Model identification for dose response signal detection

Holger Dette, Stefanie Titoff,
Stanislav Volgushev
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de

Frank Bretz
Novartis Pharma AG
Lichtstrasse 35
4002 Basel, Switzerland
e-mail: frank.bretz@novartis.com

August 27, 2012

Model identification for dose response signal detection

Abstract

We consider the problem of detecting a dose response signal if several competing regression models are available to describe the dose response relationship. In particular, we re-analyze the MCP-Mod approach from Bretz et al. (2005), which has become a very popular tool for this problem in recent years. We propose an improvement based on likelihood ratio tests and prove that in linear models this approach is always at least as powerful as the MCP-Mod method. This result remains valid in nonlinear regression models with identifiable parameters. However, for many commonly used nonlinear dose response models the regression parameters are not identifiable and standard likelihood ratio test theory is not applicable. We thus derive the asymptotic distribution of likelihood ratio tests in regression models with a lack of identifiability and use this result to simulate the quantiles based on Gaussian processes. The new method is illustrated with a real data example and compared to the MCP-Mod procedure using theoretical investigations as well as simulations.

Keywords and Phrases: dose response studies; nonlinear regression; model identification; likelihood ratio test; contrast tests

1 Introduction

An important problem in the development of any new chemical or biological entity consists of characterizing well its dose response relationship (Ruberg, 1995a,b; Bretz et al., 2008). Statistical analysis methods of dose response studies can be roughly divided into (i) test procedures to detect a dose response signal (Stewart and Ruberg, 2000) or identify statistically significant dose levels (Tamhane et al., 1997, 2001) and (ii) regression methods to estimate the dose response curve. A common regression approach is to use a parametric dose response model that assumes a functional relationship between dose, treated as a continuous variable, and response (Morgan, 1992; Pinheiro et al., 2006b). If the parametric model is correctly specified, maximum likelihood or least squares estimation is an efficient method. However, misspecifying the parametric model can lead to substantial bias in estimating the dose response curve. This creates a dilemma in practice, because it is often required to pre-specify the parametric model, when the form of the dose response model in unknown. This is particularly true for the regulated environment in which drug development takes place, where the analysis methods (including the choice of the dose response model) have to be defined at the study design stage (ICH, 1994).

Many authors have investigated semi-parametric or non-parametric approaches to alleviate the model dependency problem and enhance the robustness of the dose response estimation; see Müller and Schmitt (1988); Kelly and Rice (1990); Mukhopadhyay (2000); Dette et al. (2005); Bornkamp and Ickstadt (2009); Dette and Scheder (2010); Yuan and Yin (2011) among many others. These methods allow model-independent descriptions of a dose response relationship. However, their applicability in dose response studies is limited because they require observations on a rather dense set of different dose levels, which are rarely available in practice. Due to logistic and ethical reasons, the number of different dose levels is typically in the order of 5 (Bornkamp et al., 2007) and therefore nonparametric methods do not yield reliable results.

Extending initial ideas proposed by Tukey et al. (1985) to address model uncertainty, Bretz et al. (2005) suggested a different strategy by choosing an appropriate model from a set of pre-specified candidate parametric dose response models; see Table 1 for a typical set of competing models used in practice. Their method, abbreviated as MCP-Mod, combines multiple comparison procedures with modeling techniques and consists of two steps. To begin with, it aims at detecting a dose response signal using multiple contrast tests. Conditional on a statistically significant outcome, it then selects a suitable dose response model to produce inference on adequate doses, employing a model-based approach. The MCP-Mod approach has become a very popular tool in recent years and has been subject to several extensions. Pinheiro et al. (2006a) discussed practical considerations regarding the implementation of

this methodology. Neal (2006) and Wakana et al. (2007) extended the original approach to Bayesian methods estimating or selecting the dose response curve from a sparse dose design. Klingenberg (2009) applied the MCP-Mod approach to proof-of-concept studies with binary responses. Benda (2010) proposed a time-dependent dose finding approach with repeated binary data. Akacha and Benda (2010) investigated the impact of dropouts on the analysis with recurrent event data. Several authors investigated extensions of the original MCP-Mod approach to response-adaptive designs; see Miller (2010); Bornkamp et al. (2011); Tanaka and Sampson (2012).

In this paper we re-analyze the MCP-Mod procedure and suggest an improvement. Our proposed approach is based on likelihood ratio tests and has at least two advantages. On the one hand, it makes better use of the available information than the original MCP-Mod procedure, because it uses the complete structure of the regression models. On the other hand, it does not require knowledge of the parameters of the competing models which are needed for the MCP-Mod procedure to construct the contrast tests for dose response signal detection. A particular challenge when using the likelihood ratio approach is that for commonly used regression models, such as those specified in Table 1, the resulting contrast tests correspond to the problem of testing for a constant regression. This leads to the problem of non-identifiability of some model parameters under the null hypothesis of no dose response. Consequently, standard asymptotic theory for likelihood ratio tests is not applicable here. In the context of independent and identically distributed observations such theory has been developed by Lindsay (1995) and Liu and Shao (2003). However, to the best knowledge of the authors, the asymptotic properties of the likelihood ratio test in the case of a lack of identifiability and independent but not identically distributed observations (in particular for regression models with fixed design) have not been investigated so far.

The remaining part of the paper is organized as follows. In Section 2 we revisit the original MCP-Mod approach and introduce the corresponding likelihood ratio tests. In Section 3 we compare both methods in the case of linear regression models. We show that if the number of different dose levels coincides with the number of parameters in the regression model, the MCP-Mod procedure is in fact a likelihood ratio test. More generally, we prove that in linear models the likelihood ratio test is always at least as powerful as the original MCP-Mod method and that the amount of improvement can be substantial. These results hold also (at least asymptotically) in nonlinear models where all parameters are identifiable under the null hypothesis of no dose response. However, the results of Song et al. (2009) indicate that this is not necessarily the case under non-identifiability of the parameters. Therefore, we develop in Section 4 a method to simulate the asymptotic distribution of the likelihood ratio test under lack of identifiability of the parameters for independent non-identically distributed data. In Section 5 we present some simulation results to illustrate the differences

model	$\eta(x,\theta)$	(I)	(II)	(III)
linear	$\vartheta_{1,1}+\vartheta_{1,2}x$	(0.2,0.6)	(0.1,0.3)	(0.2,0.6)
Emax	$\vartheta_{2,1}+\frac{\vartheta_{2,2}x}{\vartheta_{2,3}+x}$	(0.2,0.7,0.2)	(0.1,0.3,0.01)	(0.2,0.612,0.021)
exponential	$\vartheta_{3,1}+\vartheta_{3,2}\exp\left(x/\vartheta_{3,3}\right)$	(0.183,0.017,0.28)	(0.085,0.006,0.333)	(0.005,0.195,0.712)
log-linear	$\vartheta_{4,1}+\vartheta_{4,2}\log\left(x+\vartheta_{4,3}\right)$	(0.74,0.33,0.2)	(0.392,0.098,0.05)	(0.795,0.175,0.033)

Table 1: *Common parametric dose response models, with different parameter specifications (I) - (III).*

between the two methods. In particular, we demonstrate that in the considered examples the procedure based on likelihood ratio tests is never less powerful than the original MCP-Mod approach. In Section 6 we illustrate the new methodology with a real data set and provide some concluding remarks in Section 7. Finally, Section 8 contains some technical details justifying the results from Section 3 and 4.

