y_n = \beta_3 x_n + \cos(\beta_4 x_n) + \epsilon_n$ is used to generate the response vector. One can note that this model adds a trend component to the proposed model along with different frequencies. The simulated data is depicted in Figures 4.10 and 4.11 for selected (β_3, β_4) values with $\sigma = .1$ and $n = 100$. The power of the test was investigated at chosen values of (β_3, β_4) with nominal size .05, and the results for 1000 simulations of each such settings are given in Tables 4.15 and 4.16.

As described in Section 4.1.2, a simple Bonferroni adjusted multiple testing procedure for detecting nonlinear between-cluster lack of fit can be useful as well. Note the empirical power of the proposed LRT when $(\beta_3, \beta_4) = (.01, 1.65)$ was very low (see Table 4.15). The light blue color curve in Figure 4.10 represents the corresponding generated data and the same color curve in Figure 4.7 gives one of the proposed models. These two curves agree closely with each other and thus low power results for the tests using either multiple maximin clusterings or a single maximin cluster. Similar arguments can be made to justify the low power for the tests with cases $(\beta_3, \beta_4) = \{(.01, .75), (.01, 2.25), (.01, 2.55)\}$.

One can observe that all plots in Figure 4.11 add some trend to the corresponding proposed model in Figure 4.7. Furthermore, some extra bending or lengthening can be noticed at the end of the curve. Therefore, as indicated in Table 4.16, excellent power results with the multiple testing procedure, as can be justified by comparing Figures 4.7 and 4.11.

4.3 Example 3: Michaelis-Menten Model

The Michaelis-Menten equation is widely used in enzyme kinetic assays according to '*curvefit.com*'. The enzyme velocity (Y) is proportional to the concentration (X) of enzyme-

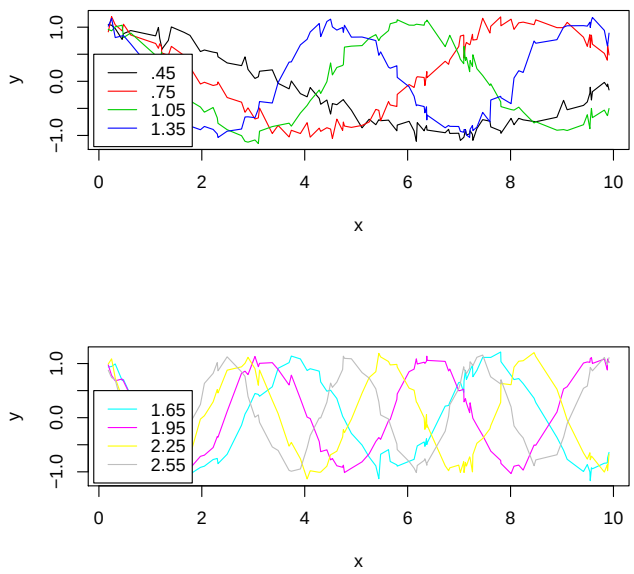

Figure 4.10: *Plots of the data generator* $y_i = \beta_3 x_i + \cos(\beta_4 x_i) + \epsilon_i$ *with* $n = 100$, $\sigma = .1$, $\beta_3 = .01$ *and selected* β_4 *values*

(β_3, β_4)	with multiple maximin clusters	with a single maximin cluster	with multiple testing
$(.01, .45)$	.68	.75	.88
$(.01, .75)$	.02	.24	.2
$(.01, 1.05)$	.79	.97	.97
$(.01, 1.35)$	.21	.2	.2
$(.01, 1.65)$	.02	.02	.03
$(.01, 1.95)$	.47	.01	.34
$(.01, 2.25)$	.06	.17	.11
$(.01, 2.55)$	.03	.41	.23

Table 4.15: *Power of the test for cosine model with data generator* $y_i = \beta_3 x_i + \cos(\beta_4 x_i) + \epsilon_i$, $n = 100$, $\sigma = .1$, $\beta_3 = .01$ *and selected* β_4 *values*

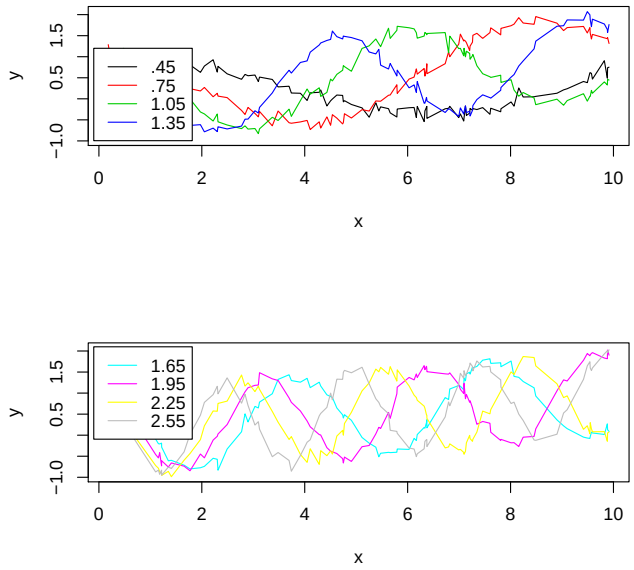

Figure 4.11: *Plots of the data generator* $y_i = \beta_3 x_i + \cos(\beta_4 x_i) + \epsilon_i$ *with* $n = 100$, $\sigma = .1$, $\beta_3 = .1$ *and selected* β_4 *values*

(β_3, β_4)	with multiple maximin clusters	with a single maximin cluster	with multiple testing
$(.1, .45)$	1	.03	1
$(.1, .75)$	0	1	1
$(.1, 1.05)$	0	1	1
$(.1, 1.35)$	1	1	1
$(.1, 1.65)$	1	1	1
$(.1, 1.95)$	0	1	1
$(.1, 2.25)$	0	1	1
$(.1, 2.55)$	.86	1	1

Table 4.16: *Power of the test for cosine model with data generator* $y_i = \beta_3 x_i + \cos(\beta_4 x_i) + \epsilon_i$, $n = 100$, $\sigma = .1$, $\beta_3 = .1$ *and selected* β_4 *values*

substrate complexes and the relationship is given by $y_i = \frac{\gamma_0 x_i}{\gamma_1 + x_i} + \epsilon_i$. This model is sometimes called a rectangular hyperbola or a binding isotherm. Note that this plot on log transformed concentration values becomes sigmoidal. A similar model is used in determining the relationship between density of crop planting and crop yield in agriculture.

A scaled version of the Michaelis-Menten equation, $y_i = \frac{x_i}{\beta_1 + \beta_2 x_i} + \epsilon_i$ was used, where $\beta_1 = \frac{\gamma_1}{\gamma_0}$ and $\beta_2 = \frac{1}{\gamma_0}$. Uniformly distributed predictor values for x on the interval $[0, 100]$ were used to analyze the data size of 100. Similar to the other two examples, the error distribution was assumed to be Normal when investigating power of the proposed LRT for detecting nonlinear between-cluster lack of fit. One run of data size 100 was generated in R (cf Appendix C). As discussed in Chapter 3, a cover C_n was specified by dividing $[0, 100]$ into the cells $\{[0, 20), [20, 40), [40, 60), [60, 80), [80, 100]\}$, with associated overlapping subsets given by $F_n = \{F_{n1}, F_{n2}, F_{n3}\} = \{[0, 40), [20, 80), [60, 100]\}$. Again, a subset comprised of *ordered* partitions from the full collection $\kappa_{n,o}$ was utilized to facilitate computational feasibility for the simulation. In particular, for the specified cover C_n, suppose in the even cells $[20, 40)$ and $[60, 80)$ there are k and l predictor values and denote the ordered values by $\{x_{11}, x_{12}, ..., x_{1k}\}$ and $\{x_{21}, x_{22}, ..., x_{2l}\}$, respectively. Now group all predictor values $< x_{1m}$ in even cell $[20, 40)$ with odd cell $[0, 20)$, while all predictor values $\geq x_{1m}$ in even cell $[20, 40)$ are grouped with odd cell $[40, 60)$ for $m = 1, 2, ..., k$. In addition, if $m = k$ group all predictor values $\leq x_{1m}$ in even cell $[20, 40)$ with odd cell $[0, 20)$. Similarly, group all predictor values $< x_{2n}$ in even cell $[60, 80)$ with odd cell $[40, 60)$, while all predictor values $\geq x_{2n}$ in even cell $[60, 80)$ are grouped with odd cell $[80, 100]$ for $n = 1, 2, ..., l$. In addition, if $n = l$ group all predictor values $\leq x_{2n}$ in even cell $[60, 80)$ with odd cell $[40, 60)$. This leads to $(k + 1)(l + 1)$ different groupings (crisp clusterings) of the predictor values based on the 5 cells chosen. Based on the specified cover C_n, 459 such crisp clusterings were produced for the data size 100. The maximin power clustering criterion, as described in Section 3.2, was then used to find the best crisp atom $Z_n^*(\beta_i)$ at each of $N = 9$ chosen β values. For illustration purposes, the values of

$\beta = (\beta_1, \beta_2)$ were selected as $\{(.1, .003), (.1, .006), (.1, .009), (.2, .003), (.2, .006), (.2, .009)\}$ and $\{(.3, .003), (.3, .006), (.3, .009)\}$. The plots of the nonlinear regression function $f(x, \beta) = \frac{x}{\beta_1 + \beta_2 x}$ for the selected $(\beta_1, \beta_2, \sigma)$ values are depicted in Figures 4.12 and 4.13.

Similar to the previous two models, the dimension of $C(B_{Z_n(\beta)})$ is equal to one for all corresponding nonlinear between-cluster lack of fit subspaces. Table 4.17 gives the crisp clusterings $Z_n^*(\beta_i)$ for each of the $N = 9$ selected β values. This table indicates that the maximin clusterings vary according to different points along the expectation surface as indexed by the parameters (β_1, β_2). However, such clusterings are nearly identical.

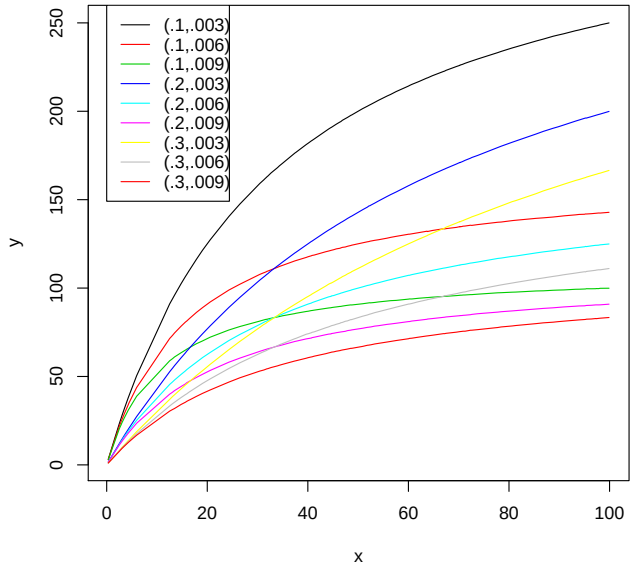

Figure 4.12: *Plots of* $y_i = \frac{x_i}{\beta_1 + \beta_2 x_i} + \epsilon_i$ *with* $n = 100$, $\sigma = .05$ *and selected* (β_1, β_2) *values*

Using the g function given in section 4.1 and the maximin crisp clusterings for the $N = 9$ chosen β values, $Z_n^*(\beta)$ can be obtained for the remaining $\beta \in B$ according to the formula given in Section 3.4. A grid of 125 β_1 values over the interval $[0, .4]$ and 125 β_2 values over the interval $[0, .01]$ were considered for approximation purposes.

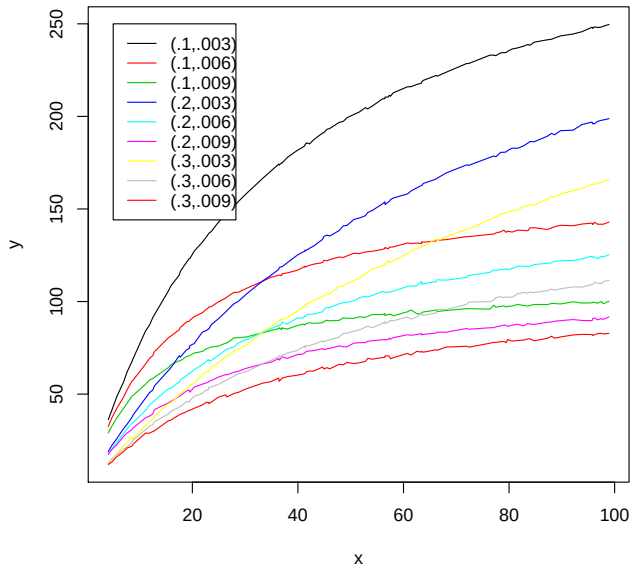

Figure 4.13: *Plots of* $y_i = \frac{x_i}{\beta_1+\beta_2 x_i} + \epsilon_i$ *with* $n = 100$, $\sigma = .5$ *and selected* (β_1, β_2) *values*

values of β	$(\beta_1, \beta_2) = \{(.1, .003), (.1, .006)\}$ $(\beta_1, \beta_2) = \{(.1, .009), (.2, .006)\}$ $(\beta_1, \beta_2) = \{(.2, .009), (.3, .006)\}$ $(\beta_1, \beta_2) = \{(.3, .009)\}$	$(\beta_1, \beta_2) = \{(2, .003), (3, .003)\}$
$Z_n^*(\beta)$	$\{x_1 : x_{22}\}$ $\{x_{23} : x_{73}\}$ $\{x_{74} : x_{100}\}$	$\{x_1 : x_{23}\}$ $\{x_{24} : x_{72}\}$ $\{x_{73} : x_{100}\}$

Table 4.17: *Maximin clustering results for Michaelis-Menten model with* $n = 100$ *and the selected* (β_1, β_2) *values*

The parameter estimation under the proposed model and the constructed full model were obtained via standard linear model projection techniques and log LRT statistic was computed for a simulated y_n, similar to the exponential and cosine model cases. Note that a two dimensional grid search is necessary to estimate parameters since the tangent space (hence the lack of fit subspace) depends on both parameters in the model. Whereas in exponential and cosine models, the tangent space (hence the lack of fit subspace) depends on only one parameter in the model i.e. β_2.

4.3.1 Data Generation by Perturbing the Proposed Michaelis-Menten Model

The simulated y_n is obtained using the model $y_n = \frac{x_n}{\beta_1+\beta_2 x_n} + \gamma B_0 + \epsilon_n$, where B_0 is a basis vector for the between cluster lack of fit subspace at selected values of β. The power of the test was investigated at particular values of $(\beta_1, \beta_2, \gamma, \sigma)$ with size $\alpha = .05$, and the results for 1000 simulations of each such setting are given in Table 4.18. The empirical power of the proposed LRT for detecting nonlinear between-cluster lack of fit is presented for both multiple maximin clusterings and a single maximin cluster i.e. the maximin clustering associated with $(\beta_1, \beta_2) = (.2, .006)$. Notably, the power of the test with the single maximin clustering closely agrees with the power attained when clusterings are allowed to vary according to points on the expectation surface. This observation is consistent with the nearly identical maximin crisp clusterings for selected β values, as indicated in Table 4.17.

4.3.2 Data Generation using a Functionally Different Model than the Proposed Michaelis-Menten Model

The model $y_n = \frac{x_n^2}{\beta_3+\beta_4 x_n^2} + \epsilon_n$ is used to generate the response vector. The simulated data is depicted in Figures 4.14 and 4.15 for selected $(\beta_3, \beta_4, \sigma)$ values when $n = 100$. The power

(β_1, β_2)	γ	with multiple maximin clusters	with a single maximin cluster
$(.1, .003)$	0	.04	.04
	.5	.11	.11
	1	.2	.2
	2	.53	.53
	3	.78	.78
$(.1, .006)$	0	.04	.04
	.5	.12	.12
	1	.24	.24
	2	.67	.67
	3	.85	.85
$(.1, .009)$	0	.04	.04
	.5	.07	.07
	1	.31	.31
	2	.87	.87
	3	1	1
$(.2, .003)$	0	.04	.04
	.5	.06	.07
	1	.21	.23
	2	.77	.73
	3	.98	.95
$(.2, .009)$	0	.04	.04
	.5	.11	.11
	1	.23	.23
	2	.76	.76
	3	.99	.99
$(.3, .003)$	0	.04	.04
	.5	.05	.07
	1	.22	.22
	2	.76	.69
	3	.98	.93
$(.3, .006)$	0	.04	.04
	.5	.11	.11
	1	.21	.21
	2	.57	.57
	3	.96	.96

Table 4.18: *Power of the test for Michaelis-Menten model with* $n = 100$, $\sigma = .5$, *selected* (β_1, β_2), *and* γ *values*

of the test was investigated at chosen values of $(\beta_3, \beta_4, \sigma)$ with nominal size .05, and the results for 1000 simulations of each such setting are given in Tables 4.19 and 4.20. The power results in these tables can be justified by comparing Figures 4.14 and 4.15 with Figures 4.12 and 4.13 i.e plots for data generator are subtantially different from plots for the proposed model for selected parameter settings.