2 Preliminaries

We consider the common nonlinear regression model

$$Y_{ij} = \eta(x_i,\theta) + \varepsilon_{ij}, \quad i = 1,\dots,N, \quad j = 1,\dots,n_i, \quad \sum_{i=1}^{N} n_i = n, \tag{2.1}$$

where η denotes the regression function, $x_1,\dots,x_N$ are different experimental conditions and n_i observations are taken at $x_i, i = 1,\dots,N$. In (2.1) the quantities ε_{ij} denote independently normally distributed random variables with mean 0 and variance $\sigma^2 > 0$. We assume $M \in \mathbb{N}$ candidate models

$$\eta_1(x,\theta_1),\dots,\eta_M(x,\theta_M) \tag{2.2}$$

to describe the regression, where $\theta_k \in \Theta_k \subset \mathbb{R}^{d_k}$ denotes a d_k-dimensional parameter in the kth model, $k = 1,\dots,M$. Throughout this paper the number of different experimental conditions is fixed, such that

$$\frac{n_i}{n} = \xi_i + o(1), \qquad i = 1,\dots,N, \tag{2.3}$$

where $n = \sum_{i=1}^{N} n_i \to \infty$ denotes the total sample size and $\xi_1,\dots,\xi_N$ positive weights with sum 1. The described situation is motivated by our interest in dose response studies, where patients are randomized to N dose levels and N is typically in the order of 4 or 5 due to logistic reasons. The assumption (2.3) is introduced for the asymptotic analysis of likelihood ratio tests in Section 4. We finally define ξ as the probability measure, which puts mass ξ_i at the point $x_i, i = 1,\dots,N$.

2.1 The MCP-Mod procedure revisited

As stated in the Introduction, Bretz et al. (2005) proposed the MCP-Mod approach to analyze dose response data under model uncertainty. First, the MCP-Mod approach investigates whether a given compound has a dose dependent effect. Second, it selects an appropriate model from the candidate model set (2.2) that most likely describes the underlying dose response curve.
Bretz et al. (2005) defined for each dose response model under consideration the vector

$$\mu_k = (\mu_{k,1}, \ldots, \mu_{k,N})^T, \tag{2.4}$$

where for $k = 1, \ldots, M$ and $i = 1, \ldots, N$ the quantity $\mu_{k,i} = \eta_k(x_i, \theta_k)$ denotes the expectation of Y_{ij} in model (2.1) if η_k is the "correct" model. That is, the vector μ_k describes the average effect of the compound at the experimental conditions $x_1, \ldots, x_N$, if the model η_k is the true one.
Bretz et al. (2005) proposed to test for each model the hypothesis

$$H_{0,k} : c_k^T \mu_k = 0 \quad \text{against} \quad H_{1,k} : c_k^T \mu_k > 0 \tag{2.5}$$

for a given contrast vector $c_k = (c_{k,1}, \ldots, c_{k,N})^T \in \mathbb{R}^N$ using the test statistic

$$K_{n,k} = \frac{c_k^T \overline{Y}}{\sqrt{\hat{\sigma}^2 \sum_{i=1}^N c_{k,i}^2 / n_i}}, \quad k = 1, \ldots, M. \tag{2.6}$$

Here, $\overline{Y}^T = (\overline{Y}_1, \ldots, \overline{Y}_N)$ denotes the vector of means $\overline{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$ at the dose levels x_i $(i = 1, \ldots, N)$ and $\hat{\sigma}^2 = \frac{1}{N-n} \sum_{i=1}^N \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i)^2$ is an estimator of the variance. Note that under the null hypothesis $H_{0,k} : c_k^T \mu_k = 0$ the statistic $K_{n,k}$ has a central t-distribution with $n - N$ degrees of freedom, while under the alternative it has a non-central t distribution with non-centrality parameter

$$\tau_k(c) = \frac{c_k^T \mu_k}{\sqrt{\sigma^2 \sum_{i=1}^N c_{k,i}^2 / n_i}} \, . \tag{2.7}$$

Bretz et al. (2005) determined for each candidate dose response model an optimal contrast c_k^* maximizing $\tau_k^2(c)$ in the class of all contrasts. These contrasts are then used in the test statistics (2.6). Note that the optimal contrasts depend on the particular model under consideration (in particular on the unknown model parameters) and the experimental conditions $x_1, \ldots, x_N$.

In order to conclude in favor of a statistically significant dose response signal, the M individual contrast test statistics $K_{n,k}$ are combined into a single decision rule. Bretz et al. (2005) suggested using the maximum statistic

$$K_{n,\max} = \max_{k=1}^{M} K_{n,k}, \tag{2.8}$$

where critical values are obtained from a multivariate t-distribution (Genz and Bretz, 2009). If statistical significance is achieved at this step, the MCP-Mod approach proceeds with selecting a suitable dose response model and fitting it to the data before estimating the target dose(s) of interest based on the fitted model. In this paper we mainly investigate the first step of the MCP-Mod procedure and propose a more powerful test to detect a dose response signal.

2.2 Likelihood ratio tests

A natural alternative to the contrast tests used by Bretz et al. (2005) is an approach based on likelihood ratio (LR) tests. In the following discussion let $\|\cdot\|_2$ denote the Euclidean norm. The LR test statistics for the hypotheses in (2.5) are given by

$$LR_{n,k} = -2\log\frac{T^k_{n,H_0}}{T^k_{n,H_1}}, \qquad k = 1,\ldots,M, \tag{2.9}$$

where

$$\begin{aligned} T^k_{n,H_0} &= \sup\Big\{\frac{1}{\sigma^n}\exp\Big(-\frac{1}{2\sigma^2}\,\|Y-\eta_k(\theta_k)\|_2^2\Big)\Big|\ \sigma>0;\ \theta_k\in\mathbb{R}^{d_k};\ c_k^T\mu_k = 0\Big\},\\ T^k_{n,H_1} &= \sup\Big\{\frac{1}{\sigma^n}\exp\Big(-\frac{1}{2\sigma^2}\,\|Y-\eta_k(\theta_k)\|_2^2\Big)\Big|\ \sigma>0;\ \theta_k\in\mathbb{R}^{d_k};\ c_k^T\mu_k \geq 0\Big\}. \end{aligned}$$

Here, $Y = (Y_{ij})_{i=1,\ldots,N}^{j=1,\ldots,n_i} \in \mathbb{R}^n$ and $\eta_k(\theta_k) = (\eta(x_{ij},\theta_k)_{i=1,\ldots,N}^{j=1,\ldots,n_i} \in \mathbb{R}^n$ denote the vectors of observations and expected responses at the different dose levels in the kth model, respectively. The LR test rejects the null hypothesis (2.5) of no dose response for large values of the statistic $LR_{n,k}$. The analogue of the statistic (2.8) is given by

$$LR_{n,\max} = \max_{k=1}^{M} LR_{n,k}, \tag{2.10}$$

where critical values have to be found by asymptotic theory in most cases of practical interest. In classical likelihood theory the statistic $LR_{n,k}$ usually converges weakly to a chi-squared type distribution, provided that the parameters characterizing the null distribution are unique [Wilks (1938) or Chernoff (1954)]. On the other hand, it is well known that classical

likelihood ratio theory does not apply to problems with a loss of identifiability [see Prakasa-Rao (1992), Lindsay (1995), Liu and Shao (2003) or Song et al. (2009) among others]. In the examples of Table 1 the problem of non-identifiability under the null hypothesis occurs naturally when testing the null hypothesis $H_{0,k} : \vartheta_{k,2} = 0$ in the kth model.
In the following sections we will compare the contrast test used in the original MCP-Mod method with the LR test proposed here. We begin the discussion with linear models for which the situation is most transparent. In this case the parameters are identifiable. The LR approach is always at least as good as the contrast test and the amount of improvement can be substantial. These results hold also in case of nonlinear regression models with identifiable parameters. For regression models with lack of identifiability the asymptotic distribution of the LR test has not been considered so far in the literature and will be presented in Section 4.

3 LR tests and MCP-Mod in linear regression models

For the sake of simplicity, we consider the test problem (2.5) for a single linear regression model

$$Y_{ij} = f^T(x_i)\theta + \varepsilon_{ij}, \qquad j = 1, \dots, n_i,\ i = 1, \dots, N, \tag{3.1}$$

with parameter vector $\theta \in \mathbb{R}^d$, where f is a given vector of regression functions. The results derived in this section can then be applied to each of the k test problems discussed in Section 2.
Because we consider only a single model, we use for now the notation $\mu^T = (\mu_1, \dots, \mu_N)$ and $c^T = (c_1, \dots, c_N)$ instead of μ_k and c_k. Let

$$X^T = (f(x_1), \dots, f(x_1), \dots, f(x_N), \dots, f(x_N)) \in \mathbb{R}^{d \times n} \tag{3.2}$$

denote the corresponding design matrix, where each vector $f(x_i)$ appears exactly n_i times in the matrix X^T and $n = \sum_{i=1}^N n_i$. We also assume that X^T has rank d, which means that there exist d linearly independent vectors $f(x_{i_1}), \dots, f(x_{i_d})$ among $f(x_1), \dots, f(x_N)$. It is easy to see that the vector μ in (2.4) can be represented in the form $\mu = AX\theta$, where

$$A^T = \operatorname{diag}\big(\frac{1}{n_1}1_{n_1}, \frac{1}{n_2}1_{n_2}, \dots, \frac{1}{n_N}1_{n_N}\big) \in \mathbb{R}^{n \times N}, \tag{3.3}$$

$1_k \in \mathbb{R}^k$ denotes the vector with all entries given by 1 and all other entries in the matrix A^T are 0. In linear models with normally distributed homoscedastic errors the LR test statistic

(2.9) for the hypotheses in (2.5) specializes to T_{n,H_0} and T_{n,H_1} are defined by

$$\begin{aligned} T_{n,H_0} &= \sup\Big\{\frac{1}{\sigma^n}\exp\Big(-\frac{1}{2\sigma^2}\ \|Y-X\theta\|_2^2\Big)\ \Big|\ \sigma>0;\ \theta\in\mathbb{R}^d;\ c^T\mu=0\Big\}, \\ T_{n,H_1} &= \sup\Big\{\frac{1}{\sigma^n}\exp\Big(-\frac{1}{2\sigma^2}\ \|Y-X\theta\|_2^2\Big)\ \Big|\ \sigma>0;\ \theta\in\mathbb{R}^d;\ c^T\mu\geq 0\Big\}. \end{aligned} \tag{3.4}$$

It now follows by straightforward calculation using Lagrangian multipliers that

$$ne^{-1}\cdot(T_{n,H_0})^{-2/n} = \|Y-X\hat\theta\|_2^2 + \frac{(\tilde c^T(X^TX)^{-1}X^TY)^2}{\tilde c^T(X^TX)^{-1}\tilde c},$$

where $Y=(Y_{11},\ldots,Y_{1n_1},\ldots,Y_{N1},\ldots,Y_{Nn_N})^T$ is the vector of all observations, $\tilde c = X^TA^Tc$ and $\hat\theta=(X^TX)^{-1}X^TY$ is the usual least squares estimate. Similarly, we have

$$ne^{-1}\cdot(T_{n,H_1})^{-2/n} = \|Y-X\hat\theta_{H_1}\|_2^2, \tag{3.5}$$

where $\hat\theta_{H_1}$ denotes the maximum likelihood estimate under the assumption $c^T\mu=\tilde c^T\theta\geq 0$; see the derivation in Section 8.1 of the Appendix. We thus have

$$\hat\theta_{H_1} = \begin{cases} \hat\theta &= (X^TX)^{-1}X^TY & \text{if}\quad \tilde c^T\hat\theta = c^TAX\hat\theta>0 \\ \hat\theta_{H_0} &= (X^TX)^{-1}X^TY-\hat\lambda(X^TX)^{-1}X^TA^Tc & \text{if}\quad \tilde c^T\hat\theta = c^TAX\hat\theta\leq 0 \end{cases},$$

where $\hat\theta_{H_0}$ denotes the maximum likelihood estimate under the null hypothesis $c^T\mu=0$ and

$$\hat\lambda = \frac{\tilde c^T(X^TX)^{-1}X^TY}{\tilde c^T(X^TX)^{-1}\tilde c} = \frac{c^TAX(X^TX)^{-1}X^TY}{c^TAX(X^TX)^{-1}X^TA^Tc}. \tag{3.6}$$

This gives for the LR test in (2.9), up to a monotone transformation,

$$L_n = \begin{cases} 0 & \text{if}\quad \tilde c^T\hat\theta\leq 0 \\ \dfrac{(\tilde c^T(X^TX)^{-1}X^TY)^2}{\tilde c^T(X^TX)^{-1}\tilde c\|Y-X\hat\theta\|_2^2} & \text{if}\quad \tilde c^T\hat\theta>0. \end{cases} \tag{3.7}$$

Consequently, the LR test rejects the null hypothesis (2.4) in the linear model (3.1) for large values of the statistic

$$\ell_n = \frac{\tilde c^T\hat\theta}{\|Y-X\hat\theta\|_2(\tilde c^T(X^TX)^{-1}\tilde c)^{1/2}}. \tag{3.8}$$

THEOREM 3.1 *For the linear regression model* (3.1) *with* $N=d$ *different experimental conditions, the LR test statistic* (3.8) *and the contrast test statistic* (2.6) *coincide up to a constant factor.*

Note that the assumption $d = N$ is crucial for Theorem 3.1. If the number of different experimental conditions is larger than the number of parameters in the linear model, the LR test can be more powerful than the contrast test. In many cases the improvement is substantial, as illustrated in the following example.