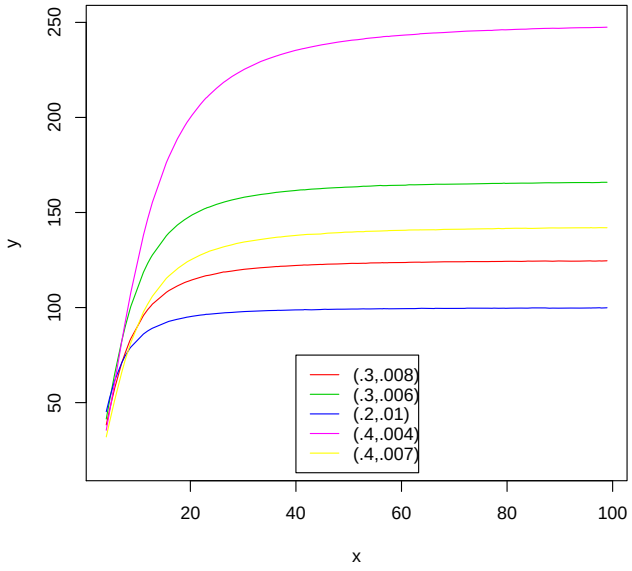

Figure 4.14: *Plots of the data generator* $y_i = \frac{x_i^2}{\beta_3+\beta_4 x_i^2} + \epsilon_i$ *with* $n = 100$, $\sigma = .05$ *and selected* (β_3, β_4) *values*

(β_3, β_4)	with multiple maximin clusters	with a single maximin cluster
$(.2, .01)$	1	1
$(.3, .006)$	1	1
$(.3, .008)$	1	1
$(.4, .004)$	1	1
$(.4, .007)$	1	1

Table 4.19: *Power of the test for Michaelis-Menten model with data generator* $y_i = \frac{x_i^2}{\beta_3+\beta_4 x_i^2} + \epsilon_i$, $n = 100$, $\sigma = .05$ *and selected* (β_3, β_4) *values*

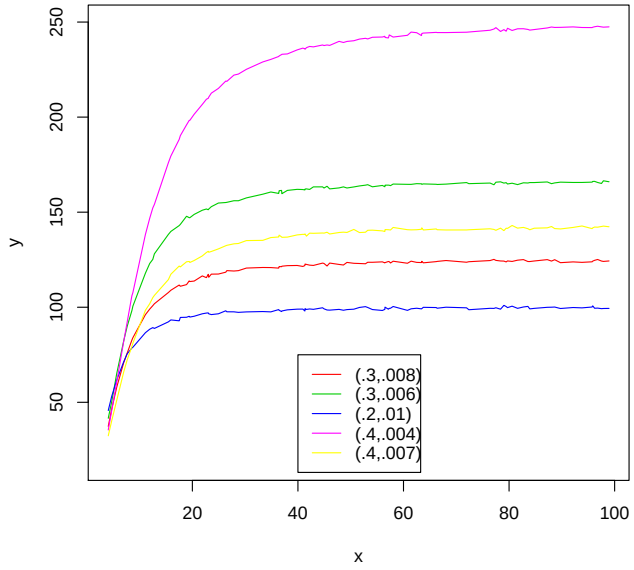

Figure 4.15: *Plots of the data generator* $y_i = \frac{x_i^2}{\beta_3+\beta_4 x_i^2} + \epsilon_i$ *with* $n = 100$, $\sigma = .5$ *and selected* (β_3, β_4) *values*

(β_3, β_4)	with multiple maximin clusters	with a single maximin cluster
$(.2, .01)$	1	1
$(.3, .006)$	1	1
$(.3, .008)$	1	1
$(.4, .004)$	1	1
$(.4, .007)$	1	1

Table 4.20: *Power of the test for Michaelis-Menten model with data generator* $y_i = \frac{x_i^2}{\beta_3+\beta_4 x_i^2} + \epsilon_i$, $n = 100$, $\sigma = .5$ *and selected* (β_3, β_4) *values*

In addition, the model $y_n = \log(\beta_3 + \beta_4 x_n) + \epsilon_n$ was used to generate the response vector. The simulated data is depicted in Figure 4.16 for selected (β_3, β_4) values with $\sigma = .5$ and $n = 100$. The power of the test was investigated at chosen values of (β_3, β_4) with nominal size .05, and the results for 1000 simulations of each such setting are given in Table 4.21. Considering scaling with corresponding parameter setting, the exceptionally low power results in this table can be justified by comparing Figure 4.16 with Figure 4.13 i.e plots for data generator closely agree with plots for proposed model for selected parameter settings.

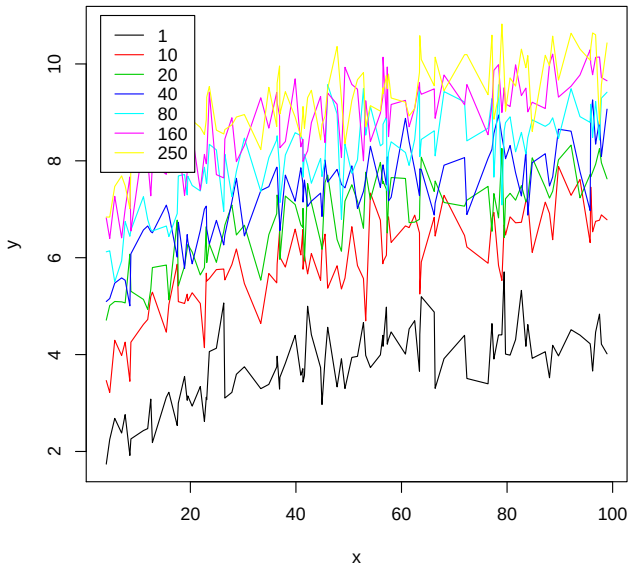

Figure 4.16: *Plots of the data generator* $y_i = \log(\beta_3 + \beta_4 x_i) + \epsilon_i$ *with* $n = 100$, $\sigma = .5$, $\beta_3 = 2$ *and selected* β_4 *values*

(β_3, β_4)	with multiple maximin clusters	with a single maximin cluster
$(2, 1)$	0	0
$(2, 10)$	0	0
$(2, 20)$	0	0
$(2, 40)$	0	0
$(2, 80)$	0	0
$(2, 160)$	0	0
$(2, 250)$	0	0

Table 4.21: *Power of the test for Michaelis-Menten model with data generator* $y_i = \log(\beta_3 + \beta_4 x_i) + \epsilon_i$, $n = 100$, $\sigma = .5$ *and selected* (β_3, β_4) *values*

Chapter 5

Asymptotics

This chapter is devoted to proving the asymptotic non-central Chi-square distribution property of the log LRT statistic for testing $H_o : \gamma = 0$ versus $H_a : \gamma \neq 0$ in a sequence of BCLFEs (i.e. Theorem 1 below). In addition, the use of MMBCLFEs in testing lack of fit is justified in terms of providing optimal power (i.e. Theorem 2 below).

A proof of Theorem 1 is based in part on the requirement that the sequence of BCLFEs be locally asymptotic normal (LAN). A key condition to ensure LAN is that the sequence of BCLFEs satisfy the following definition.

Definition:

A sequence of BCLFEs $E_n\left(x_{n1}, ..., x_{nn}; Z_n\left(\beta\right) \in \kappa_n, \beta \in B\right)$ is *uniformly differentiable in quadratic mean* (i.e. uniform q.m.d.) at θ if for each $1 \leq i \leq n$ and $n \geq 1$ there exists a measurable function $\ell'_{ni,\theta}$ such that

$$\int \left(\sqrt{p_{ni,\theta+\tilde{h}}} - \sqrt{p_{ni,\theta}} - \frac{1}{2}\tilde{h}^T \ell'_{ni,\theta} \sqrt{p_{ni,\theta}} \right)^2 d\mu = o\left(\left\| \tilde{h} \right\|^2 \right)$$

independent of n and i as $\tilde{h} \to 0$ in $\Re^{c+s}$, where $p_{ni,\theta}$ denotes the density of y_{ni}. $\diamondsuit$

Lemma 1 below gives mild conditions under which uniform q.m.d. obtains. Note that $\theta = (\vartheta, \eta)$ where $\vartheta = (\beta, \gamma)$ with dimension c and the dimension of η is s. Define

$\Phi_n(\vartheta) = f_n(\beta) + B_{Z_n(\beta)}\gamma$ for $n \geq 1$. Under conditions of Lemma 1, $\ell'_{ni,\theta} = \frac{\partial \log p_{ni,\theta}}{\partial \theta}$.

Lemma 1:

Let $s_{ni}(y,\theta) = s(y - \Phi_{ni}(\vartheta), \eta)$ for $1 \leq i \leq n$ and $n \geq 1$ where $s(\epsilon,\eta) = \sqrt{p(\epsilon,\eta)}$ and $p(\epsilon,\eta)$ is the error density. Fix $\theta \in \Theta$ and $\delta > 0$ and let $M > 0$ such that for any $\phi = \left(\widetilde{\vartheta}, \widetilde{\eta}\right)$ with $\|\phi - \theta\| < \delta$,

$$\left|\Phi_{ni}\left(\widetilde{\vartheta}\right)\right| \leq M, \left|\frac{\partial \Phi_{ni}\left(\widetilde{\vartheta}\right)}{\partial \vartheta_j}\right| \leq M \text{ and } \left|\frac{\partial^2 \Phi_{ni}\left(\widetilde{\vartheta}\right)}{\partial \vartheta_j \partial \vartheta_l}\right| \leq M,$$

for $j, l = 1, ..., c$, for all n and i. In addition, let

$$T(y) = \sup_{|\xi| < M, \|\zeta - \eta\| < \delta} \Big(|s_1(y - \xi, \zeta)| + |s_{1,1}(y - \xi, \zeta)|$$

$$+ \sum_{k=1}^{s} |s_{1,k+1}(y - \xi, \zeta)| + \sum_{k=1}^{s} \sum_{m=1}^{s} |s_{k+1,m+1}(y - \xi, \zeta)| \Big)$$

where $s_j(z_1, ..., z_{s+1}) = \frac{\partial s(z_1,...,z_{s+1})}{\partial z_j}$ and $s_{j,l}(z_1, ..., z_{s+1}) = \frac{\partial^2 s(z_1,...,z_{s+1})}{\partial z_j \partial z_l}$, and suppose $\int T^2 d\mu < \infty$. Then the corresponding sequence of BCLFEs is uniformly q.m.d. at θ. $\diamondsuit$

Proof:

Note that

$$\frac{\partial s_{ni}(y,\theta)}{\partial \vartheta_j} = s_1(y - \Phi_{ni}(\vartheta), \eta)\left(\frac{-\partial \Phi_{ni}(\vartheta)}{\partial \vartheta_j}\right), \text{ for } j = 1, ..., c,$$

$$\frac{\partial s_{ni}(y,\theta)}{\partial \eta_k} = s_{k+1}(y - \Phi_{ni}(\vartheta), \eta), \text{ for } k = 1, ..., s,$$

$$\frac{\partial^2 s_{ni}(y,\theta)}{\partial \vartheta_j \partial \vartheta_l} = s_1(y - \Phi_{ni}(\vartheta), \eta)\left(\frac{-\partial \Phi_{ni}(\vartheta)}{\partial \vartheta_j \partial \vartheta_l}\right) + s_{1,1}(y - \Phi_{ni}(\vartheta), \eta)\left(\frac{-\partial \Phi_{ni}(\vartheta)}{\partial \vartheta_j}\right)\left(\frac{-\partial \Phi_{ni}(\vartheta)}{\partial \vartheta_l}\right),$$

for $j, l = 1, ..., c$,

$$\frac{\partial^2 s_{ni}(y,\theta)}{\partial \vartheta_j \partial \eta_k} = s_{1,k+1}(y - \Phi_{ni}(\vartheta), \eta)\left(\frac{-\partial \Phi_{ni}(\vartheta)}{\partial \vartheta_j}\right), \text{ for } j = 1, ..., c \text{ and } k = 1, ..., s,$$

$$\frac{\partial^2 s_{ni}(y,\theta)}{\partial \eta_k \partial \eta_l} = s_{k+1,l+1}(y - \Phi_{ni}(\vartheta), \eta), \text{ for } k, l = 1, ..., s.$$

Let $r(t) = s_{ni}(y, \theta + th)$ for $t \in \Re$, $h \in \Re^{c+s}$. Then

$$r'(t) = \sum_{j=1}^{c+s} \frac{\partial s_{ni}(y, \theta + th)}{\partial \theta_j} h_j, \text{ and}$$

$$r''(t) = \sum_{j=1}^{c+s} \sum_{l=1}^{c+s} \frac{\partial^2 s_{ni}(y, \theta + th)}{\partial \theta_j \partial \theta_l} h_j h_l.$$

Now $r(t) = r(0) + r'(0)\, t + \frac{1}{2} r''(\tau)\, t^2$ for some $0 \leq \tau \leq t$. Thus, setting $t = 1$ and rearranging terms we obtain

$$s_{ni}(y, \theta + h) - s_{ni}(y, \theta) - \sum_{j=1}^{c+s} \frac{\partial s_{ni}(y, \theta)}{\partial \theta_j} h_j = \frac{1}{2} \sum_{j=1}^{c+s} \sum_{l=1}^{c+s} \frac{\partial^2 s_{ni}(y, \theta + \tau h)}{\partial \theta_j \partial \theta_l} h_j h_l$$

for some $0 \leq \tau \leq 1$. Next consider the matrix $\left[\frac{\partial^2 s_{ni}(y,\phi)}{\partial \theta_j \partial \theta_l} \right]$ as a linear operator $L_{ni}(y, \phi)$, say, from $\Re^{c+s}$ into $\Re^{c+s}$ so that

$$\left(\sum_{j=1}^{c+s} \sum_{l=1}^{c+s} \frac{\partial^2 s_{ni}(y, \theta + \tau h)}{\partial \theta_j \partial \theta_l} h_j h_l \right)^2 \leq \| L_{ni}(y, \theta + \tau h) \|^2 \, \|h\|^4 .$$

Now let $T_{ni}(y, \theta) = \sup_{\|\phi - \theta\| < \delta} \|L_{ni}(y, \phi)\|$ and note for $\|h\| < \delta$ that we have $\|\theta + \tau h - \theta\| < \delta$. Thus,

$$\left(s_{ni}(y, \theta + h) - s_{ni}(y, \theta) - \sum_{j=1}^{c+s} \frac{\partial s_{ni}(y, \theta)}{\partial \theta_j} h_j \right)^2 \leq \frac{1}{4} T_{ni}^2(y, \theta) \, \|h\|^4 .$$

By hypothesis, for $\|\phi - \theta\| < \delta$,

$$\left| \frac{\partial^2 s_{ni}(y,\phi)}{\partial \theta_j \partial \theta_l} \right| \leq \Big(\left| s_1 \left(y - \Phi_{ni} \left(\tilde{\vartheta} \right), \tilde{\eta} \right) \right| + \left| s_{1,1} \left(y - \Phi_{ni} \left(\tilde{\vartheta} \right), \tilde{\eta} \right) \right| +$$

$$\sum_{k=1}^{s} \left| s_{1,k+1} \left(y - \Phi_{ni} \left(\tilde{\vartheta} \right), \tilde{\eta} \right) \right| + \sum_{k=1}^{s} \sum_{m=1}^{s} \left| s_{k+1,m+1} \left(y - \Phi_{ni} \left(\tilde{\vartheta} \right), \tilde{\eta} \right) \right| \Big) M^2.$$

Thus, by the equivalence of norms in finite dimensional vector spaces, there exists a $K > 0$ such that for $\|\phi - \theta\| < \delta$,

$$\|L_{ni}(y,\phi)\| \leq \left(\left| s_1\left(y - \Phi_{ni}\left(\tilde{\vartheta}\right), \tilde{\eta}\right)\right| + \left| s_{1,1}\left(y - \Phi_{ni}\left(\tilde{\vartheta}\right), \tilde{\eta}\right)\right| \right.$$

$$\left. + \sum_{k=1}^{s} \left| s_{1,k+1}\left(y - \Phi_{ni}\left(\tilde{\vartheta}\right), \tilde{\eta}\right)\right| + \sum_{k=1}^{s}\sum_{m=1}^{s} \left| s_{k+1,m+1}\left(y - \Phi_{ni}\left(\tilde{\vartheta}\right), \tilde{\eta}\right)\right| \right) K.$$

Hence, by hypothesis for $\|\phi - \theta\| < \delta$, $\|L_{ni}(y,\phi)\| \leq T(y) K$ so that $T_{ni}(y,\theta) \leq T(y) K$ for all n and i. The preceding finally gives

$$\left(s_{ni}(y, \theta + h) - s_{ni}(y,\theta) - \sum_{j=1}^{c+s} \frac{\partial s_{ni}(y,\theta)}{\partial \theta_j} h_j \right)^2 \leq \frac{1}{4} T^2(y) K^2 \|h\|^4$$

from which we obtain

$$\int \left(s_{ni}(y, \theta + h) - s_{ni}(y,\theta) - \sum_{j=1}^{c+s} \frac{\partial s_{ni}(y,\theta)}{\partial \theta_j} h_j \right)^2 \mu(dy) \leq \frac{1}{4} \|h\|^4 K^2 \int T^2(y)\, \mu(dy)$$

whenever $\|h\| < \delta$ for all n and i. $\diamondsuit$

The following lemma gives conditions under which a uniform q.m.d. sequence of BCLFEs can be shown to be LAN.