EXAMPLE 3.1 Consider the model $\eta(x, \theta) = ax$ and the case $N = 2$, where observations are taken at two different experimental conditions, say x_1, x_2, where $x_2 > x_1$. In this case, contrast coefficients are uniquely determined (up to the sign) and we have $c = (-1, 1)^T/\sqrt{2}$. The test statistic in (2.6) becomes

$$K_n = \frac{\overline{Y}_2 - \overline{Y}_1}{\sqrt{\hat{\sigma}^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

and follows a t-distribution with $n - 2$ degrees of freedom if $a = 0$. We reject the null hypothesis in (2.5) whenever $K_n > t_{n-2,1-\alpha}$, where $t_{n-2,1-\alpha}$ denotes the $(1 - \alpha)$-quantile of the t-distribution with $n - 2$ degrees of freedom. Moreover, let Φ denote the standard normal distribution function. Since $c^T\mu = a(x_2 - x_1)/\sqrt{2}$, the power of this contrast test is approximately given by

$$P_{c^T\mu>0}(K_n > t_{n-2,1-\alpha}) \approx \Phi\Big(\frac{a}{\sigma}(x_2 - x_1)\sqrt{\frac{n_1 n_2}{n_1 + n_2}} - u_{1-\alpha}\Big), \tag{3.9}$$

where $u_{1-\alpha}$ denotes the $(1 - \alpha)$-quantile of the standard normal distribution.
Consider now the LR test and note that $d = 1$, which implies $\tilde{c} \in \mathbb{R}$. Therefore, the LR test (3.8) rejects the null hypothesis in (2.5) whenever

$$L_n = \frac{n_1 x_1 \overline{Y}_1 + n_2 x_2 \overline{Y}_2}{\tilde{\sigma}(n_1 x_1^2 + n_2 x_2^2)^{1/2}} > t_{n-1,1-\alpha},$$

where the estimator of the variance is now given by

$$\tilde{\sigma}^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} \Big(Y_{ij} - \frac{n_1 x_1^{3-i} x_2^{i-1} \overline{Y}_1 + n_2 x_1^{2-i} x_2^i \overline{Y}_2}{n_1 x_1^2 + n_2 x_2^2}\Big)^2.$$

The power of this test is approximately given by

$$P_{c^T\mu>0}(L_n > t_{n-1,1-\alpha}) \approx \Phi\Big(\frac{a(n_1 x_1^2 + n_2 x_2^2)^{1/2}}{\sigma} - u_{1-\alpha}\Big). \tag{3.10}$$

Since $(x_2 - x_1)\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \leq \sqrt{n_1 x_1^2 + n_2 x_2^2}$ for all $x_2 > x_1$, it follows that the LR test is always more powerful than the contrast test. The power difference can be rather substantial. For example, if $x_1 = 1,\ x_2 = 2$, and $n_1 = n_2$, the power in (3.9) becomes $\Phi(\frac{a}{\sigma}\frac{\sqrt{n}}{\sqrt{2}} - u_{1-\alpha})$, while the corresponding term in (3.10) is $\Phi(\frac{a}{\sigma}\sqrt{10}\frac{\sqrt{n}}{\sqrt{2}} - u_{1-\alpha})$. Thus, in this example the LR test gives approximately the same power as the contrast test using only 10% of the sample size.

We now discuss a more general result showing the general superiority of the LR test in case of linear models. For a precise statement we assume that the variance σ^2 is known. A corresponding statement in cases where the variance has to be estimated holds asymptotically, as indicated in the previous example.
If the variance is known, the LR test rejects the null hypothesis in (2.5) if

$$\ell_n(\sigma^2) = \frac{\tilde{c}^T\hat{\theta}}{\sigma(\tilde{c}^T(X^TX)^{-1}\tilde{c})^{1/2}} > u_{1-\alpha}, \tag{3.11}$$

where $\tilde{c} = X^TA^Tc$ and the matrices X^T and A^T are defined in (3.2) and (3.3), respectively. The corresponding contrast test rejects whenever

$$k_n(\sigma^2) = \frac{c^T\overline{Y}}{\sigma\sqrt{\sum_{i=1}^N c_i^2/n_i}} > u_{1-\alpha}. \tag{3.12}$$

Our next results shows that in linear regression models with known variance the LR test is at least as powerful as the contrast test, regardless of the choice of the contrast vector. For an unknown variance these results hold at least if the sample size is sufficiently large (see Example 3.2 below).

THEOREM 3.2 *Consider the linear regression model* (3.1) *with known variance. Whenever* $c^T\mu \geq 0$ *we have*

$$P(\ell_n(\sigma^2) > u_{1-\alpha}) \;\geq\; P(k_n(\sigma^2) > u_{1-\alpha}). \tag{3.13}$$

Moreover, the inequality is strict if and only if $A^Tc \notin \text{range}(X)$*, where the matrix* A^T *is defined in* (3.3)*.*

EXAMPLE 3.2 Consider the linear regression model $\eta(x,\theta) = \vartheta_0 + \vartheta_1 x$ with equal sample sizes $n_1 = \ldots = n_N = m$. Up to the sign, the optimal contrast is given by

$$c^* = \Big(\frac{x_i - \overline{x}}{(\sum_{j=1}^N (x_j - \overline{x})^2)^{1/2}}\Big)_{i=1}^N, \tag{3.14}$$

where $\overline{x} = \frac{1}{N}\sum_{j=1}^N x_j$ (Bretz et al., 2005). Thus,

$$\begin{aligned} A^Tc^* &= \big(\sum_{j=1}^N (x_j - \overline{x})^2\big)^{-1/2}(1_m^T(x_1 - \overline{x}), \ldots, 1_m^T(x_N - \overline{x}))^T \\ &\in \text{range}(X) = \{a1_n + b\,(1_m^Tx_1, \ldots, 1_m^Tx_m)^T \mid a, b \in \mathbb{R}\}. \end{aligned}$$

Consequently, by Theorem 3.2 the LR test and the optimal contrast test have the same power. Note that this result holds also if the number of different experimental conditions

exceeds the dimension of the parameter. However, if a different contrast vector is used, the range inclusion is not necessarily satisfied and the LR rest is usually more powerful.
In the following we report the results of a small simulation study to illustrate these facts. We assume the models $\eta(x) = 0.2$ and $\eta(x) = 0.2 + 0.6x$ under the null and the alternative hypothesis, respectively. Further, we generated data with standard deviation $\sigma = 1.478$, while the sample size is $n = 200$ with $n_1 = \ldots = n_N$. All results are based on 5000 simulation runs. The simulated level and power of the LR and contrast tests (with estimated standard deviation) are shown in Table 2. The observed differences between the two methods are within the simulation error, as predicted by Theorem 3.1.

Design		LR test		contrast test	
x_1	x_2	level	power	level	power
0	0.2	0.0436	0.1436	0.0446	0.1426
0	0.5	0.0522	0.4214	0.0494	0.4096
0	1	0.0568	0.8852	0.0470	0.8810

Table 2: *Level and power of the LR and contrast tests for $N = 2$.*

Next, we consider $N = 4$ different experimental conditions and the three different contrasts

$$c^*, \quad c_1 = (-1, -1, 1, 1)^T, \quad \text{and} \quad c_2 = (-1, 0, 0, 1)^T, \tag{3.15}$$

where c^* denotes the optimal contrast defined in (3.14). Table 3 displays the simulation results. Note that for all three contrasts under consideration the hypotheses (2.5) are equivalent to $H_0 : \vartheta_1 = 0$ and $H_1 : \vartheta_1 > 0$ and therefore only one LR test is displayed in Table 3. In fact, one can show after tedious calculations that all vectors yield the same LR test statistic in (3.7). The LR test is more powerful than the contrast tests based on c_1 and c_2, as they do not satisfy the range inclusion in Theorem 3.2. For the optimal contrast test based on $c^\star$, however, we observe no power difference compared with the LR test. Note that this accordance is a specific coincidence of the linear regression model $\vartheta_0 + \vartheta_1 x$ used in this example.

Design				LR test		contrast test			
x_1	x_2	x_3	x_4	level	power	level	power		
							$c^\star$	c_1	c_2
0	0.05	0.1	0.2	0.0522	0.1100	0.0536	0.1078	0.0988	0.1080
0	0.1	0.4	0.5	0.0540	0.3288	0.0506	0.3206	0.2994	0.2666
0	0.25	0.75	1	0.0494	0.7342	0.0482	0.7414	0.6902	0.6464

Table 3: *Level and power of the LR and contrast tests for $N = 4$.*

In the case of nonlinear regression models the superiority of the LR test holds at least asymptotically, provided that all parameters of the model are identifiable. This follows by

the usual linearization arguments [see Seber and Wild (1989)]. In the case of a lack of identifiability, however, the results of Song et al. (2009) indicate that the superiority of the LR test is not granted. We will investigate this in more detail in Section 4. We complete this section with a nonlinear regression example satisfying the identifiability condition.

EXAMPLE 3.3 Consider the model

$$\eta(x, \theta) = \vartheta_0 - e^{-\vartheta_1 x}; \quad \vartheta_0 \in \mathbb{R},\ \vartheta_1 \in \mathbb{R}_0^+$$

with $N = 2$ different experimental conditions. In this case the contrast vector c is proportional to $(-1, 1)$ and we obtain for $x_1 < x_2$ $c^T\mu = c^{-\vartheta_1 x_1} - e^{-\vartheta_1 x_2}$. Therefore the hypothesis (2.5) is equivalent to $H_0 : \vartheta_1 = 0$ versus $H_1 : \vartheta_1 > 0$. Table 4 displays the simulated level and power of the LR and contrast tests, again for a sample size of $n = 200$, $n_1 = n_2$ and $\sigma = 1.478$. We observe that both tests have very similar power properties. Next we consider

Design		LR test		contrast test	
x_1	x_2	level	power	level	power
0	1	0.0502	0.2306	0.0472	0.2206
0	2	0.0502	0.4780	0.0484	0.4666
0	5	0.0482	0.9106	0.0484	0.9148

Table 4: *Level and power of the LR and contrast tests for $N = 2$.*

the case of $N = 4$ with the contrasts c_1 and c_2 given in (3.15) and the optimal contrast c^*. Table 5 displays the simulation results. We observe substantial power advantages for the LR for the two contrasts c_1, c_2 and a similar behaviour for the optimal contrast c^*.

Design				LR test		contrast test			
x_1	x_2	x_3	x_4	level	power	level	power		
							$c^\star$	c_1	c_2
0	0.25	0.75	1	0.0496	0.1720	0.0456	0.1638	0.1592	0.1398
0	0.5	1	2	0.0566	0.3154	0.0504	0.2974	0.2542	0.2876
0	1	3	5	0.0514	0.7540	0.0504	0.7524	0.6968	0.6900

Table 5: *Level and power of the LR and contrast tests for $N = 4$.*

4 LR tests under lack of identifiability

In this section we investigate the asymptotic distribution of the LR test for the hypotheses in (2.5) under more general dose response models, such as those presented in Table 1. To begin with, we demonstrate that under the null hypothesis of no dose response certain model parameters are not identifiable and standard LR test theory is not applicable. Nevertheless, we show in Section 8.2 of the Appendix that the quantiles of the limiting distribution can be

obtained using non-standard asymptotic theory for LR tests in regression models with a lack of identifiability. These results are used in Section 4.2 where we explain how the quantiles of the limiting process can be obtained by simulation.

4.1 The problem of identifiability

All models from Table 1 can be written as a nonlinear regression model of the form

$$\eta(x,\theta) = \vartheta_0 + \vartheta_1 \tilde{\eta}(x,\theta_{(2)}), \tag{4.1}$$

where $\theta = (\vartheta_0, \vartheta_1, \theta_{(2)}^T)^T \in \mathbb{R}^d$ and $\theta_{(2)} = (\vartheta_2, \ldots, \vartheta_{d-1})^T \in \mathbb{R}^{d-2}$. Assume without loss of generality that we are interested in an increasing trend $\vartheta_1 \geq 0$, i.e. $\tilde{\eta}(x_1,\theta_{(2)}) \leq \ldots \leq \tilde{\eta}(x_N,\theta_{(2)})$ for $x_1 < x_2 < ... < x_N$. Using the Lagranges multiplier device and letting $\overline{\mu} = \frac{1}{n}\sum_{i=1}^N n_i\mu_i$, one can show by similar arguments as in (Bornkamp, 2006, p. 88) that the solution of

$$c_\ell^* = n_\ell \frac{\mu_\ell - \overline{\mu}}{\sum_{i=1}^N c_i^* \mu_i}, \qquad \ell = 1, \ldots, N, \tag{4.2}$$

maximizes the non-centrality parameter $\tau_k^2(c)$ in (2.7) and is thus optimal. Note that $\tau_k^2(c)$ is invariant with respect to scalings of the vector c and that c^* satisfies $\sum_{\ell=1}^N \frac{(c_\ell^*)^2}{n_\ell} = 1$. Since $\mu_\ell = \eta_k(x_\ell,\theta)$, we have

$$c_\ell^* = n_\ell \frac{\eta_k(x_\ell,\theta) - \overline{\eta}}{\sum_{i=1}^N c_i^*\mu_i} = \vartheta_1 n_\ell \frac{\tilde{\eta}(x_\ell,\theta_{(2)}) - \tilde{\eta}}{\sum_{i=1}^N c_i^*\mu_i},$$

where $\overline{\eta} = \frac{1}{n}\sum_{\ell=1}^N n_\ell \eta(x_\ell,\theta)$ and $\tilde{\eta} = \frac{1}{n}\sum_{\ell=1}^N n_\ell \tilde{\eta}(x_\ell,\theta_{(2)})$. From the normalizing condition it finally follows that

$$c_\ell^* = \frac{\vartheta_1}{|\vartheta_1|} \frac{n_\ell(\tilde{\eta}(x_\ell,\theta_{(2)}) - \tilde{\eta})}{\left(\sum_{i=1}^N n_i(\tilde{\eta}(x_i,\theta_{(2)}) - \tilde{\eta})^2\right)^{1/2}}, \qquad \ell = 1, \ldots, N.$$

Since we assumed $\vartheta_1 \geq 0$, we obtain

$$c^{*T}\mu \;=\; \vartheta_1 \sum_{\ell=1}^N c_\ell^* \tilde{\eta}_\ell(x_\ell,\theta_{(2)}) = \vartheta_1 \cdot \Big(\sum_{i=1}^N n_\ell(\tilde{\eta}(x_\ell,\theta_{(2)}) - \tilde{\eta})^2\Big)^{1/2}.$$

Using the optimal contrast for the hypotheses in (2.5) under model (4.1) is thus equivalent to testing the hypotheses

$$H_0 : \vartheta_1 = 0 \qquad \text{versus} \qquad H_1 : \vartheta_1 > 0. \tag{4.3}$$

Therefore, the parameter $\theta_{(2)}$ is not identifiable under the null hypothesis $H_0 : \vartheta_1 = 0$ whenever $d > 2$. Nevertheless, the quantiles for the corresponding likelihood ratio test can

be obtained by non-standard asymptotic theory. Because these arguments are complicated, we defer the detailed discussion to Section 8.2 in the Appendix and explain in the following section how the quantiles of the asymptotic distribution of the likelihood ratio test can be calculated numerically by simulating non-standard Gaussian processes.