Lemma 2:

Assume the conditions of Lemma 1 hold. In addition, suppose that

$$p_1^2/p^2, p_1 p_{j+1}/p^2, p_{j+1}p_{k+1}/p^2, p_1^4/p^4, p_1^3 p_{j+1}/p^4,$$

$$p_1^2 p_{j+1}p_{k+1}/p^4, p_1 p_{j+1}p_{k+1}p_{l+1}/p^4 \text{ and } p_{j+1}p_{k+1}p_{l+1}p_{m+1}/p^4$$

for $j,k,l,m = 1,...,s$ are P-integrable ($dP = p$) where $p_j(z_1,...,z_{s+1}) = \frac{\partial p(z_1,...,z_{s+1})}{\partial z_j}$ for $j = 1,...,s+1$.

Let $\kappa^* = \max\left(\left|\kappa_1^{(2)}\right|, \left|\kappa_{1,j}^{(2)}\right|, \left|\kappa_{j,k}^{(2)}\right| \right)$ and $\kappa^{**} = \max\left(\left|\kappa_1^{(4)}\right|, \left|\kappa_{1,j}^{(4)}\right|, \left|\kappa_{1,j,k}^{(4)}\right|, \left|\kappa_{1,j,k,l}^{(4)}\right|, \left|\kappa_{j,k,l,m}^{(4)}\right| \right)$ where

$$\kappa_1^{(2)} = P\left(\frac{p_1^2}{p^2}\right), \kappa_{1,j}^{(2)} = P\left(\frac{p_1 p_{j+1}}{p^2}\right), \kappa_{j,k}^{(2)} = P\left(\frac{p_{j+1} p_{k+1}}{p^2}\right),$$

$$\kappa_1^{(4)} = P\left(\frac{p_1^4}{p^4}\right), \kappa_{1,j}^{(4)} = P\left(\frac{p_1^3 p_{j+1}}{p^4}\right), \kappa_{1,j,k}^{(4)} = P\left(\frac{p_1^2 p_{j+1} p_{k+1}}{p^4}\right),$$

$$\kappa_{1,j,k,l}^{(4)} = P\left(\frac{p_1 p_{j+1} p_{k+1} p_{l+1}}{p^4}\right), \kappa_{j,k,l,m}^{(4)} = P\left(\frac{p_{j+1} p_{k+1} p_{l+1} p_{m+1}}{p^4}\right)$$

for $j,k,l,m = 1,...,s$. Then $P_{ni,\theta} g_{ni,\theta}^2 \leq \|h\|^2 \kappa^* M^2$ and $P_{ni,\theta} g_{ni,\theta}^4 \leq \|h\|^4 \kappa^{**} M^4$ for all $n \geq 1$ and $1 \leq i \leq n$, where $g_{ni,\theta} = h^T \ell'_{ni,\theta}$. $\diamondsuit$

Proof:

Note that

$$\ell'_{ni,\theta}(y) = \frac{1}{p_{ni,\theta}(y - \Phi_{ni}(\vartheta), \eta)} \Big(p_1(y - \Phi_{ni}(\vartheta), \eta) \left(-\frac{\partial \Phi_{ni}(\vartheta)}{\partial \vartheta_1}\right), \ldots,$$

$$p_1(y - \Phi_{ni}(\vartheta), \eta) \left(-\frac{\partial \Phi_{ni}(\vartheta)}{\partial \vartheta_c}\right), p_2(y - \Phi_{ni}(\vartheta), \eta), \ldots, p_{s+1}(y - \Phi_{ni}(\vartheta), \eta) \Big)^T.$$

Since $\left|\frac{\partial \Phi_{ni}(\vartheta)}{\partial \vartheta_j}\right| \leq M, j = 1,...,c$, for all n and i, and by the Cauchy-Schwarz inequality and the definitions of κ^* and κ^{**} we have that

$$P_{ni,\theta} g_{ni,\theta}^2 = P_{ni,\theta} \sum_{j=1}^{c+s} \sum_{k=1}^{c+s} h_j h_k \left(\ell'_{ni,\theta} \left(\ell'_{ni,\theta}\right)^T \right)_{jk}$$

$$\leq \sum_{j=1}^{c+s} \sum_{k=1}^{c+s} |h_j| |h_k| \kappa^* M^2 \leq \|h\|^2 \kappa^* M^2$$

and

$$P_{ni,\theta} g_{ni,\theta}^4 = P_{ni,\theta} \sum_{j=1}^{c+s} \sum_{k=1}^{c+s} h_j h_k \left(\ell'_{ni,\theta} \left(\ell'_{ni,\theta}\right)^T \right)_{jk} \sum_{l=1}^{c+s} \sum_{m=1}^{c+s} h_l h_m \left(\ell'_{ni,\theta} \left(\ell'_{ni,\theta}\right)^T \right)_{lm}$$

$$\leq \sum_{j=1}^{c+s} \sum_{k=1}^{c+s} |h_j| |h_k| \sum_{l=1}^{c+s} \sum_{m=1}^{c+s} |h_l| |h_m| \kappa^{**} M^4 \leq \|h\|^4 \kappa^{**} M^4$$

for all n and i. $\diamondsuit$

That a sequence of BCLFEs satisfying the conditions of Lemmas 1 and 2 is LAN follows from Rieder (1994), Theorem 2.3.9. In particular, it suffices to establish that the following Lindeberg and differentiability conditions hold for a uniform q.m.d. sequence of BCLFEs satisfying the integrability conditions specified in Lemma 2. In addition, it is presently only assumed that $\sqrt{n}I_{n,\theta}^{-\frac{1}{2}}$ is bounded. Here, $I_{n,\theta} = \sum_{i=1}^{n} I_{ni,\theta}$ is a $(c+s)\times(c+s)$ matrix, assumed to be positive definite, and $I_{ni,\theta} = P_{ni,\theta}\ell'_{ni,\theta}\left(\ell'_{ni,\theta}\right)^T$ denotes the Fisher information matrix with $dP_{ni,\theta} = p_{ni,\theta}$.

Condition 1 (Lindeberg):

For all $t \in \Re^{c+s}$ and for all $\epsilon \in (0,\infty)$,

$$\lim_{n\to\infty}\sum_{i=1}^{n}\int_{\left\{\left|t^T I_{n,\theta}^{-1/2}\ell'_{ni,\theta}\right|>\epsilon\right\}}\left(t^T I_{n,\theta}^{-1/2}\ell'_{ni,\theta}\right)^2 dP_{ni,\theta} = 0$$

Proof:

Note for a fixed $\epsilon > 0$ and $t \in \Re^{c+s}$,

$$\sum_{i=1}^{n}\int_{\left\{\left|t^T I_{n,\theta}^{-\frac{1}{2}}\ell'_{ni,\theta}\right|>\epsilon\right\}}\left(t^T I_{n,\theta}^{-\frac{1}{2}}\ell'_{ni,\theta}\right)^2 dP_{ni,\theta}$$

$$= \sum_{i=1}^{n}\int_{\left\{\left|t^T\sqrt{n}I_{n,\theta}^{-\frac{1}{2}}\ell'_{ni,\theta}\right|>\sqrt{n}\epsilon\right\}}\left(t^T\tfrac{1}{\sqrt{n}}\sqrt{n}I_{n,\theta}^{-\frac{1}{2}}\ell'_{ni,\theta}\right)^2 dP_{ni,\theta}.$$

Let $h = \sqrt{n}I_{n,\theta}^{-\frac{1}{2}}t$ and note that h is bounded since $\sqrt{n}I_{n,\theta}^{-\frac{1}{2}}$ is bounded and t is fixed. Then the above expression becomes

$$\frac{1}{n}\sum_{i=1}^{n}\int_{\left\{\left|h^T\ell'_{ni,\theta}\right|>\sqrt{n}\epsilon\right\}}\left(h^T\ell'_{ni,\theta}\right)^2 dP_{ni,\theta}.$$

By Lemma 2 $P_{ni,\theta}g_{ni,\theta}^4$, with $g_{ni,\theta} = h^T\ell'_{ni,\theta}$, are uniformly bounded in n and i. Thus, the collection of random variables $g_{ni,\theta}^2$ is uniformly integrable (cf Billingsley (1995), p.338). Therefore, for sufficiently large n,

$$\int_{\{g^2_{ni,\theta} > n\epsilon^2\}} g^2_{ni,\theta} dP_{ni,\theta}.$$

is arbitrarily small for all such n and $1 \leq i \leq n$. Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_{\{g^2_{ni,\theta} > n\epsilon^2\}} g^2_{ni,\theta} dP_{ni,\theta} = 0$$

for all $\epsilon \in (0, \infty)$ Thus, the Lindeberg condition holds. $\diamondsuit$

Condition 2 (differentiability):

For all $b \in (0, \infty)$,

$$\lim_{n\to\infty} \sup_{\|t\| \leq b} \sum_{i=1}^{n} \int \left(\sqrt{p_{ni,\theta_n(t)}} - \sqrt{p_{ni,\theta}} \left(1 + \frac{1}{2} t^T I^{-1/2}_{n,\theta} \ell'_{ni,\theta} \right) \right)^2 dP_{ni,\theta} = 0$$

Proof:

The local parameter alternatives $\theta_n(t)$ about θ are given by $\theta_n(t) = \theta + I^{-\frac{1}{2}}_{n,\theta} t$. Let $\tilde{h} = \frac{1}{\sqrt{n}}(\sqrt{n} I^{-\frac{1}{2}}_{n,\theta} t)$ so that $\theta_n(t) = \theta + \tilde{h}$. Note $\tilde{h} \to 0$ uniformly with $\|t\|$ bounded. Note that

$$\sup_{\|t\| \leq b} \sum_{i=1}^{n} \int \left(\sqrt{p_{ni,\theta_n(t)}} - \sqrt{p_{ni,\theta}} \left(1 + \frac{1}{2} t^T I^{-1/2}_{n,\theta} \ell'_{ni,\theta} \right) \right)^2 dP_{ni,\theta}$$

$$= \sup_{\|t\| \leq b} \sum_{i=1}^{n} \int \left(\sqrt{p_{ni,\theta+\tilde{h}}} - \sqrt{p_{ni,\theta}} - \frac{1}{2} \tilde{h}^T \ell'_{ni,\theta} \sqrt{p_{ni,\theta}} \right)^2 dP_{ni,\theta} = \sup_{\|t\| \leq b} \sum_{i=1}^{n} o\left(\left\| \tilde{h} \right\|^2 \right).$$

Note that the last equality comes from the uniform q.m.d. of the sequence of BCLFEs at $\theta \in \Theta$ as provided by Lemma 1. Also, note that $\left\| \tilde{h} \right\|^2 = O\left(\frac{1}{n}\right)$ since $\sqrt{n} I^{-\frac{1}{2}}_{n,\theta}$ is bounded and t is fixed, and thus the differentiability condition holds. $\diamondsuit$

Conditions 1 and 2 imply the L_2 differentiability of the parametric array $\{P_{ni,\theta} : \theta \in \Theta\}$ at a fixed θ (cf Rieder(1994) Definition 2.3.8), and thus the desired log likelihood expansion

indicated below.

In particular, we now suppose that $\frac{1}{n}I_{n,\theta} \rightarrow I_\theta$ where I_θ is positive definite, and therefore $\sqrt{n}I_{n,\theta}^{-\frac{1}{2}} \rightarrow I_\theta^{-\frac{1}{2}}$. Let $h_n = \sqrt{n}I_{n,\theta}^{-\frac{1}{2}}t_n$ where $t_n \rightarrow t \in \Re^{c+s}$. Thus, $h_n \rightarrow h \in \Re^{c+s}$ where $h = I_\theta^{-\frac{1}{2}}t$. Consequently, $t = I_\theta^{\frac{1}{2}}h$ so that $\|t\|^2 = h^T I_\theta h$ and $t^T I_{n,\theta}^{-\frac{1}{2}} = h^T I_\theta^{\frac{1}{2}}(\sqrt{n}I_{n,\theta}^{-\frac{1}{2}})(\frac{1}{\sqrt{n}})$. The local parameter alternatives $\theta_n(t_n)$ about θ are given by $\theta_n(t_n) = \theta + I_{n,\theta}^{-\frac{1}{2}}t_n$, and with the preceding identification for h_n, we have $\theta_n(t_n) = \theta + \frac{h_n}{\sqrt{n}}$. Accordingly, by Theorem 2.3.9 of Rieder (1994), a L_2 differentiable sequence of BCLFEs is LAN. That is,

$$\log \frac{dP_{n,\theta+h_n/\sqrt{n}}}{dP_{n,\theta}} = t^T I_{n,\theta}^{-\frac{1}{2}} \sum_i \ell'_{ni,\theta}(x_{ni}) - (1/2)\|t\|^2 + o_{P_{n,\theta}}(1)$$

$$= h^T I_\theta^{\frac{1}{2}} \sqrt{n} I_{n,\theta}^{-\frac{1}{2}} \Delta_{n,\theta} - (1/2) h^T I_\theta h + o_{P_{n,\theta}}(1)$$

where $\Delta_{n,\theta} = \frac{1}{\sqrt{n}} \sum_i \ell'_{ni,\theta}(x_{ni})$ are random vectors such that $\Delta_{n,\theta} \Rightarrow N_{c+s}(0, I_\theta)$ as $n \rightarrow \infty$ under $P_{n,\theta}$.

In the following let $I_n(\beta, \gamma, \eta)$ denote the Fisher information metric with respect to all of the parameters, including the nuisance parameter vector η. It is assumed for Theorem 1 below that the nuisance parameters satisfy the orthogonality condition introduced by Cox and Reid(1987). That is, the nuisance parameter directions are perpendicular to the other parameter directions with respect to the information metric. For example, this condition is satisfied when the errors are Normally distributed and the nuisance parameters are the elements of the covariance matrix, as discussed by Skovgaard(1984). Returning to the general case, it follows that at a point $\theta = (\beta, 0, \eta)$, the γ-direction is orthogonal to the (β, η)-direction with respect to the information metric I_n. Theorem 1 can now be stated as follows.

Theorem 1:

Suppose the conditions that were indicated under Lemma 1 and Lemma 2 hold at $\theta =$

$(\beta, 0, \eta)$ for a sequence of BCLFEs

$$E_n\left(x_{n1}, ..., x_{nn}; Z_n\left(\beta\right) \in \kappa_n, \beta \in B\right)$$

and that the $(1/n) I_n\left(\beta, 0, \eta\right)$ converge to $I\left(\beta, 0, \eta\right)$, say. In addition, suppose the matrices $\ell''_{ni,\theta}$ of second-order partial derivatives are locally bounded. If the unrestricted and restricted (by $H_o : \gamma = 0$) maximum likelihood estimators (MLEs) $\hat{\theta}_n$ and $\hat{\theta}_{n,o}$, respectively, are $\sqrt{n}$-consistent under $(\beta, 0, \eta)$, then the sequence of log LRT statistics Λ_n for testing $H_o : \gamma = 0$ versus $H_a : \gamma \neq 0$ converges under $(\beta, \gamma/\sqrt{n}, \eta)$ in distribution to a random variable distributed according to the $\chi^2\left(c - p, \delta\right)$ distribution with noncentrality parameter $\delta = (0, \gamma, 0)^T I\left(\beta, 0, \eta\right)(0, \gamma, 0).\diamondsuit$

Before giving the proof of Theorem 1, we establish conditions for $\sqrt{n}$-consistency of the MLE of θ. Recall that $p_{ni,\theta}\left(y\right) = p\left(y - \Phi_{ni}\left(\vartheta\right), \eta\right)$ is the density of y_{ni} where $\Phi_{ni}\left(\vartheta\right) = \left(f_n\left(\beta\right)\right)_i - \left(B_{Z_n(\beta)}\gamma\right)_i$. Also, $\theta = \left(\vartheta, \eta\right)^T$ where $\vartheta = (\beta, \gamma)$ has dimension c, and η is an s dimensional nuisance parameter vector. Next we introduce some notation and assumptions that will be used to establish $\sqrt{n}-$ consistency of the MLE of θ under the proposed model and the constructed full model.

Notation

- $\ell_{n,\theta} = \sum_{i=1}^{n} \ell_{ni,\theta}$ where $\ell_{ni,\theta} = \log p\left(y - \Phi_{ni}\left(\vartheta\right), \eta\right)$.
- $\ell'_{n,\theta} = \sum_{i=1}^{n} \ell'_{ni,\theta}$ where $\ell'_{ni,\theta} = \frac{\partial}{\partial\theta} \log p\left(y - \Phi_{ni}\left(\vartheta\right), \eta\right)$. Specifically, the components of the $(c + s)$ dimensional vector $\ell'_{ni,\theta}$ have the following forms where $p_j(z_1, \ldots, z_{s+1}) = \frac{\partial}{\partial z_j} p(z_1, \ldots, z_{s+1})$ for $j = 1, \ldots, s + 1$.

  1. $\frac{\partial}{\partial\vartheta_j}\ell_{ni,\theta} = \frac{-1}{p(y-\Phi_{ni}(\vartheta),\eta)} p_1\left(y - \Phi_{ni}\left(\vartheta\right), \eta\right) \frac{\partial}{\partial\vartheta_j}\Phi_{ni}\left(\vartheta\right)$ for $j = 1, \ldots, c$.

  2. $\frac{\partial}{\partial\eta_j}\ell_{ni,\theta} = \frac{1}{p(y-\Phi_{ni}(\vartheta),\eta)} p_{j+1}\left(y - \Phi_{ni}\left(\vartheta\right), \eta\right)$ for $j = 1, \ldots, s$.

- $\ell''_{n,\theta} = \sum_{i=1}^{n} \ell''_{ni,\theta}$ where $\ell''_{ni,\theta} = \frac{\partial}{\partial\theta}\ell'_{ni,\theta}$. Specifically, the components of the $(c+s)\times(c+s)$ matrix $\ell''_{ni,\theta}$ have the following forms where $p_{j,k}(z_1,\ldots,z_{s+1}) = \frac{\partial^2}{\partial z_j \partial z_k} p(z_1,\ldots,z_{s+1})$ for $j,k = 1,2,...,(s+1)$.

$$1.\frac{\partial^2}{\partial\vartheta_j\partial\vartheta_k}\ell_{ni,\theta} = \frac{1}{[p(y-\Phi_{ni}(\vartheta),\eta)]^2} p_1^2\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)\left(-\frac{\partial}{\partial\vartheta_j}\Phi_{ni}\left(\vartheta\right)\frac{\partial}{\partial\vartheta_k}\Phi_{ni}\left(\vartheta\right)\right)$$

$$-\frac{1}{p(y-\Phi_{ni}(\vartheta),\eta)} p_{1,1}\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)\left(-\frac{\partial}{\partial\vartheta_j}\Phi_{ni}\left(\vartheta\right)\frac{\partial}{\partial\vartheta_k}\Phi_{ni}\left(\vartheta\right)\right)$$

$$-\frac{1}{p(y-\Phi_{ni}(\vartheta),\eta)} p_1\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)\frac{\partial^2}{\partial\vartheta_j\partial\vartheta_k}\Phi_{ni}\left(\vartheta\right)$$

$$= a\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)\left(\frac{\partial}{\partial\vartheta_j}\Phi_{ni}\left(\vartheta\right)\frac{\partial}{\partial\vartheta_k}\Phi_{ni}\left(\vartheta\right)\right) - b\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)\frac{\partial^2}{\partial\vartheta_j\partial\vartheta_k}\Phi_{ni}\left(\vartheta\right)$$

where $a\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right) = \frac{p_{1,1}(y-\Phi_{ni}(\vartheta),\eta)}{p(y-\Phi_{ni}(\vartheta),\eta)} - \frac{p_1^2(y-\Phi_{ni}(\vartheta),\eta)}{[p(y-\Phi_{ni}(\vartheta),\eta)]^2}$

and $b\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right) = \frac{p_1(y-\Phi_{ni}(\vartheta),\eta)}{p(y-\Phi_{ni}(\vartheta),\eta)}$.

$$2.\frac{\partial^2}{\partial\vartheta_j\partial\eta_k}\ell_{ni,\theta} = \frac{1}{[p(y-\Phi_{ni}(\vartheta),\eta)]^2} p_1\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)p_{k+1}\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)\frac{\partial}{\partial\vartheta_j}\Phi_{ni}\left(\vartheta\right)$$

$$-\frac{1}{p(y-\Phi_{ni}(\vartheta),\eta)} p_{1,k+1}\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)\frac{\partial}{\partial\vartheta_j}\Phi_{ni}\left(\vartheta\right).$$

$$3.\frac{\partial^2}{\partial\eta_j\partial\eta_k}\ell_{ni,\theta} = -\frac{1}{[p(y-\Phi_{ni}(\vartheta),\eta)]^2} p_{j+1}\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)p_{k+1}\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right)$$

$$+\frac{1}{p(y-\Phi_{ni}(\vartheta),\eta)} p_{j+1,k+1}\left(y-\Phi_{ni}\left(\vartheta\right),\eta\right).$$

- $I_{n,\theta} = \sum_{i=1}^{n} I_{ni,\theta}$ where $I_{ni,\theta} = E_\theta\left[-\ell''_{ni,\theta}\right] \approx E_\theta\left[\left(-\ell'_{ni,\theta}\right)\left(-\ell'_{ni,\theta}\right)^T\right]$ is the Fisher Information matrix.
- $\bar{I}_{n,\theta} = \frac{1}{n}\sum_{i=1}^{n} I_{ni,\theta}$.

Assumptions(along with the conditions for Lemmas 1 and 2)

1. $\bar{I}_{n,\theta} \to I_\theta$, say, as $n\to\infty$ where I_θ is positive definite.

2. $\hat{\theta}_n \xrightarrow{P_{n,\theta}} \theta$ where $\hat{\theta}_n$ is the MLE of θ and thus $\ell'_{n,\hat{\theta}_n} = 0$.

3. $\frac{1}{n}\left(\ell'''_{n,\tilde{\theta}_n}\right)_{ghk} = O_{P_{n,\theta}}(1)$ for $g,h,k = 1,2,...,(c+s)$ for all $\tilde{\theta}_n$ in a neighborhood of θ.

Proposition 1:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) = I_\theta^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell'_{ni,\theta} + o_{P_{n,\theta}}(1)$$

under $P_{n,\theta}$.$\diamondsuit$

Proof:

For fixed $g = 1,\ldots,(c+s)$, consider the second order Taylor expansion of $\left(\ell'_{n,\hat{\theta}_n}\right)_g$ about the true value $(\theta)_g$ so that

$$0 = \left(\ell'_{n,\hat{\theta}_n}\right)_g = \left(\ell'_{n,\theta}\right)_g + \sum_{h=1}^{c+s}\left(\hat{\theta}_n - \theta\right)_h\left(\ell''_{n,\theta}\right)_{gh} + \frac{1}{2}\sum_{h=1}^{c+s}\sum_{k=1}^{c+s}\left(\hat{\theta}_n - \theta\right)_h\left(\hat{\theta}_n - \theta\right)_k\left(\ell'''_{n,\theta_n^*}\right)_{ghk}$$

where $\|\theta_n^* - \theta\| < \|\hat{\theta}_n - \theta\|$. The preceding can be written in matrix form as

$$\frac{1}{\sqrt{n}}\ell'_{n,\theta} = \left[-\frac{1}{n}\ell''_{n,\theta} - \frac{1}{2n}\left(\hat{\theta}_n - \theta\right)^T\ell'''_{n,\theta_n^*}\right]\sqrt{n}\left(\hat{\theta}_n - \theta\right)$$

where ℓ'''_{n,θ_n^*} is a $(c+s)$ dimensional vector of $(c+s)\times(c+s)$ matrices depending on the third-order derivatives $\left(\ell'''_{n,\theta}\right)_{ghk}$. Given the assumptions above, it suffices to show that

$$-\frac{1}{n}\ell''_{n,\theta} - \bar{I}_{n,\theta} \xrightarrow{P_{n,\theta}} 0.$$

For then, it follows that

$$\frac{1}{\sqrt{n}}\ell'_{n,\theta} = \left[I_\theta + o_{P_{n,\theta}}(1) + \frac{1}{2}\left(\hat{\theta}_n - \theta\right)^T O_{P_{n,\theta}}(1)\right]\sqrt{n}\left(\hat{\theta}_n - \theta\right) = \left[I_\theta + o_{P_{n,\theta}}(1)\right]\sqrt{n}\left(\hat{\theta}_n - \theta\right).$$

Since the probability that the matrix $I_\theta + o_{P_{n,\theta}}(1)$ is invertible tends to one, applying $\left[I_\theta + o_{P_{n,\theta}}(1)\right]^{-1}$ left and right, and using the fact that $\frac{1}{\sqrt{n}}\ell'_{n,\theta}$ converges in law, establishes the proposition.

The convergence $-\frac{1}{n}\ell''_{n,\theta} - \bar{I}_{n,\theta} \xrightarrow{P_{n,\theta}} 0$ will be established using a weak law of large numbers (WLLN) for triangular arrays (cf Billingsley (1995), p. 86).

For fixed $g, h = 1, 2, ..., (c+s)$, consider the triangular array of random variables

$$\left(-\ell''_{n1,\theta}\right)_{gh}, \left(-\ell''_{n2,\theta}\right)_{gh}, ..., \left(-\ell''_{nn,\theta}\right)_{gh}$$

where $n = 1, 2, \ldots$ and let $S_n = \left(-\ell''_{n,\theta}\right)_{gh} = \sum_{i=1}^{n} \left(-\ell''_{ni,\theta}\right)_{gh}$. Thus, if $\frac{Var_\theta(S_n)}{n^2} \to 0$ holds as $n \to \infty$, then by Chebyshev's inequality it follows that $\frac{S_n - E_\theta[S_n]}{n} \xrightarrow{P_{n,\theta}} 0$ where $E_\theta[S_n] = (I_{n,\theta})_{gh} = \sum_{i=1}^{n} (I_{ni,\theta})_{gh}$.

In particular, for any $\epsilon > 0$,

$$P_{n,\theta}\left(|S_n - E_\theta[S_n]| \geq n\epsilon\right) \leq \frac{1}{n^2\epsilon^2} Var_\theta[S_n] \leq \frac{1}{n^2\epsilon^2} \sum_{i=1}^{n} E_\theta\left[\left(-\ell''_{ni,\theta}\right)_{gh}\right]^2.$$

Next it is shown that $E_\theta\left[\left(\ell''_{ni,\theta}\right)_{gh}\right]^2$ is bounded. Let the error density p be defined with respect to a (translation invariant) measure μ (e.g. Lebesgue measure) and first suppose $\ell''_{ni,\theta}$ has the form $\frac{\partial^2}{\partial\vartheta_j \partial\vartheta_k}\ell_{ni,\theta}$. Then $E_\theta\left[\left(\ell''_{ni,\theta}\right)_{gh}\right]^2 =$

$$\int \left[a\left(y - \Phi_{ni}(\vartheta), \eta\right)\left(\frac{\partial}{\partial\vartheta_j}\Phi_{ni}(\vartheta)\frac{\partial}{\partial\vartheta_k}\Phi_{ni}(\vartheta)\right) - b\left(y - \Phi_{ni}(\vartheta), \eta\right)\frac{\partial^2}{\partial\vartheta_j\partial\vartheta_k}\Phi_{ni}(\vartheta)\right]^2 \times p\left(y - \Phi_{ni}(\vartheta), \eta\right) d\mu(y)$$

$$= \int \left[a(y,\eta)\left(\frac{\partial}{\partial\vartheta_j}\Phi_{ni}(\vartheta)\frac{\partial}{\partial\vartheta_k}\Phi_{ni}(\vartheta)\right) - b(y,\eta)\frac{\partial^2}{\partial\vartheta_j\partial\vartheta_k}\Phi_{ni}(\vartheta)\right]^2 p(y,\eta)\, d\mu(y)$$

$$= \left[\int (a(y,\eta))^2 p(y,\eta)\, d\mu(y)\right]\left(\frac{\partial}{\partial\vartheta_j}\Phi_{ni}(\vartheta)\right)^2\left(\frac{\partial}{\partial\vartheta_k}\Phi_{ni}(\vartheta)\right)^2 -$$

$$2\left[\int a(y,\eta)\, b(y,\eta)\, p(y,\eta)\, d\mu(y)\right]\left(\frac{\partial}{\partial\vartheta_j}\Phi_{ni}(\vartheta)\right)\left(\frac{\partial}{\partial\vartheta_k}\Phi_{ni}(\vartheta)\right)\left(\frac{\partial^2}{\partial\vartheta_j\partial\vartheta_k}\Phi_{ni}(\vartheta)\right) +$$

$$\left[\int (b(y,\eta))^2 p(y,\eta)\, d\mu(y)\right]\left(\frac{\partial^2}{\partial\vartheta_j\partial\vartheta_k}\Phi_{ni}(\vartheta)\right)^2.$$

Therefore, $0 \leq E_\theta \left[\left(\ell''_{ni,\theta} \right)_{gh} \right]^2 \leq$

$$\left[\int \left(a\left(y,\eta \right) \right)^2 p\left(y,\eta \right) d\mu(y) \right] \left(\frac{\partial}{\partial \vartheta_j} \Phi_{ni}\left(\vartheta \right) \right)^2 \left(\frac{\partial}{\partial \vartheta_k} \Phi_{ni}\left(\vartheta \right) \right)^2 +$$

$$2 \left| \int a\left(y,\eta \right) b\left(y,\eta \right) p\left(y,\eta \right) d\mu(y) \right| \left| \frac{\partial}{\partial \vartheta_j} \Phi_{ni}\left(\vartheta \right) \right| \left| \frac{\partial}{\partial \vartheta_k} \Phi_{ni}\left(\vartheta \right) \right| \left| \frac{\partial^2}{\partial \vartheta_j \partial \vartheta_k} \Phi_{ni}\left(\vartheta \right) \right| +$$

$$\left[\int \left(b\left(y,\eta \right) \right)^2 p\left(y,\eta \right) d\mu(y) \right] \left(\frac{\partial^2}{\partial \vartheta_j \partial \vartheta_k} \Phi_{ni}\left(\vartheta \right) \right)^2 .$$

Now θ is fixed, and the x_{ni} take values in a closed and bounded region on which the partial derivatives of $\Phi_{ni}\left(\vartheta \right)$ are bounded. Thus, all the partial derivative terms are $\leq M$, say, so that

$$0 \leq E_\theta \left[\left(\ell''_{ni,\theta} \right)_{gh} \right]^2 \leq \Big[\int \left(a\left(y,\eta \right) \right)^2 p\left(y,\eta \right) d\mu(y)$$

$$+2 \left| \int a\left(y,\eta \right) b\left(y,\eta \right) p\left(y,\eta \right) d\mu(y) \right| + \int \left(b\left(y,\eta \right) \right)^2 p\left(y,\eta \right) d\mu(y) \Big] M$$

for all n and i. (Note that this bound can be obtained for all n and i and uniformly in θ for θ in a closed and bounded region.) The remaining forms for $\ell''_{ni,\theta}$, $\frac{\partial^2}{\partial \vartheta_j \partial \eta_k} \ell_{ni,\theta}$ and $\frac{\partial^2}{\partial \eta_j \partial \eta_k} \ell_{ni,\theta}$, are similarly bounded.