4.2 Simulating quantiles

We assume that the vector $\theta_{(2)}$ varies in some set, say Ψ, and define $Z_1, \ldots, Z_N, Y_1, \ldots, Y_N$ as independent identically distributed standard normal random variables. It is shown in Section 8.2 that the asymptotic distribution of the likelihood ratio test for the hypothesis (4.3) can be described by a functional of the stochastic process

$$\{\mathcal{W}_{\mathcal{S}} = \mathcal{W}(\beta_0, \beta_1, \sigma, \theta^{(2)}) \mid \beta_0^2 + \beta_1^2 + \sigma^2 = 1,\ \beta_1 \geq 0,\ \sigma > 0,\ \theta^{(2)} \in \Psi\},$$

where $\mathcal{S} = (\beta_0, \beta_1, \sigma, \theta^{(2)})$ and

$$\mathcal{W}_{\mathcal{S}} = \mathcal{W}(\beta_0, \beta_1, \sigma, \theta^{(2)}) \quad = \quad \frac{\sum_{i=1}^N \sqrt{\xi_i}((\beta_0 + \beta_1 \tilde{\eta}(x_i, \theta_{(2)}))Z_i + \sqrt{2}\sigma Y_i)}{(\sum_{i=1}^N \xi_i(\beta_0 + \beta_1 \tilde{\eta}(x_i, \theta_{(2)})^2 + 2\sigma^2))^{1/2}}. \tag{4.4}$$

To be precise, define

$$\begin{aligned} \mathcal{M}_2 &= \Big\{(\beta_0, \beta_1, \sigma, \theta_{(2)}) \in \mathbb{R}^{d+1} \,\Big|\, \beta_0^2 + \beta_1^2 + \sigma^2 = 1,\ \beta_1 \geq 0,\ \theta_{(2)} \in \Psi\Big\}, \\ \mathcal{M}_1 &= \Big\{(\beta_0, 0, \sigma, \theta_{(2)}) \in \mathbb{R}^{d+1} \,\Big|\, \beta_0^2 + \sigma^2 = 1\Big\}. \end{aligned}$$

It is shown in the Appendix that under the null hypothesis (4.3) the LR statistic converges weakly to the random variable

$$L = \sup_{\mathcal{S} \in \mathcal{M}_2} (\mathcal{W}_{\mathcal{S}} \vee 0)^2 - \sup_{\mathcal{S} \in \mathcal{M}_1} (\mathcal{W}_{\mathcal{S}} \vee 0)^2. \tag{4.5}$$

The quantiles of this limiting distribution can now be obtained by simulation.
So far, we focused on the specific case of testing a single model. In general, if M competing models are considered and the test statistic (2.10) is used, the corresponding quantile can be simulated in a similar way as described above. At each simulation step and for each model η_k under consideration a random variable L_k defined in (4.5) is simulated on the basis of the same data $Z_1, \ldots, Z_N, Y_1, \ldots, Y_N \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, resulting in the simulated random variable

$$L_{\max} = \max_{k=1}^{M} L_k. \tag{4.6}$$

The quantiles of the limiting distribution of the statistic (2.10) are then calculated by repeating this simulation step 10.000 times, say. Note that the resulting quantiles depend on

designs						statistic			
						$K_{n,\max}$			$L_{\max}$
						total sample size			
	x_1	x_2	x_3	x_4	x_5	$n = 50$	$n = 125$	$n = 250$	
A	0	0.05	0.2	0.6	1	1.91719	1.90372	1.88322	3.84865
B	0	0.05	0.1	0.2	0.5	1.87306	1.83116	1.81475	3.93500
C	0.5	0.7	0.9	0.95	1	1.79709	1.78592	1.77634	3.16702

Table 6: *Simulated* 95%*-quantiles of the statistic* $K_{n,\max}$ *defined in* (2.8) *and simulated* 95%*-quantiles of the random variable* $L_{\max}$ *defined in* (4.6)

the models and the design under consideration. Exemplarily we display in Table 6 these quantiles for the models from Table 1 and three uniform designs A, B, C with five different experimental conditions each. Table 6 also contains the quantiles of the statistic (2.8) for different values of n, based on the multivariate t-distribution with $n - 5$ degrees of freedom.

5 Simulation study

In this section we compare via a simulation study the original MCP-Mod from Bretz et al. (2005) using contrast tests with a modified version using the LR tests developed in this paper. We focus on the first step of the MCP-Mod approach, where the statistics (2.8) and (2.10) are used to test the null hypothesis in (2.5). We investigate the three designs A, B and C from Table 6. We use the four models in Table 1 as the candidate models for both procedures. These four models plus the constant model serve also as the data-generating "true" models in the simulations. The residual errors in (2.1) are normal distributed with standard deviation $\sigma = 1.478$. All results are based on 5.000 simulation runs.

In Table 7 we display the power for both tests and various sample sizes for each of the five data-generating regression models. The data were generated under the scenario (I) from Table 1. Note that the statistic (2.8) requires the specification of the optimal contrasts c_k^* and we used the "true" parameter values for their calculation. The LR test does not require this knowledge.

For all designs the nominal level is well approximated by the (asymptotic) LR test. For the contrast test, the significance level is maintained by construction also for finite sample sizes. In terms of power, we do not observe substantial differences between both procedures. Only if the true model is the Emax one, the LR test has slightly more power than the contrast test. Note that the power to detect the Emax model for design C is generally very small because of the choice of an inefficient design: observations are only taken at points larger than 0.5. In this region the derivative varies between 0.1 and 0.25 and the function is almost not distinguishable from the constant function.

Because the contrast test requires the knowledge of the model parameters for the calculation

			true regression function				
design	test	n	constant	Emax	log-linear	linear	exponential
A	contrast	50	0.0536	0.2444	0.2742	0.2772	0.2684
		125	0.0440	0.4718	0.4958	0.5034	0.4784
		250	0.0566	0.7280	0.7706	0.7692	0.7636
	LR	50	0.0546	0.2710	0.3084	0.2802	0.2666
		125	0.0548	0.4936	0.5186	0.5104	0.4820
		250	0.0586	0.7468	0.7672	0.7622	0.7396
B	contrast	50	0.0532	0.1840	0.1586	0.1294	0.0606
		125	0.0570	0.3450	0.2834	0.1862	0.0840
		250	0.0580	0.5508	0.4522	0.2994	0.0988
	LR	50	0.0520	0.1944	0.1628	0.1262	0.0670
		125	0.0488	0.3346	0.2662	0.1794	0.0780
		250	0.0534	0.5454	0.4224	0.2750	0.0872
C	contrast	50	0.0498	0.0650	0.0952	0.1336	0.2212
		125	0.0454	0.0744	0.1204	0.2028	0.4002
		250	0.0470	0.0828	0.1756	0.3206	0.6428
	LR	50	0.0602	0.0688	0.1038	0.1362	0.2472
		125	0.0474	0.0856	0.1304	0.2180	0.4146
		250	0.0546	0.0940	0.1714	0.3274	0.6422