Thus, $E_\theta \left[\left(-\ell''_{ni,\theta} \right)_{gh} \right]^2$ is bounded for all $g, h = 1, 2, ..., (c+s)$. This implies $\frac{Var_\theta(S_n)}{n^2} \to 0$ as $n \to \infty$. Therefore,

$$P_{n,\theta} \left(\left| S_n - E_\theta \left[S_n \right] \right| \geq n\epsilon \right) \to 0$$

as $n \to \infty$. That is,

$$\frac{1}{n} \left(-\ell''_{n,\theta} \right)_{gh} - \left(\bar{I}_{n,\theta} \right)_{gh} \overset{P_{n,\theta}}{\to} 0$$

for all $g, h = 1, 2, ..., (c+s)$.$\diamondsuit$

In the proof of Theorem 1, the log LRT statistic for testing $H_0 : \gamma = 0$ versus $H_a : \gamma \neq 0$ is given by

$$\Lambda_n = 2 \log \frac{\sup_{\theta \in \Theta} dP_{n,\theta}}{\sup_{\theta \in \Theta_0} dP_{n,\theta}} (y_n)$$

where $\Theta = B \times R^{c-p} \times R^s$ and $\Theta_0 = B \times \{0\} \times R^s$. The LAN property will be used to derive the asymptotic distribution of Λ_n. First, define the local parameter spaces $H_n = \sqrt{n}(\Theta - \theta)$ and $H_{n,0} = \sqrt{n}(\Theta_0 - \theta)$ where $\theta = (\beta, 0, \eta)$, and rewrite Λ_n as

$$\Lambda_n = 2 \sup_{h \in H_n} \log \frac{dP_{n,\theta+h/\sqrt{n}}}{dP_{n,\theta}} (y_n) - 2 \sup_{h \in H_{n,0}} \log \frac{dP_{n,\theta+h/\sqrt{n}}}{dP_{n,\theta}} (y_n). \tag{5.1}$$

In addition, the following definition of convergence of sets as indicated in van der Vaart (1998) will be useful. Write $H_n \to H$ if H is the set of all limits of converging sequences (h_n) where $h_n \in H_n$ for every n, and, moreover, the limit of every converging sequence (h_{n_i}) where $h_{n_i} \in H_{n_i}$ for every i is contained in H. Similarly, $H_{n,0} \to H_0$ for some limit set H_0. In particular, with $\theta = (\beta, 0, \eta)$, the local parameter spaces are such that $H_n \to R^p \times R^{c-p} \times R^s$ and $H_{n,0} \to R^p \times \{0\} \times R^s$. The method of proof for Theorem 1 follows van der Vaart (Theorem 16.7, 1998), and represents generalization to the case of triangular arrays.

Proof of Theorem 1:

Let $h = (0, \gamma, 0)$, and consider specific local parameter alternatives $\theta + h/\sqrt{n}$. By the LAN property of BCLFEs, $\Delta_{n,\theta} \Rightarrow \Delta \sim N_{c+s}(0, I_\theta)$ under $P_{n,\theta}$, and

$$\log \frac{dP_{n,\theta+h/\sqrt{n}}}{dP_{n,\theta}} = h^T \Delta_{n,\theta} - \frac{1}{2} h^T I_\theta h + o_{p_{n,\theta}}(1),$$

so that $\left(\Delta_{n,\theta}, \log \frac{dP_{n,\theta+h/\sqrt{n}}}{dP_{n,\theta}}\right)^T \Rightarrow \left(\Delta, h^T\Delta - \frac{1}{2} h^T I_\theta h\right)^T$

$$\sim N\left(\begin{pmatrix} 0 \\ -\frac{1}{2} h^T I_\theta h \end{pmatrix}, \begin{pmatrix} I_\theta & I_\theta h \\ h^T I_\theta & h^T I_\theta h \end{pmatrix} \right).$$

Thus, the sequences $P_{n,\theta+h/\sqrt{n}}$ and $P_{n,\theta}$ are mutually contiguous (cf example 6.5 of van der Vaart(1998)).

As indicated in Lemma 3 below,

$$\sup_{\|h\|\leq M} |Z_n(h)| \overset{P_{n,\theta}}{\to} 0$$

for every M where $\theta = (\beta, 0, \eta)$ is fixed and

$$Z_n(h) = \log \frac{dP_{n,\theta+h/\sqrt{n}}}{dP_{n,\theta}} - h^T \Delta_{n,\theta} + \frac{1}{2} h^T I_\theta h \tag{5.2}$$

with $\Delta_{n,\theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell'_{ni,\theta}$. The above convergence holds for every sequence (M_n) where $M_n \to \infty$ sufficiently slowly (see Remark 1 below). Fix such a sequence so that, using $\sqrt{n}-$consistency, $\hat{\theta}_n$ and $\hat{\theta}_{n,0}$ are contained in the ball with radius $M_n/\sqrt{n}$ around $\theta = (\beta, 0, \eta)$ with $P_{n,\theta}$ probability converging to one. Thus, with contiguity, Λ'_n and Λ_n have the same limit distribution where Λ'_n is the statistic obtained by replacing H_n and $H_{n,0}$ in the definition of Λ_n with $H'_n = H_n \cap \texttt{ball}(0, M_n)$ and $H'_{n,0} = H_{n,0} \cap \texttt{ball}(0, M_n)$, respectively. Also note that $H'_n \to R^p \times R^{c-p} \times R^s$ and $H'_{n,0} \to R^p \times \{0\} \times R^s$. Now, by the uniform convergence of $Z_n(h)$ to zero on H'_n and $H'_{n,0}$,

$$\Lambda'_n = 2 \sup_{h \in H'_n} \left(h^T \Delta_{n,\theta} - \frac{1}{2} h^T I_\theta h \right) - 2 \sup_{h \in H'_{n,0}} \left(h^T \Delta_{n,\theta} - \frac{1}{2} h^T I_\theta h \right) + o_{P_{n,\theta}}(1)$$

$$= \inf_{h \in H'_{n,0}} \left(I_\theta^{-1} \Delta_{n,\theta} - h \right)^T I_\theta \left(I_\theta^{-1} \Delta_{n,\theta} - h \right) - \inf_{h \in H'_n} \left(I_\theta^{-1} \Delta_{n,\theta} - h \right)^T I_\theta \left(I_\theta^{-1} \Delta_{n,\theta} - h \right) + o_{P_{n,\theta}}(1)$$

$$\begin{aligned} &= \left\| I_\theta^{\frac{-1}{2}} \Delta_{n,\theta} - I_\theta^{\frac{1}{2}} H'_{n,0} \right\|^2 - \left\| I_\theta^{\frac{-1}{2}} \Delta_{n,\theta} - I_\theta^{\frac{1}{2}} H'_n \right\|^2 + o_{P_{n,\theta}}(1) \\ &= \left\| I_\theta^{\frac{-1}{2}} \Delta_{n,\theta} - I_\theta^{\frac{1}{2}} H_0 \right\|^2 - \left\| I_\theta^{\frac{-1}{2}} \Delta_{n,\theta} - I_\theta^{\frac{1}{2}} H \right\|^2 + o_{P_{n,\theta}}(1). \end{aligned}$$

Note that the last equality follows by Lemma 7.13 of van der Vaart(1998) with $H_0 = R^p \times \{0\} \times R^s$ and $H = R^p \times R^{c-p} \times R^s$ (this lemma is stated as Lemma 4 below for

completeness).

Thus, by contiguity and Le Cam's third lemma (cf example 6.7 of van der Vaart(1998)),

$$\Delta_{n,\theta} \Rightarrow N_{c+s}\left(I_\theta h, I_\theta\right)$$

under $P_{n,\theta+h/\sqrt{n}}$. Hence, by the continuous mapping theorem, $\Lambda_n^{'} \overset{\theta+h/\sqrt{n}}{\rightarrow} \Lambda$ where

$$\Lambda = \left\| I_\theta^{\frac{1}{2}} X - I_\theta^{\frac{1}{2}} H_0 \right\|^2 - \left\| I_\theta^{\frac{1}{2}} X - I_\theta^{\frac{1}{2}} H \right\|^2$$

with X distributed as $N_{c+s}\left(h, I_\theta^{-1}\right)$. Since H is the full space R^{c+s}, Λ reduces to

$$\Lambda = \left\| Z + I_\theta^{\frac{1}{2}} h - I_\theta^{\frac{1}{2}} H_0 \right\|^2$$

where Z is distributed as a standard normal vector. Thus, Λ is distributed according to the noncentral chi-square distribution with $c-p$ degrees of freedom and noncentrality parameter δ, say, given by

$$\left\| I_\theta^{\frac{1}{2}} h - I_\theta^{\frac{1}{2}} H_0 \right\|^2 = (0,\gamma,0)^T I\left(\beta,0,\eta\right)(0,\gamma,0).$$

Note that the orthogonality of the nuisance parameters is used to establish the preceding equality. In particular,

$$\left\| I_\theta^{\frac{1}{2}} h - I_\theta^{\frac{1}{2}} H_0 \right\|^2 = \inf_{h_0 \in H_0} \left\| I_\theta^{\frac{1}{2}} h - I_\theta^{\frac{1}{2}} h_0 \right\|^2 = \inf_{h_0 \in H_0} \|h - h_0\|_I^2 = \left\| h - P_{H_0}^I h \right\|_I^2$$

where the subscript I refers to the norm calculated using the information metric. Taking $h = (0,\gamma,0)$ we obtain $P_{H_0}^I h = 0$ using the orthogonality of the γ direction to the (β,η) direction with respect to the information metric. Thus,

$$\left\| I_\theta^{\frac{1}{2}} h - I_\theta^{\frac{1}{2}} H_0 \right\|^2 = \|h\|_I^2 = (0,\gamma,0)^T I\left(\beta,0,\eta\right)(0,\gamma,0).$$

Finally, since Λ_n and $\Lambda_n^{'}$ have the same limit distribution, conclude that

$$\Lambda_n \overset{\left(\beta,\gamma/\sqrt{n},\eta\right)}{\rightarrow} \chi_{c-p}^2\left(\delta\right). \diamondsuit$$

Lemma 3:

Suppose the conditions of Theorem 1 hold at $\theta = (\beta, 0, \eta)$, and let $E_\theta = \{\theta + h/\sqrt{n} : h \in T\}$ where $T = \{h \in R^{c+s} : \|h\| \leq 1\}$. If $\left\|\ell''_{ni,\phi}\right\| \leq M_{ni}$ for every $\phi \in E_\theta$ where $\frac{1}{n}\sum_{i=1}^{n} M_{ni} = O_{P_{n,\theta}}(1)$ then

$$\sup_{\|h\|\leq 1} |Z_n(h)| \overset{P_{n,\theta}}{\to} 0.\diamondsuit$$

Proof:

Consider $Z_n(h)$ to be a sequence of stochastic processes indexed by h. As shown first, the sequence Z_n converges weakly in the space $l^\infty(T)$ of bounded functions on T with the uniform norm.

Note by LAN the sequence $Z_n(h)$ converges marginally in distribution to zero, and for s and t in T, $Z_n(s) - Z_n(t) =$

$$\log dP_{n,\theta+s/\sqrt{n}} - \log dP_{n,\theta+t/\sqrt{n}} + (t-s)^T \Delta_{n,\theta} + \tfrac{1}{2}\left((s-t)^T I_\theta (s-t) + 2(s-t)^T I_\theta t\right).$$

Since $[\theta + s/\sqrt{n}, \theta + t/\sqrt{n}] \subset E_\theta$, by the mean value theorem there is a point $\theta + r_i/\sqrt{n} \in [\theta + s/\sqrt{n}, \theta + r/\sqrt{n}]$ with $r_i \in T$ such that

$$\log P_{ni,\theta+s/\sqrt{n}} - \log P_{ni,\theta+t/\sqrt{n}} = \left(\ell'_{ni,\theta+r_i/\sqrt{n}}{}^T\right)\left(\frac{s-t}{\sqrt{n}}\right)$$

for each $i = 1, ..., n$. Thus,

$$\log dP_{n,\theta+s/\sqrt{n}} - \log dP_{n,\theta+t/\sqrt{n}} = \left(\frac{s-t}{\sqrt{n}}\right)^T \sum_{i=1}^{n} \ell'_{ni,\theta+r_i/\sqrt{n}}$$

$$= (s-t)^T \left(\Delta_{n,\theta} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\ell'_{ni,\theta+r_i/\sqrt{n}} - \ell'_{ni,\theta}\right)\right).$$

Also,

$$\left|(s-t)^T \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\ell'_{ni,\theta+r_i/\sqrt{n}} - \ell'_{ni,\theta}\right)\right|$$

$$\leq \|s-t\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\|\ell'_{ni,\theta+r_i/\sqrt{n}} - \ell'_{ni,\theta}\right\| \leq \|s-t\| \frac{1}{n}\sum_{i=1}^{n} M_{ni}$$

where $\frac{1}{n}\sum_{i=1}^{n} M_{ni} = O_{P_{n,\theta}}(1)$ by hypothesis. Since $\Delta_{n,\theta} = O_{P_{n,\theta}}(1)$ as well, it follows that for every $\epsilon > 0$ and $\delta > 0$ there exists a partition of T into finitely many sets $T_1, T_2, ..., T_k$ such that

$$\limsup_{n\to\infty} P_{n,\theta}\left(\sup_{1\leq j\leq k}\sup_{s,t\in T_j}|Z_n(s) - Z_n(t)| \geq \epsilon\right) \leq \delta.$$

Hence, by Theorem 18.14 of van der Vaart(1998) (this theorem is stated as Lemma 5 below for completeness), the sequence Z_n converges weakly in the space $l^\infty(T)$. Finally, since taking a supremum is a continuous operation and the marginal convergence is to zero, the weak limit is then necessarily zero. $\diamondsuit$

Lemma 4:

(This is lemma 7.13 listed on page 102 in van der Vaart (1998).)

If the sequence of subsets H_n of $\Re^k$ converges to a nonempty set H and the sequence of random vectors X_n converges in distribution to a random vector X, then

1. $\|X_n - H_n\| \Rightarrow \|X - H\|$.
2. $\|X_n - H_n \cap F\| \geq \|X_n - H \cap F\| + o_p(1)$, for every closed set F.
3. $\|X_n - H_n \cap G\| \leq \|X_n - H \cap G\| + o_p(1)$, for every open set G.

$\diamondsuit$

Lemma 5:

(This is theorem 18.14 listed on page 261 in van der Vaart (1998).)

A sequence of arbitrary maps $X_n : \Omega_n \mapsto l^\infty(T)$ converges weakly to a tight random element if and only if both of the following conditions hold:

1. The sequence $(X_{n,t_1}, X_{n,t_2}, ..., X_{n,t_k})$ converges in distribution in $\Re^k$ for every finite set of points $t_1, t_2, ..., t_k$ in T;

2. for every $\epsilon, \eta > 0$ there exists a partition of T into finitely many sets $T_1, T_2, ..., T_k$ such that

$$\limsup_{n\to\infty} P\left(\sup_i \sup_{s,t\in T_i} |X_{n,s} - X_{n,t}| \geq \epsilon\right) \leq \eta.$$

$\diamondsuit$

Remark 1:

Given a sequence of functions $g_n : Z^+ \to R$ such that $g_n(m) \geq 0$ for all n, m, and for all $m \in Z^+$, $\lim_{n\to\infty} g_n(m) = 0$, there exists a sequence of positive integers $M_1 \leq M_2 \leq M_3 \leq ...M_k \leq ...$ such that $\lim_{k\to\infty} M_k = \infty$ and $\lim_{k\to\infty} g_k(M_k) = 0.\diamondsuit$

Proof:

Note that there exists N_2 such that $n \geq N_2$ implies $g_n(2) < \frac{1}{2}$, and
there exists $N_3 > N_2$ such that $n \geq N_3$ implies $g_n(3) < \frac{1}{3}$, and
there exists $N_4 > N_3 > N_2$ such that $n \geq N_4$ implies $g_n(4) < \frac{1}{4}$.
Continue to obtain $N_2 < N_3 < N_4 < ... < N_k < ...$ such that $n \geq N_k$ implies $g_n(k) < \frac{1}{k}$.
Now consider the sequence $g_1(1), g_2(1), ..., g_{N_2-1}(1), g_{N_2}(2), g_{N_2+1}(2), ...,$
$g_{N_3-1}(2), g_{N_3}(3), g_{N_3+1}(3), ..., g_{N_4-1}(3), g_{N_4}(4), g_{N_4+1}(4), ...$
Note that if $g_m(k)$ is a term of this sequence, $k \geq 2$ gives $g_m(k) < \frac{1}{k}$. Thus, the limit of the sequence is zero.
(Note this remark is used in the proof of Theorem 1 with

$$g_n(m) = P_{n,\theta}\left(\left\{w : \sup_{\|h\|\leq m} |Z_n(h)(w)| > \epsilon\right\}\right) \text{ for } \epsilon > 0.)$$

$\diamondsuit$

Based on Theorem 1, an asymptotic size α test of H_o versus H_a rejects H_o whenever $\Lambda_n > \chi^2(1-\alpha, c-p)$, the upper α probability point of a $\chi^2(c-p)$ distribution. In addition,

the noncentrality parameter δ depends only on the information metric in the expectation surface parameter directions. In particular,

$$(0,\gamma,0)^T I_n(\beta,0,\eta)(0,\gamma,0) = (0,\gamma)^T I_n^E(\beta,0)(0,\gamma)$$

where I_n^E denotes the Fisher information metric for the expectation surface parameters (β,γ) at the n^{th} stage.