Table 7: *Power of the MCP-Mod procedure based on the LR test* (2.10) *and the contrast test* (2.8) *at level* 5%. *Data was generated according to the constant model and the four regression models from scenario (I) in Table 1. The "true" parameters have been used for the calculation of the optimal contrasts in MCP-Mod.*

		true regression function				
design	n	constant	Emax	log-linear	linear	exponential
A	50	0.0524	0.2462	0.2396	0.2450	0.2236
	125	0.0482	0.4598	0.4744	0.4652	0.4286
	250	0.0478	0.7122	0.7520	0.7444	0.7136
B	50	0.0464	0.1786	0.1470	0.1116	0.0588
	125	0.0552	0.3234	0.2584	0.1770	0.0674
	250	0.0500	0.5348	0.4208	0.2594	0.0856
C	50	0.0520	0.0702	0.0918	0.1196	0.2156
	125	0.0526	0.0818	0.1236	0.2034	0.4110
	250	0.0468	0.0934	0.1840	0.3404	0.6534

Table 8: *Power of the MCP-Mod procedure based on the contrast test* (2.8) *at level* 5%. *data were generated according to the constant model and four "true" regression models from scenario (I) of Table 1. The parameters from scenario (II) have been used for the calculation of the optimal contrasts in MCP-Mod.*

of the optimal contrasts, we also investigate the properties of both tests if the contrasts are misspecified due to wrong assumptions about the model parameters. In our first example we generated data according to the same models as in Table 7, but with the contrasts being slightly misspecified based on the parameters from scenario (II) in Table 1. Table 8 displays the power results for the contrast test. The results for the LR test reported in Table 7 remain the same because the data were generated according to the same models in both tables. In almost all cases we observe a slight decrease in power for the contrast test as compared to the LR test.

Finally, we study a further case of misspecification, where we generated data according to scenario (III) in Table 1, while the optimal contrasts were calculated for the parameter constellations given in scenario (I). Table 9 displays these results. We observe no power differences between both procedures under design C. On the other hand, for designs A and B the LR test approach is more powerful in most cases under consideration, in particular for the Emax model.

			true regression function				
design	test	n	constant	Emax	log-linear	linear	exponential
A	contrast	50	0.0488	0.1958	0.2480	0.2688	0.2698
		125	0.0504	0.3674	0.4686	0.4954	0.5000
		250	0.0534	0.5990	0.7328	0.7758	0.7553
	LR	50	0.0574	0.2548	0.2788	0.2928	0.2846
		125	0.0614	0.4914	0.4932	0.4984	0.5070
		250	0.0590	0.7516	0.7414	0.7540	0.7552
B	contrast	50	0.0472	0.1762	0.1692	0.1202	0.0886
		125	0.0510	0.3448	0.3092	0.1962	0.1376
		250	0.0534	0.5630	0.5192	0.3062	0.1842
	LR	50	0.0496	0.2356	0.1932	0.1192	0.0864
		125	0.0516	0.4488	0.3264	0.1792	0.1188
		250	0.0544	0.6998	0.5176	0.2728	0.1706
C	contrast	50	0.0562	0.0512	0.0754	0.1272	0.1812
		125	0.0482	0.0506	0.0908	0.2094	0.2976
		250	0.0498	0.0546	0.1188	0.3202	0.4802
	LR	50	0.0572	0.0600	0.0900	0.1572	0.1860
		125	0.0554	0.0567	0.0944	0.2188	0.3088
		250	0.0478	0.0572	0.1159	0.3256	0.4882

Table 9: *Power of the MCP-Mod procedure based on the LR test* (2.10) *and the contrast test* (2.8) *at level* 5%. *Data were generated according to the constant model and the 4 regression models from scenario (III) in Table 1. The parameters from scenario (I) have been used for the calculation of the optimal contrasts in MCP-Mod.*

6 A real dose finding trial example

In this section we illustrate the proposed LR test with a real dose finding trial example. Biesheuvel and Hothorn (2002) investigated a dose ranging trial on a compound for the treatment of the irritable bowel syndrome. Patients were randomized to either placebo or one

of four active dose levels, corresponding to doses 0, 1,2,3, and 4. Note that the original dose levels have been blinded for confidentiality. The primary endpoint was a baseline adjusted abdominal pain score with larger values corresponding to a better treatment effect. In total 369 patients completed the study, with nearly balanced allocation across the doses. For the purpose of the calculations below, we ignore the gender information and investigated dose response for the complete data set. The data are available, for example, with the DoseFinding package from Bornkamp et al. (2010).

We consider three competing regression models of the form (4.1): linear, Emax, and exponential. Note that for the Emax and exponential models the parameter ϑ_2 is non-identifiable. When simulating the quantiles as described in Section 4.2, we apply a polar coordinate transformation to the vector of identifiable parameters $(\vartheta_0, \vartheta_1, \sigma)$ for numerical efficiency. For similar reasons, we restrict the search for the non-identifiable parameter ϑ_2 to the interval $[0, 6]$, because larger intervals lead to essentially the same results. The quantiles of the limiting distribution of the statistic (2.10) are then obtained by simulating 10.000 random variables of the form (4.6).

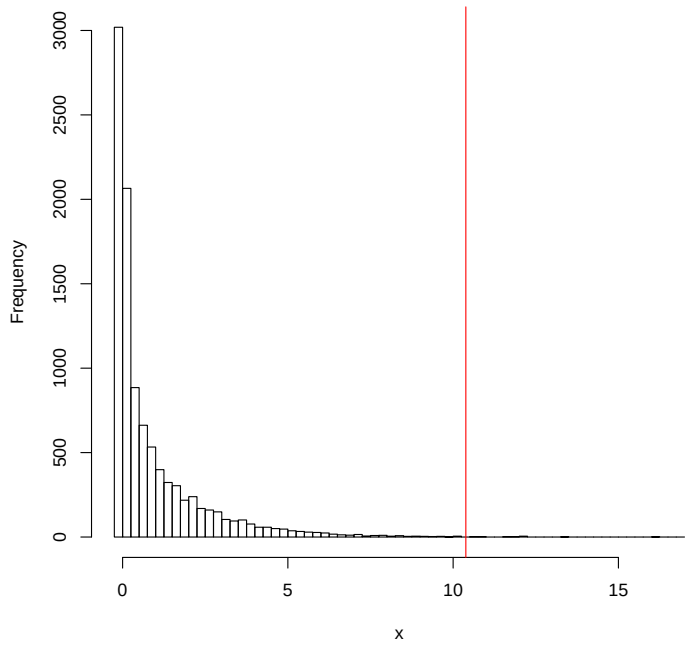

Figure 1: *Histogram of* 10.000 *simulated random variables* (4.6). *Vertical line: observed maximum LR test statistic*

Figure 1 displays the resulting histogram from which we immediately obtain the simulated quantiles. For this example, we obtain the simulated 90% (95%, 99%) quantiles as 2.78 (3.98,

6.86). The observed maximum LR test statistic (2.10) for the given data is 10.3844, which is displayed as a vertical line in Figure 1. Thus, we can safely reject the null hypothesis and conclude in favor of a significant dose response signal. We obtain the same test decision by computing the associated p-value, which is 0.0015 in this example