Since the power of the preceding LRT is an increasing function of the noncentrality parameter, Theorem 1 indicates the importance of the $\hat{l}_{Z_n(\beta)}$. In particular, the convergence $(1/n) I_n^E(\beta,0) \rightarrow I^E(\beta,0)$ implies the convergence $\hat{l}_{Z_n(\beta)} \rightarrow \hat{l}(\beta)$, say. Now consider the MMBCLFE sequence $E_n\left(x_{n1},...,x_{nn}; Z_n^M(\beta) \in \kappa_n, \beta \in B\right)$ with its information metrics $(1/n) I_n^{E,M}(\beta,\gamma)$ and $\hat{l}_{Z_n^M(\beta)}$. Also, let $\hat{l}^M(\beta) = \liminf \hat{l}_{Z_n^M(\beta)}$. The following theorem, given in Neill and Miller(2003), together with Theorem 1 above, justifies the conclusion that for any n, the MMBCLFE is a good choice to use in testing for lack of fit in terms of optimal power.

Theorem 2:

If $E_n(x_{n1},...,x_{nn}; Z_n(\beta) \in \kappa_n, \beta \in B)$ is any sequence of BCLFEs as described above satisfying the hypotheses of Theorem 1, then $\hat{l}^M(\beta) \geq \hat{l}(\beta)$.$\diamondsuit$

Chapter 6

Conclusions and Future Research

Using the exponential, cosine and Michaelis-Menten nonlinear models, it was demonstrated that maximin clustering can vary substantially across different points along the expectation surface of a proposed model. Such leads to constructed alternative models, and hence associated LRTs, which can yield good power for detecting nonlinear between-cluster lack of fit.

However, depending on curvature considerations, certain models did not provide substantially different maximin clusterings across the expectation surface. Accordingly, for computational simplicity, it was of interest to study the case of constructed alternative models based on a single maximin clustering. This approach was investigated in parallel with the case in which the clusters were allowed to vary across the expectation surface. The use of a single clustering was observed to be competitive in certain cases but not all cases.

In practice, it is of course not known which approach to necessarily use due to the unknown potential lack of fit involved. Thus, a multiple testing approach utilizing multiple and single maximin clustering strategies was also considered. The Bonferroni adjusted testing procedure was observed to be reasonably successful in detecting between-cluster lack of fit.

The LRTs were based on constructed alternative models which account for between-cluster lack of fit. Data was generated accordingly and the performance of the proposed tests was

observed to be reasonably good. However, it was also of interest to subject these tests to data generators which were functionally different than the proposed models. Such an investigation was carried out as well, using the multiple and single maximin clustering strategies, along with the multiple testing approach. The simulated powers were consistent with the similarity or departure from the proposed models.

In addition to the use of maximin clusterings, it was also of interest to consider cluster-based LRTs using structured but non-maximin clusterings. The simulated power results indicate that the use of maximin clusterings in constructing alternative models is generally superior to that of using non-maximin clusterings.

Finally, the proposed cluster-based LRTs were compared to the generalized LRT with semiparametric alternative model as presented by Ciprian and Ruppert (2004). Notably, the maximin clustering based LRTs were markedly better in simulated power performance. In fact, the non-maximin cluster-based LRTs were generally observed to possess better simulated power than that associated with the test given by Ciprian and Ruppert. In addition, Ciprian and Ruppert compared their test with a lack of fit test proposed by Fan and Huang (2001). For the cases considered in the comparison simulation study, the test presented by Ciprian and Ruppert performed better then the test given by Fan and Huang. The test presented by Fan and Huang employs the adaptive Neyman test on the Fourier transform of the residual vector to assess the adequacy of a parametric regression model against nonparametric alternatives. Thus, by implication, the proposed maximin cluster-based LRTs are superior to the test given by Fan and Huang, as well, at least for the simulation cases considered.

Future research of interest would involve the following broad points.

1. Additional comparison with other nonlinear lack of fit tests in order to investigate more

fully the effectiveness of the proposed maximin cluster based LRTs. It is of interest in particular to compare the proposed tests with tests based on nonparametric smoothed alternatives. Included in such comparisons would be nonlinear models with more than one predictor.

2. Extend the proposed testing procedure to cover the case of nonlinear within-cluster lack of fit, as well as mixtures of nonlinear between- and within-cluster lack of fit. The multiple testing procedures presented by Miller and Neill (2008) may be useful for such extension.

3. Generalize the testing procedures investigated in this dissertation to accommodate exponential family regression models (e.g. Kass and Vos (1997)). Such generalization would include nonlinear models without additive error, parametric regression models with heterogeneous variance and generalized linear models.

Bibliography

[1] Aerts, M., Claeskens, G. and Hart, J.D. (1999). Testing the Fit of a Parametric Function, *Journal of the American Statistical Association*, 94, 869-879.

[2] Aerts, M., Claeskens, G. and Hart, J.D. (2000). Testing Lack of Fit in Multiple Regression, *Biometrika*, 82, 405-424.

[3] Aerts, M., Claeskens, G. and Hart, J.D. (2004). Bayesian-Motivated Tests of Function Fit and Their Asymptotic Frequentist Properties, *Annals of Statistics*, 32, 2580-2615.

[4] Akritas, M. G. and Papadatos, N. (2004). Heteroscedastic One-Way ANOVA and Lack of Fit Tests, *Journal of the American Statistical Association*, 99, 368-382.

[5] Akritas, M. G. and Wang, L. (2009). Lack-of-Fit Tests in Nonparametric Random Design Regression, *Submitted for Publication*.

[6] Billinglsey, P. (1995). *Probability and Measure, 3rd edition*, John Wiley ans Sons Inc., New York.

[7] Caouder, N. and Huet, S. (1997). Testing Goodness-of-Fit for Nonlinear Regression Models with Heterogeneous Variances, *Computational Statistics and Data Analysis*, 23, 491-507.

[8] Christensen, R.R. (1989). Lack of Fit Based on Near or Exact Replicates, *Annals of Statistics*, 17, 673-683.

[9] Christensen, R.R. (1991). Small Sample Characterizations of Near Replicate Lack of Fit Tests, *Journal of the American Statistical Association*, 86, 752-756.

[10] Christensen, R.R. (2002). *Plane Answers to Complex Questions:The Theory of Linear Models, 3rd edition*, Springer, New York.

[11] Ciprian, C. and Ruppert, D. (2004). Likelihood Ratio Tests for Goodness of Fit of a Nonlinear Model, *Journal of Multivariate Analysis*, 91, 35- 52.

[12] Claeskens, G. (2004). Restricted Likelihood Ratio Lack of Fit Tests using Mixed Spline Models, *Journal of the Royal Statistical Society Series B*, 66, 4, 909- 926.

[13] Cook, R.D. and Weisburg, S. (1997). Graphics for Assessing the Adequacy of Regression Models, *Journal of the American Statistical Association*, 92, 490- 499.

[14] Cox, D.R. and Reid, N. (1987). Parameter Orthogonality and Approximate Conditional Inference (with discussion), *Journal of the Royal Statistical Society Series B.* 49, 1-39.

[15] Dette, H. (1999). A Consistent Test for the Functional Form of a Regression Based on a Difference of Variance Estimators, *Annals of Statistics*, 27, 1012-1040.

[16] Diebolt, J. and Zuber, J. (1999). Goodness of Fit Tests for Nonlinear Heteroscedastic Regression Models, *Statistics and Probability Letters*, 42, 53-60.

[17] Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York.

[18] Eubank, R.L., Chin-Shang, L. and Wang, S. (2005). Testing Lack-of-Fit of Parametric Regression Models Using Nonparametric Regression Techniques, *Statistica Sinica*, 15, 135-152.

[19] Fan, J. and Huang, L.-S. (2001). Goodness of Fit Tests for Parametric Regression Models, *Journal of the American Statistical Association*, 96, 640-652.

[20] Fan, J., Zhang, C. and Zhang, J. (2001). Generalized Likelihood Ratio Statistics and Wilks Phenomenon, *Annals of Statistics*, 29, 153-193.

[21] Fisher, R.A. (1922). The Goodness of Fit of Regression Formulas and the Distribution of Regression Coefficients, *Journal of the Royal Statistical Society*, 85, 597-612.

[22] Guo, H. and Koul, H.L. (2008). Asymptotic Inference in Some Heteroscedastic Regression Models with Long Memory Design and Errors, *Annals of Statistics*, 36, 458-487.

[23] Guerre, E. and Lavergne, P. (2005). Data-Driven Rate-Optimal Specification Testing in Regression Models, *Annals of Statistics*, 33, 840-870.

[24] Hardle, W. and Mammen, E. (1993). Comparing Nonparametric Versus Parametric Regression Fits, *Annals of Statistics*, 21, 1926-1947.

[25] Hart, J.D. (1997). *Nonparametric Smoothing and Lack of Fit Tests*, Springer-Verlag, New York.

[26] Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning*, Springer, New York.

[27] Kass, R. E. and Vos, P. W. (1997). *Geometrical Foundations of Asymptotic Inference*, Wiley, New York.

[28] Khmaladze, E.V. and Koul H.L. (2004). Martingale Transforms Goodness of Fit Tests in Regression Models, *Annals of Statistics*, 32, 995-1034.

[29] Koul, H.L. and Ni, P. (2004). Minimum Distance Regression Model Checking, *Journal of Statistical Planning and Inference*, 119, 109-141.

[30] Koul, H. L. and Song, W. (2008). Model Checking with Berkson Measurement Error, *Journal of Statistical Planning and Inference*, 138, 1615-1628.

[31] Koul, H. L. and Song, W. (2009). Minimum Distance Regression Model Checking with Berkson Error, *Annals of Statistics*, 37, 132-156.

[32] Le Cam, L. M. (1972). Limits of Experiments, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 245-261, University of California Press, Berkeley.

[33] Le Cam, L. M. and Yang, G. L. (2001). *Asymptotics in Statistics:Some Basic Concepts, 2nd edition*, Springer, New York.

[34] Lee, G. and Hart, J. (1998). An L_2 Error Test with Order Selection and Thresholding, *Statistics and Probability Letters*, 39, 61-72.

[35] Li, C. S. (2005). Using Local Linear Kernel Smoothers to Test the Lack of Fit of Nonlinear Regression Models, *Statistical Methodology*, 2, 267- 284.

[36] Loader, C. (1999). *Local Regression and Likelihood*, Springer, New York.

[37] Lopez, O. and Patilea, V. (2009). Nonparametric Lack-of Fit Tests for Parametric Mean-Regression Models with Censored Data, *Journal of Multivariate Analysis*, 100, 210-230.

[38] Matsumoto, C. and Wakaki, H. (2006). Asymptotic Expansion for Lack of Fit Test Under Nonnormality, *Hiroshima Mathematics Journal*, 36, 29-38.

[39] Miller, F.R., Neill, J.W. and Sherfey, B.W. (1998). Maximin Clusters for Near Replicate Regression, *Annals of Statistics*, 26, 1411-1433.

[40] Miller, F.R., Neill, J.W. and Sherfey, B.W. (1999). Implementation of a Maximin Power Clustering Criterion to Select Near Replicates for Regression Lack of Fit Tests, *Journal of the American Statistical Association*, 94, 610-620.

[41] Miller, F.R. and Neill, J.W. (2008). General Lack of Fit Tests Based on Families of Groupings, *Journal of Statistical Planning and Inference*, 138, 2433-2449.

[42] Neill, J.W. (1988). Testing for Lack of Fit in Nonlinear Regression, *Annals of Statistics*, 16, 733-740.

[43] Neill, J.W. and Miller, F.R. (2003). Limit Experiments, Lack of Fit Tests and Fuzzy Clusterings, Technical Report II-03-1, Department of Statistics, Kansas State University.

[44] Neill, J.W., Miller, F.R. and Brown, D. D. (2002). Maximin Clusters for Nonreplicated Multiresponse Lack of Fit Tests, *Journal of Statistical Planning and Inference*, 102, 359-375.

[45] Presnell, B. and Boos, D. D. (2004). The IOS Test for Model Misspecification, *Journal of the American Statistical Association*, 99, 216-227.

[46] Rieder, H. (1994). *Robust Asymptotic Statistics*, Springer-Verlag, New York.

[47] Rohde, A. (2008). Adaptive Goodness-of-Fit Tests Based on Signed Ranks, *Annals of Statistics*, 36, 1346-1374.

[48] Rudin, W. R. (1991). *Functional Analysis, 2nd edition*, McGraw-Hill, New York.

[49] Schucany, W.R. (2004). Kernel Smoothers: An Overview of Curve Estimators for the First Graduate Course in Nonparametric Statistics, *Statistical Science*, 19, 663- 675.

[50] Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*, Wiley, New York.

[51] Schimek, M. G. (2000). *Smoothing and Regression*, Wiley, New York.

[52] Shiryaev, A. N. and Spokoiny, V. G. (2000). *Statistical Experiments and Decisions*, World Scientific Publishing Company, New Jersey.

[53] Skovgaard, L.T. (1984). A Riemannian Geometry of the Multivariate Normal Model, *Scandinavian Journal of Statistics*, 11, 211-223.

[54] Stute, W. (1997). Nonparametric Model Checks for Regression, *Annals of Statistics*, 25, 613-641.

[55] Stute, W., Manteiga, G.W. and Quindimil, P.M. (1998a). Bootstrap Approximations in Model Checks for Regression, *Journal of the American Statistical Association*, 93, 141-149.

[56] Stute, W., Thies, S. and Zhu, L. (1998b). Model Checks for Regression: an Innovation Process Approach, *Annals of Statistics*, 26, 1916-1934.

[57] van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.

[58] Wang, L., Akritas, M. G. and van Keilegom, I. (2008). ANOVA-Type Nonparametric Diagnostic Tests for Heteroscedastic Regression Models, *Journal of Nonparametric Statistics*, 20, 365-382.

[59] Zheng, J.X. (1996). A Consistent Test of Functional Form Via Nonparametric Estimation Technique, *Journal of Econometrics*, 75, 263-289.

Appendix A

Generated Data Set 1

0.06524904	0.99421674	1.73514893	2.05203975	2.12335588	2.26580093
2.42592366	2.50049595	3.36624638	4.23561146	4.40980336	5.15355938
5.19080359	5.31842467	5.42894020	5.44499426	5.45463420	6.01582744
6.29080697	6.56368798	7.43721835	7.45316243	7.94172121	8.38902923
8.46787080	8.61031176	9.01535931	9.13918048	9.50638337	9.55264658

Table A.1: *Generated and sorted* $n = 30$ *Uniform [0,10] values*

Appendix B

Generated Data Set 2

0.1443391	0.1704265	0.1967341	0.3001186	0.3188816	0.3928139	0.5190332
0.7557029	0.8380201	0.8513597	0.9308813	1.0290017	1.1350898	1.1446759
1.6118328	1.6660910	1.6980304	1.7015172	1.9156087	2.0364770	2.0795168
2.1913855	2.2286743	2.2543662	2.2890394	2.3958913	2.4586639	2.4900805
2.4988047	2.5734157	2.6417767	2.7230507	2.7431890	2.7453052	2.7548386
3.0676851	3.5432806	3.5556869	3.5589774	3.5804998	3.7276310	3.8162769
3.9698674	3.9879073	3.9983256	4.0584997	4.2263761	4.2690767	4.2880942
4.2967153	4.4750805	4.5947319	4.6439198	4.6743405	4.6893548	4.7314146
4.8610035	4.8850594	5.0050323	5.0528560	5.0747820	5.1715688	5.2573932
5.3442678	5.3559704	5.4060103	5.4585984	5.6773775	5.7335645	5.8322197
5.9193492	5.9592531	6.1535242	6.1582931	6.2354960	6.4135658	6.5165567
6.8544451	6.9310208	7.0664691	7.1140908	7.2896456	7.4383940	7.4774647
7.6112174	7.7077153	7.7510990	7.9799301	8.0154700	8.2265258	8.3361919
8.3477833	8.3613414	8.3828767	8.6138244	8.6472123	8.9983245	9.3643254
9.5465365	9.6130241					

Table B.1: *Generated and sorted* $n = 100$ *Uniform [0,10] values*

Appendix C

Generated Data Set 3

3.014575	3.780258	5.638315	9.151028	12.348723	12.523909	13.028887
17.026205	17.142021	18.040725	19.867907	20.461216	20.771390	20.981677
21.140856	22.990589	23.569430	24.509259	24.874240	25.339065	25.767250
27.488666	27.772376	28.035384	30.073652	30.708590	30.776611	32.772587
33.052985	33.066053	34.869198	35.752485	35.777378	35.813799	35.947511
37.032054	38.943930	39.848790	41.223704	42.010145	44.514802	44.541398
44.829914	45.573146	45.952549	46.370118	46.854928	48.377074	48.830599
49.123182	51.745975	53.581115	53.834870	54.655860	54.909671	55.232243
57.256477	57.793740	59.132105	59.757529	59.939816	60.332436	62.469772
62.499648	62.939085	64.710119	66.177902	66.902171	67.639817	69.029053
69.435071	69.527414	71.080385	73.331417	73.346670	74.897223	76.255108
77.030161	77.180549	77.477889	77.758444	78.035851	81.240262	82.730345
83.402710	85.665304	86.512055	88.195359	88.216552	88.422703	88.945354
90.642643	90.695438	91.972191	92.850507	95.415771	96.057309	96.699908
98.282408	98.956414					

Table C.1: *Generated and sorted* $n = 100$ *Uniform [0,100] values*

Appendix D

R Code for Implementing Nonlinear Between-Cluster Lack-of-Fit Tests for Exponential Model with Additive Normal Errors

```
 n<-100
 k<-0
 l<-10
 b<-.1
 m<-5
 #set.seed(1235)# for n=30#
set.seed(10)# for n=100#

x<-runif(n, k,l) #create n unifrom(k,l) random variables#
xs<-sort(x*b) #scaling & sorting x#

cover<-function(n,b,k,l,m,xs)
{
#cover i.e. [k,l] divided into m(=5) cells#
# here we consider cells based on values c1:[0-2), c2:[2-4)... c5:[8-10]#
```

```
c1<-ifelse(xs<((l-k)/m)*(m-4)*b, c1<-1, c1<-0)
c2<-ifelse(xs<((l-k)/m)*(m-3)*b,
ifelse(xs>=((l-k)/m)*(m-4)*b, c2<-1, c2<-0), c2<-0)
c3<-ifelse(xs<((l-k)/m)*(m-2)*b,
ifelse(xs>=((l-k)/m)*(m-3)*b, c3<-1, c3<-0), c3<-0)
c4<-ifelse(xs<((l-k)/m)*(m-1)*b,
ifelse(xs>=((l-k)/m)*(m-2)*b, c4<-1, c4<-0), c4<-0)
c5<-ifelse(xs<((l-k)/m)*m*b,
ifelse(xs>=((l-k)/m)*(m-1)*b, c5<-1, c5<-0), c5<-0)

#getting overlapping subsets based on odd cells created by cover#
#here we have 3 overlapping subsets Fn1:[0-4),Fn2:[2-8), Fn3:[6-10]#
os1<-ifelse(xs<=((l-k)/m)*(m-4)*b, 1, NA)
Fn1<-{if(length(sort(os1))<1) NA else c1+c2}
os2<-ifelse(xs<=((l-k)/m)*(m-2)*b,
ifelse(xs>((l-k)/m)*(m-3)*b, 1, NA),NA)
Fn2<-{if(length(sort(os2))<1) NA else c2+c3+c4}
os3<-ifelse(xs<=((l-k)/m)*m*b,
ifelse(xs>((l-k)/m)*(m-1)*b, 1, NA),NA)
Fn3<-{if(length(sort(os3))<1) NA else c4+c5}

z<-data.matrix(data.frame(Fn1,Fn2,Fn3))
#matrix indicating overlapping subsets based on selected cover#
ca<-(sum(c2)+1)*(sum(c4)+1)#cardinality of Kn,0#
B<-matrix(rep(0, (n*(n-1))), nrow=n) #matrix for the edge sets#
out<-list(z,B,ca,c1,c2,c3,c4,c5)
return(out)
```

```
}
cover.out<-cover(n,1,k,l,5,xs)#

#####################################################################
#Finding tangent space to the expectation surface at each x.
#Exponential model f(x, beta1, beta2)=beta1*exp(-x*beta2) is used.
#####################################################################
tm<-function(beta1, beta2,xs)
{
b1<-exp(-x*beta2) #partial derivative wrt beta1#
b2<-(-x)*beta1*exp(-x*beta2) #partial derivative wrt beta2#
b1n<-b1/sqrt(sum(b1^2))
b2n<-b2/sqrt(sum(b2^2)) #normalized b1 & b2#
TM<-cbind(b1n,b2n)
TM
}

#####################################################################
#Finding perpendicular projection operator for a vector space.
#####################################################################
#c=matrix that we want to project onto#
#need to assign the corresponding matrix to c before using this function#
ppo<-function(c)
{
library(MASS) #need for g-inverse function#
PPO<-c%*%ginv(t(c)%*%c)%*%t(c)
PPO
```

```
}

#####################################################################
#Finding basis set for the matrix based on output from ppo function.
#####################################################################
basis<-function(ppo.out)
{
Ba<-ppo.out%*%t(ppo.out)
m<-nrow(Ba)
e<-eigen(Ba,symmetric=T )#getting eigen values & vectos#
evals<-e$values
evects<-e$vectors
nevals<-evals[abs(evals)>10e-6]
#avoid the problem of decemal places used in R (to get correct d.f.)#
n<-length(nevals)
BASIS<-evects[1:m,1:n] #final basis set#
BASIS
}

#####################################
#####################################
#Selecting maximin power clusterings
#####################################
#####################################

#########################################################################
#New method to create crisp atoms using c(=3) covers and ordered dial
```

```
#setting values corresponds to even cells.
#Let k obs in 1st even cell i.e. {x(11), x(12),... x(1k)} &
#l obs in 2nd even cell i.e. {x(21), x(22),... x(2l)}.
#1st pointer in x(11), 2nd pointer moves from x(21) to x(2l) sequentially
#1st pointer in x(12), 2nd pointer moves from x(21) to x(2l) sequentially
# so on & so forth....
#########################################################################
# z=output for overlapping subsets from cover function#
# beta1 & beta2 are exponential model parameters#
# ca=cardinality of Kn,0 from cover function#
# c1,c2,c3,c4,c5=elements corresponding to the 5 cells,
# {c1,c3,c5}=odd cells, {c2,c4}=even cells#

z<-cover.out[[1]]
ca<-cover.out[[3]]
c1<-cover.out[[4]]
c2<-cover.out[[5]]
c3<-cover.out[[6]]
c4<-cover.out[[7]]
c5<-cover.out[[8]]

atoms<-function(z,ca,c1,c2,c3,c4,c5,beta1,beta2)
{
library(MASS)
ZZ<-matrix(rep(0,((dim(z)[1])*(dim(z)[2]))),nrow=dim(z)[1],ncol=dim(z)[2])
#create an empty crisp atoms matrix#
for (h in 1:ca) #loop goes through each crisp atom#
```

```
{
ZZa<-ZZ
for (i in 1:dim(z)[1]) #fixing elements in odd cells#
{
ZZa[i,1]<-c1[i]
ZZa[i,2]<-c3[i]
ZZa[i,3]<-c5[i]
}
  for (j in (sum(c1)):(sum(c1)+sum(c2)))
#consider elements in 1st even cell corresponds to 1st column#
{
ZZa[j,1]<-1
for (k in (j+1):(sum(c1)+sum(c2)))
#consider elements in 1st even cell corresponds to 2nd column#
ZZa[k,2]<-1
if (j>(sum(c1)))
#cleaning up extra 1's in second column before the required element#
{
for (l in (sum(c1)):j)
ZZa[l,2]<-0
}
for (n in (sum(c1)+sum(c2)+sum(c3)):(sum(c1)+sum(c2)+sum(c3)+sum(c4)))
#consider elements in 2nd even cell corresponds to 2nd column#
{
ZZa[n,2]<-1
for (o in (n+1):(sum(c1)+sum(c2)+sum(c3)+sum(c4)))
#consider elements in 2nd even cell corresponds to 3rd column#
```

```
ZZa[o,3]<-1
if (n>(sum(c1)+sum(c2)+sum(c3)))
#cleaning up extra 1's in third column before the required element#
{
for (p in (sum(c1)+sum(c2)+sum(c3)):n)
ZZa[p,3]<-0
}
for (q in 1:dim(z)[1])
#cleaning up extra 1's in second column after the required elements#
{
if (sum(ZZa[q,])>1)
ZZa[q,2]<-0
}
cat(h,file="atoms.txt","\n",append=TRUE)
#assign a label for the atom stored in R#
for (v in 1:(dim(z)[1]))#consider the current atom in stored file in R#
{
cat(ZZa[v,],file="atoms.txt","\n",append=TRUE)
}
PZZ<-ppo(ZZa)-ppo(ppo(ZZa)%*%tm(beta1,beta2,xs))
#gives perpendicular projection of tm(beta1, beta2) onto zza (crisp atoms)#
PZZb<-basis(PZZ)#basis set for pzz, use for calculation of tau#
cat(PZZb,file="Basis.txt","\n",append=TRUE) #stores the basis set in R#
   }
 }
 }
}
```

```
#########################################################################
#Running atoms function for different values for beta2 i.e. 1,3,5,7,9#
atoms(z,ca,c1,c2,c3,c4,c5,1,1)
atoms(z,ca,c1,c2,c3,c4,c5,1,3)
atoms(z,ca,c1,c2,c3,c4,c5,1,5)
atoms(z,ca,c1,c2,c3,c4,c5,1,7)
atoms(z,ca,c1,c2,c3,c4,c5,1,9)
#########################################################################

#########################################################
#Function to get w values using correct x values
#########################################################
# v=z=output for overlapping subsets from cover function#
# vb=xs=the set of sorted random variables, X,#
# vc=B=the matrix B created by the edges function#

w.values<-function(v,vb ,vc)
{
for (a in 1:(dim(v)[1]-1))
{
if (sum(vc[,a])>1)
break
}
for (ab in 1:dim(v)[1])
{
if (vc[ab,a]==1)
```

```
break
}
for (ac in ab+1:dim(v)[1])
{
if (vc[ac,a]==1)
break
}
w<-(((vb[ab])-(vb[ac]))^2)
#weight for a particular element in edge set#
cat(w,file="weights.txt","\n",append=TRUE)
#need to calculate W_Z_n values later#
}

#############################################################################
#recursive function to create the edge set (w.values function used in here)
#############################################################################
# z=output for overlapping subsets from cover function#
# B=matrix B which is created by cover function#
# f=row identifier for z matrix#
# g=column identifier for z matrix#
# h=column identifier for 'b' matrix#
# b= matrices which are created by edges function#

z<-cover.out[[1]]
B<-cover.out[[2]]
edges<-function(z,b,f,g,h)
{
```

```
#primary recursion loop for calculation of edge set#
for (j in 1:(dim(z)[1]))
#indices where to start in the rows of Z (lengths of the different columns)#
if (z[j,g]==1)
{
break
}
#identifying the observations which can be clustered into different subsets#
if (g<(dim(z)[2])) #used for stopping criteria#
{
for (ja in 1:(dim(z)[1]))
if (z[ja,g]+z[ja,g+1]==2) #find next contiguous obs consistent with a cover#
{
break
}
}
else ja<-j
if (z[f,g]==1) #find if obs in the same overlapping subset are contiguous#
{
if (z[f+1,g]==1)
{
b[f,h]<-1 #insert 1 into empty edge set#
if (h>1)
{
#takes care of diagonals before the lone clustering#
for (jb in 1:h-1)
{
```

```
b[jb,jb]<-1
}
}
#the main body of the algorithm which supplies the most criteria#
for (i in (f+1):(sum(z[,g])+(j-1)))
{
b[i,h]<-1

if (i==f+1)
{
for (jc in i:(sum(z[,g])+(j-1)))
#ensure edge remain consistent with cover#
b[jc,jc-1]<-1
}
if (i>f+1)
{
b[i-1,i-1]<-1
}
if (i>f+1)#cleaning up extra 1's in the matrix#
{
for(jd in (f+1):(i-1))
b[jd,h]<-0
}
for (je in (sum(z[,g])+(j-1)):(dim(z)[1]))
{
b[je,je-1]<-1
}
```

```
if (i!=f+1) b[i,i-1]<-0 #cleaning up the edge matrix#
{
w.values(z,xs,b) #find weight component for the edge#
}
}
if (g<=(dim(z)[2]) || f<(dim(z)[1]) || h<(dim(b)[2]))
# instructions about what to do next#
{
if (g<(dim(z)[2]))
# if there is another overlapping subset to investigate do this#
 {
 for (y in 1:(dim(z)[1])) #create identifier for rows of z#
 if (z[y,g+1]==1)
 {
 break
 }
 if ((f+1)<y) edges(z,B,f+1,g,h+1) #decision for recursion#
 else edges(z,B,y,g+1,y) #decision for recursion#
  }
  else
  {
#Overall stopping criteria#
  if (h>=(dim(b)[2]) || f>=(dim(z)[1]-1))
  date() #gives date/time group at the stopping point#
  else edges(z,B,f+1,g,h+1)# go with next row of z & next col of b if not stopped#
  }
 }
```

```
}
  else
  if (g<=(dim(z)[2]) || f>(dim(z)[1]) || h<(dim(b)[2]))
  # push to search to the next overlapping subset#
  {
  for (y in 1:(dim(z)[1]))#next overlapping subset to build edge clusters for#
  if (z[y,g+1]==1)
   {
   break
   }
   edges(z,B,y,g+1,y)#next recursion#
   }
}
}

#############################################################################
# function to calculate weights (W_Z_n) for Z_n that belongs to edge set
#using edges function Note: calculates the weights once the recursive algorithm
#'edges' has stored in the output
#############################################################################
weights<-function()
{
edges(z, B, 1,1,1)
#call function that creates the edge set, this changes as f, g, h values changes#
for (ii in 1:dim(as.matrix(read.table("weights.txt")))[1])
{
ww<-(as.matrix(read.table("weights.txt")))[ii,]/
```

```
(sum(as.matrix(read.table("weights.txt"))))
#calculates W_Z_n for each atom in Kn,0#
cat(ww,file="totalweights.txt","\n",append=TRUE)#use to calculate l_Z_n values later#
}
}

##################################
#Runing weights function#
weights()
##################################
weight<-as.matrix(read.table("totalweights.txt"))
#creates a matrix of weights to use later#
basis<-as.matrix(read.table("Basis.txt"))
#creates a matrix of which each row is a basis for corresponding atom#

#########################################################
#Function which creates tau (lzn function used in here)
#########################################################
# z=output for overlaping subsets from cover function#
# B=matrix B which is created by cover function#
# f=row identifier for z matrix#
# g=column identifier for z matrix#
# h=column identifier for 'B' matrix#
# vv=index counter for the 'weights' matrix from weights function#
# vvv=index counter for the basis's rows from atoms function#

z<-cover.out[[1]]
```

```
B<-cover.out[[2]]
edges_beta<-function(z,b,f,g,h,vv,vvv)
{
if (vv==1)
ZZZ<<-0 #set the starting value of the tau at 0

################################################################
#Function to calculate the set of l_Z_n values (for each atom)
################################################################
#d=the matrix Z from cover function
#da=the matrix B which is calculated by the recursive algorithm (edges function)
#db= created as an index counter for edge sets
# 'weight' W_Z_n values based on edge sets#

lzn<-function(d,da,db)
{
tau_1<-t(as.matrix(vvv))%*%(ppo(da)-ppo(ppo(da)%*%tm(1,1,xs)))%*%(as.matrix(vvv))
#parameters in tm function need to change based on the situation
tau<-weight[db]%*%tau_1
ZZZ<<-ZZZ+tau
#keep the running total of tau while moving through the edge set
if (db>=(length(weight)))
#what to do when the entire edge set has been run #
#(before 'apply'takes this to the next atom)
{
#print((dim(d)[1])*(1/(ZZZ)))
cat(((dim(d)[1])*(1/ZZZ)),file="lzn(1,1).txt","\n",append=TRUE)
```

```
#store the values of l_Z_n before iterating to next atom in K_n_0
}
}
##primary recursion loop for the calculation of the edge set
{
for (j in 1:(dim(z)[1]))
#indicates where to start searching in the rows of z (lengths of the different cols)
if (z[j,g]==1)
{
break
}
if (g<(dim(z)[2])) #used for stoping criteria (last column if g=column dim of z)
{
for (ja in 1:(dim(z)[1]))
if (z[ja,g]+z[ja,g+1]==2)
#identifies the 'next' contiguous observations consistent with a cover
{
break
}
}
else ja<-j
if (z[f,g]==1)
#for identifing if the observations in the same overlap subset are contiguous
{
if (z[f+1,g]==1)
#identifies if the observations in the same overlap subset are contiguous
{
```

```
b[f,h]<-1 #inserts a 1 into the empty edge matrix, b
if (h>1)
{
for (jb in 1:h-1)
#places lone observations before the lone clustering
{b[jb,jb]<-1}
}
for (i in (f+1):(sum(z[,g])+(j-1)))
#for 1s at the clustering and observations after (obs are sorted)
{
b[i,h]<-1
if (i==f+1)
{
for (jc in i:(sum(z[,g])+(j-1)))
#for ensuring edge remains consistent with cover
b[jc,jc-1]<-1
}
if ((i-f-1)>0) #for ensuring edge remains consistent with cover
{
b[i-1,i-1]<-1
}
if (i>f+1) #'cleaning up' the extra 1s in the matrix
{
for(jd in (f+1):(i-1))
b[jd,h]<-0
}
for (je in (sum(z[,g])+(j-1)):(dim(z)[1]))
```

```
{
b[je,je-1]<-1
}
if (i!=f+1) b[i,i-1]<-0 #'cleaning up' the edge matrix
#print(b)
lzn(z,b,vv) #call to the lz function
vv<-vv+1
#increase the counter which is used to reset the ZZZ
#(running weight value within an edge set)
}
if (g<=(dim(z)[2]) || f<(dim(z)[1]) || h<(dim(b)[2]))
#instructions for the recursion (what to do next)
{
if (g<(dim(z)[2]))
#if there is another overlapping subset to investigate then....
{
for (t in 1:(dim(z)[1])) #create identifier for the rows of z
if (z[t,g+1]==1)
{
break
}
if ((f+1)<t) edges_beta(z,B,f+1,g,h+1,vv,vvv) #decisions for the recursion
else edges_beta(z,B,t,g+1,t,vv,vvv) #decisions for the recursion
}
else
{
if (h>=(dim(b)[2]) || f>=(dim(z)[1]-1)) #stopping criteria for the recursion
```

```
date() #prints date/time group at stopping criteria
else edges_beta(z,B,f+1,g,h+1,vv,vvv)
#keep going with next row of z and next col of B if stopping criteria isn't reached!
}
}
}
else
if (g<=(dim(z)[2]) || f>(dim(z)[1]) || h<(dim(b)[2]))
#criteria to push the search to the next overlapping subset
{
for (t in 1:(dim(z)[1]))
#identifies the next overlapping subset to build edge clusters for
if (z[t,g+1]==1)
{
break
}
edges_beta(z,B,t,g+1,t,vv,vvv) #instructions to the recursion
}
}
}
}



result<-function(v)
{
edges_beta(z,B,1,1,1,1,v)
}
```

```
#############################################################################
#Runing above 2 functions & finding lzn values#
#'basis' basis set for crisp atoms (basis changes as beta changes but atoms)#
apply(basis,1,result)
#############################################################################
lzn_beta<-as.matrix(read.table("lzn(1,1).txt"))
print(max(lzn_beta)) #prining max of lzn(beta) values
index<-seq(1:ca)
test<-cbind(index, lzn_beta)
z_n_star_beta<-test[(test[,2])==(max(lzn_beta))]
#finding crisp atom label corresponds to max of lzn(beta) value
basis_z_n_star_beta<-as.matrix(basis[(z_n_star_beta[1]),])
#finding basis corresponds to max of lzn(beta) value

#####################################################################
#####################################################################
#Simulation study for two dimensional (parameter space) case with
additive Normal errors. Specially for between cluster lack of fit
#using exponential model and g(x) function given in Miller & Neill's
#technical report
#####################################################################
#####################################################################

#Note: We have a special situation when we use exponential model.
#Since tangent vector wrt beta2 is a multiple of beta1, we get same
#maximin crisp clusterings & corresponding basis for any beta1 values
```

```
#along with a specific beta2 value. i.e. when beta1={1,3,5,7,9} & beta2=1
#we get same maximin clusters & basises
##############################################################################
#Reading best clusters based on maximin power clustering method for each
#(beta1,beta2) settings#
#for n=100#
zb11<-read.table("zb11100.txt")
zb13<-read.table("zb13100.txt")
zb15<-read.table("zb15100.txt")
zb17<-read.table("zb17100.txt")
zb19<-read.table("zb19100.txt")

#Reading corresponding basis, should change for different betas#
#for n=100#
Bb0<-read.table("Bb11100.txt")
Bb0<-read.table("Bb13100.txt")
Bb0<-read.table("Bb15100.txt")
Bb0<-read.table("Bb17100.txt")
Bb0<-read.table("Bb19100.txt")

#Bb0 <-(sqrt(n)/sqrt(sum(Bb0^2)))* Bb0  #scale to get root(n) length

#for n=100, for a single cluster
zb11<-read.table("zb15100.txt")
zb13<-read.table("zb15100.txt")
zb15<-read.table("zb15100.txt")
zb17<-read.table("zb15100.txt")
```

```
zb19<-read.table("zb15100.txt")

##################################################
# function to create LRT and test the hypothesis
##################################################
##v=soutput matrix created in 'Mn_star'
##h=snull matrix created in 'Mn_star'
##w=sZbasis matrix created in 'Mn_star'
##y=y created in 'Mn_star' (simulated observed random output)
##xs=vector of x values (xs from 'cover' function)...the sorted, scaled x values

lrt<-function(v,h,w,y,xs)
{
slrtbasis<-c(rep(0,((dim(w)[2])-2)))
#global call to create a vector of zeros to store the basis corresponding to the
closest' point to yn on Mn**. Also, find the minimum norm and associated value
#of beta for the alternative model
library(stats) #used for the chi-square quantiles computations
 ystar<-as.vector(c(0,0,max(v[,3])))
 #starting value for the search of the alternative model
for (i in 1:(dim(v)[1]))  #loop which searches through the souput matrix (for Mn**)
{
if (v[i,3]<ystar[3])
{
  ystar<-v[i,]
  }
  else next
```

```
}
cat(ystar,file="alt_result.05.txt","\n",append=TRUE)
 #store the minimum norm and beta values for the alternative model
 hstar<-as.vector(c(0,0,max(h[,3])))
 #starting value for the search in the null model
 for (j in 1:(dim(h)[1]))
 #loop which searches through the snull matrix (for null model)
{
if (h[j,3]<hstar[3])
{
hstar<-h[j,]
}
else next
}
 cat(hstar,file="null_result.05.txt","\n",append=TRUE)
 #store the minimum norm and beta values for the null model
for (k in 1:(dim(v)[1]))
#loop which finds the basis associated with the alternative
{
if ((w[k,1]==ystar[1]) & (w[k,2]==ystar[2]))
{
for (l in 3:(dim(w)[2]))
#starts with 3 since the first 2 position has the value of beta assiciated
{
b<-l-2
slrtbasis[b]<-w[k,l]
}
```

```
 }
 else next
 }
 yy<-(hstar[1])*(exp(-xs*hstar[2]))
  #nonlinear least squares approximation for the null model
yy<-as.vector(yy)
 y<-as.vector(y) #simulated yn#
 sslrtbasis<-as.matrix(slrtbasis)
 ystar_c<-((ystar[1])*(exp(-xs*ystar[2])))+((sslrtbasis%*%t(sslrtbasis))%*%(y-yy))
#nonlinear least square approximation for the alternative model
ystar_c<-as.vector(ystar_c)
 est_std_dev1<-(sum((y-ystar_c)^2))/(length(y))
 #estimated sd of errors for alt. model#
#print(est_std_dev1)
 est_std_dev2<-(sum((y-yy)^2))/(length(y))
 #estimated sd of errors for null model#
#print(est_std_dev2)
 lrtnum<-prod(dnorm(y-ystar_c,sd=sqrt(est_std_dev1)))
 #calculate the numerator of the LRT
lrtden<-prod(dnorm(y-yy,sd=sqrt(est_std_dev2)))
#calculate the denominator of the LRT
 if(lrtnum==0)
 {
 if (lrtden==0) r<-NaN
 else r<-FALSE
 }
else
```

```
 {
 lrtstat<-(2*(log(lrtnum/lrtden)))
 if(lrtstat>qchisq(.05,1,lower.tail=FALSE)) r<-TRUE
 else r<-FALSE
 }
cat(r,file="TEST.05.txt","\n",append=TRUE)
#store the value of TRUE if Ho is rejected and FALSE if Ho is NOT rejected
cat(lrtstat, file="lrt.05.txt","\n",append=TRUE) #store the LRT value #
cat(c(est_std_dev2,est_std_dev1), file="sd.05.txt","\n",append=TRUE)
#store the estimated SD values under null & alternative models#
 }


N<-100
 sZbasis<-matrix(rep(0,((length(xs)+2)*(N*N))),ncol=(length(xs)+2))
#global call to create a matrix which stores associated bases from M_n**
soutput<-matrix(rep(0,((N*N)*4)),ncol=4)
#global call to store the value of beta and associated norm
#for alternative model model
snull<-matrix(rep(0,((N*N)*3)),ncol=3)
#global call to store the value of beta and associated norm for null model
 y1<-seq(0,10, by=10/N)
y1<-y1[-(1)]
y2<-rep(y1,N) #beta2 values for the grid
soutput[,2]<-y2
snull[,2]<-y2
sZbasis[,2]<-y2
 y5<-read.table("y42.txt")
```

```
 #beta1 values for the grid
 #Note: need to rearrange values such that it is in a row or column
soutput[,1]<-t(y5)
snull[,1]<-t(y5)
sZbasis[,1]<-t(y5)

#################################################################
##Generated data
#b10,b20,g0,stdev values changes as parameter settings changes
#Bb0 changes as b10, b20 values changes
#################################################################
 Data<- function(b10,b20,g0,xs,stdev,T) {
 y<-matrix(rep(0,n*T), nrow=n, ncol=T)
 for(i in 1:T){
 eps<-rnorm(length(xs),mean=0,sd = stdev)
 y[,i]<-as.matrix(t((exp(-1*xs+g0*(xs^2))) +eps))
 y
 }
#Simulated response yn i.e. model with LOF.
#Here, Bb0 is depend on the dimension of beta vector
}

Mn_star<-function(N,xs,Gen_data,T)
{
for (p in 1:T) #loop for the number of hypothesis tests desired
 {
  sZbasis<-sZbasis
```

```
 soutput<-soutput
 snull<-snull
 library(MASS)
 y<-Gen_data[,p] #response data for one simulation run
 for (i in 1:(N*N))
 #fix the value of beta1 & get N different beta2 vlues on fibers (V_beta) &
 #compute corresponding fuzzy atoms
{
b1<-soutput[,1][i] #determine beta1 for V_beta
b2<-soutput[,2][i] #determine beta2 for V_beta
 fb<-as.matrix(b1*(exp(-xs*b2))) #create fibers across M_n based on null model
 null_norm<-sqrt(sum((((y))-((fb)))^2))
#null_norm<-sqrt(sum(((as.vector(y))-(as.vector(fb)))^2))
#distance between the null and true yn at #beta
 snull[i,3]<-null_norm #store null model norm into snull
if (b2<1) {Z<-zb11} else
 {
 if (b2>=1 & b2<3)
 {Z_1<-zb11
  Z_2<-zb13
  x2<- (b2-1)/2
  #g2<-.25*(-1*(((2*x2)-1)^3)+3*((2*x2)-1))+.5
 g2<-(1/32)*(-5*(((2*x2)-1)^7)+21*((2*x2-1)^5)-35*((2*x2-1)^3)+35*((2*x2)-1))+.5
 #function interpolating the fuzzy cluster for all beta values
 #within the corresponding interval
   Z<-(1-g2)*Z_1+g2*Z_2} else
 {
```

```
 if (b2>=3 & b2<5)
 {Z_1<-zb13
  Z_2<-zb15
  x2<- (b2-3)/2
    g2<-(1/32)*(-5*(((2*x2)-1)^7)+21*((2*x2-1)^5)-35*((2*x2-1)^3)+35*((2*x2)-1))+.5
    Z<-(1-g2)*Z_1+g2*Z_2} else
 {
 if (b2>=5 & b2<7)
 {Z_1<-zb15
  Z_2<-zb17
  x2<- (b2-5)/2
    g2<-(1/32)*(-5*(((2*x2)-1)^7)+21*((2*x2-1)^5)-35*((2*x2-1)^3)+35*((2*x2)-1))+.5
    Z<-(1-g2)*Z_1+g2*Z_2} else
 {
 if (b2>=7 & b2<9)
 {Z_1<-zb17
                Z_2<-zb19
  x2<- (b2-7)/2
    g2<-(1/32)*(-5*(((2*x2)-1)^7)+21*((2*x2-1)^5)-35*((2*x2-1)^3)+35*((2*x2)-1))+.5
    Z<-(1-g2)*Z_1+g2*Z_2}
 else {Z<-zb19}
 }
 }
 }

PZ<-ppo(as.matrix(Z))-ppo(ppo(as.matrix(Z))%*%tm(b1,b2,xs))
```

```
#projection associated with the expectation surface and fuzzy
#clusters for beta values on V_beta
Zbasis<-basis(PZ) #basis for V_beta
 for (k in 3:(length(Zbasis)+2))
 {
 sZbasis[i,k]<-Zbasis[(k-2)]
#store this basis (1xn) row-wise into a matrix with each row
#beginning with the beta values the basis corresponds
 }
  Zbasisnorm<-Zbasis/sqrt(sum(Zbasis^2))
  trans<-as.matrix(Zbasisnorm%*%t(Zbasisnorm))
 #use of the normalized basis for the projection onto V_beta
test<-((y)-as.matrix(fb)) # shift the  y vector to origin
 gamma<- t(Zbasisnorm)%*%test #estimated gamma
soutput[i,4]<-gamma  #store gamma hat into soutput
 transc<-trans%*%test #the translation of V_beta and yn
 y_star<-fb+transc  #global call to create point on V_beta closest to y
 norm<-sqrt(sum((((y))-(as.vector(y_star)))^2))
#distance between y_star and the true yn at beta
 soutput[i,3]<-norm #store alternative model norm into soutput
}
 lrt(soutput,snull,sZbasis,y,xs)
 #call to find the LRT and hypothesis test
 #(this must be read before running Mn_star function)
}
}
```

```
 #for n=100#
 zb11<-read.table("zb11100.txt")
zb13<-read.table("zb13100.txt")
zb15<-read.table("zb15100.txt")
zb17<-read.table("zb17100.txt")
zb19<-read.table("zb19100.txt")

#Bb0<-read.table("Bb11100.txt")
 Gen_data<-Data(1,1,.2,xs,.05,1000)
Mn_star(100,xs,Gen_data,1000)

 #for n=100# for a single cluster
 zb11<-read.table("zb15100.txt")
zb13<-read.table("zb15100.txt")
zb15<-read.table("zb15100.txt")
zb17<-read.table("zb15100.txt")
zb19<-read.table("zb15100.txt")

#Note: Repeat "lrt" & "Mn_star" functions after changing file
#names to store output like "alt_result_sing.05", "null_result_sing.05,
#"TEST_sing.05", "lrt_sing.05" & "sd_sing.05"

 Mn_star(100,xs,Gen_data,1000)

 #Power with multiple maximin clusterings
count<-read.table("TEST.05.txt")
count<-ifelse((count==TRUE),1,0)
```

```
power<-sum(count)/length(count) #power of the test
 power
 summary(read.table("alt_result.05.txt"))
 #Power with single maximin cluster
 count<-read.table("TEST_sing.05.txt")
count<-ifelse((count==TRUE),1,0)
power<-sum(count)/length(count) #power of the test
 power
 summary(read.table("alt_result_sing.05.txt"))
  #Power with multiple testing
 lrt1<-as.matrix(read.table("lrt.05.txt"))
lrt2<-as.matrix(read.table("lrt_sing.05.txt"))
lrt<-rep(0,length(lrt1))
for (i in 1:length(lrt1)){
lrt[i]<-max(lrt1[i],lrt2[i])
lrt}
count<-ifelse(lrt>qchisq(.025,1,lower.tail=FALSE),1,0)
power<-sum(count)/length(count) #power of the test
 power
```