7 Conclusions

A common problem in modeling dose response relationships is the identification of an appropriate model to detect a dose response signal. In many cases there exist several competing parametric regression models to describe the dose response relationship. As indicated in the Introduction, the MCP-Mod approach has been advocated by several authors in the literature. This procedure combines multiple comparison procedures with modeling techniques but ignores the specific structure of the regression models.
In this paper we investigate an alternative procedure which is based on the likelihood ratio concept and uses all information from the models under investigation. We modify the first step of MCP-Mod for detecting a dose response signal by replacing the contrast tests through suitable LR tests. Unlike the original MCP-Mod procedure, which requires knowledge about the unknown parameters of the candidate regression models in order to specify the contrast coefficients, the new method does not require such knowledge. It is demonstrated that in linear models the LR test is always at least as powerful as the originally proposed contrast tests. These results can be transferred to nonlinear regression models, where all parameters are identifiable.
However, the commonly used nonlinear regression models for describing dose response relationships suffer from a lack of identifiability of the parameters, and as a consequence, standard asymptotic theory is not applicable for calculating quantiles of the likelihood ratio test. In order to solve this problem we derive the asymptotic distribution of the likelihood ratio test in regression models with independent but not identically distributed observations under lack of identifiability of the parameters. It turns out that the quantiles of the limiting distribution can be obtained by numerical simulation. The results are illustrated by means of a simulation study. In particular, we demonstrated that the likelihood ratio test is always comparable to the contrast tests although it does not require knowledge of any model parameters. Moreover, in many cases we also observe an improvement compared to the original contrast tests with respect to the power which can be substantial.

Acknowledgements The authors thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823,

Teilprojekt C1, C2) of the German Research Foundation (DFG).

References

Akacha, M. and Benda, N. (2010). The impact of dropouts on the analysis of dose-finding studies with recurrent event data. *Statistics in Medicine*, 29(15):1635–1646.

Benda, N. (2010). Model-based approaches for time-dependent dose finding with repeated binary data. *Statistics in Medicine*, 29(10):1096–1106.

Biesheuvel, E. and Hothorn, L. A. (2002). Many-to-one comparisons in stratified designs. *Biometrical Journal*, 44:101–116.

Bornkamp, B. (2006). Comparison of model-based and model-free approaches for the analysis of dose-response studies. Diplomarbeit, Fakultät Statistik, Technische Universität Dortmund, www.statistik.tu-dortmund.de/~bornkamp/diplom.pdf.

Bornkamp, B., Bretz, F., Dette, H., and Pinheiro, J. C. (2011). Response-adaptive dose-finding under model uncertainty. *Annals of Applied Statistics*, 5:1611–1631.

Bornkamp, B., Bretz, F., Dmitrienko, A., Enas, G., Gaydos, B., Hsu, C.-H., König, F., Krams, M., Liu, Q., Neuenschwander, B., Parke, T., Pinheiro, J. C., Roy, A., Sax, R., and Shen, F. (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials. *Journal of Biopharmaceutical Statistics*, 17:965–995.

Bornkamp, B. and Ickstadt, K. (2009). Bayesian nonparametric estimation of continuous monotone functions with applications to dose-response analysis. *Biometrics*, 65:198–205.

Bornkamp, B., Pinheiro, J., and Bretz, F. (2010). *DoseFinding: Planning and Analyzing Dose Finding experiments*. R package version 0.4-1.

Bretz, F., Hsu, J. C., Pinheiro, J. C., and Liu, Y. (2008). Dose finding - a challenge in statistics. *Biometrical Journal*, 50:480–504.

Bretz, F., Pinheiro, J., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrika*, 61:738–748.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics*, 25(3):573–578.

Dette, H., Neumeyer, N., and Pilz, K. F. (2005). Aa note on nonparametric estimation of the effective dose in quantal bioassay. *Journal of the American Statistical Association*, 100:503–510.

Dette, H. and Scheder, R. (2010). A finite sample comparison of nonparametric estimates of the effective dose in quantal bioassay. *Journal of Statistical Computation and Simulation*, 80:527–544.

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Springer.

ICH (E4 1994). *Topic E4: Dose-response information to support drug registration*.

Kelly, C. and Rice, J. (1990). Monotone smoothing with application to doseresponse curves and the assessment of synergism. *Biometrics*, 46:1071–1085.

Klingenberg, B. (2009). Proof of concept and dose estimation with binary responses under model uncertainty. *Statistics in Medicine*, 28(2):274–292.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. IMS, Hayward, California.

Liu, X. and Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Annals of Statistics*, 31(3):807–832.

Miller, F. (2010). Adaptive dose-finding: Proof of concept with type I error control. *Biometrical Journal*, 52(5):577–589.

Morgan, B. (1992). *Analysis of Quantal Response Data*. Chapman and Hall, New York.

Mukhopadhyay, S. (2000). Bayesian nonparametric inference on the dose level with specified response rate. *Biometrics*, 56:220–226.

Müller, H. and Schmitt, T. (1988). Kernel and probit estimates in quantal bioassay. *Journal of the American Statistical Association*, 83:750–759.

Neal, T. (2006). Hypothesis testing and Bayesian estimation using a sigmoid E(max) model applied to sparse dose-response designs. *Journal of Biopharmaceutical Statistics*, 16(5):657–677.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition.

Pinheiro, J. C., Bornkamp, B., and Bretz, F. (2006a). Design and analysis of dose finding studies combining multiple comparisons and modeling procedures. *Journal of Biopharmaceutical Statistics*, 16:639–656.

Pinheiro, J. C., Bretz, F., and Branson, M. (2006b). Analysis of dose-response studies – modeling approaches. In Ting, N., editor, *Dose Finding in Drug Development*, pages 146–171. Springer, New York.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer Series in Statistics. Springer-Verlag, New York.

Prakasa-Rao, B. L. S. (1992). *Identifiability in Stochastic Models: Characterization of Probability Distributions*. Academic Press, London.

Ruberg, S. J. (1995a). Dose response studies. I. Some design considerations. *Journal of Biopharmaceutical Statistics*, 5(1):1–14.

Ruberg, S. J. (1995b). Dose response studies. II. Analysis and interpretation. *JBS*, 5(1):15–42.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. John Wiley and Sons Inc., New York.

Song, R., Kosorok, M. R., and Fine, J. P. (2009). On asymptotically optimal tests under loss of identifiability in semiparametric models. *Annals of Statistics*, 37:2409–2444.

Stewart, W. H. and Ruberg, S. (2000). Detecting dose-response with contrasts. *Statistics in Medicine*, 19:913–921.

Tamhane, A. C., Hochberg, Y., and Dunnett, C. W. (1997). Multiple test procedures for dose finding. *Biometrics*, 52:21–37.

Tamhane, A. C., Hochberg, Y., and Dunnett, C. W. (2001). Multiple test procedures for identifying the maximum safe dose. *Journal of the American Statistical Association*, 96:835–843.

Tanaka, Y. and Sampson, A. (2012). An adaptive two-stage dose-response design method for establishing proof of concept. *Journal of Biopharmaceutical Statistics, in press*.

Tukey, J. W., Ciminera, J. L., and Heyse, J. F. (1985). Testing the statistical certainty of a response to increasing doses of a drug. *Biometrics*, 41:295–301.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Wakana, A., Yoshimura, I., and Hamada, C. (2007). A method for therapeutic dose selection in a phase II clinical trial using contrast statistics. *Statistics in Medicine*, 26(3):498–511.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9(1):60–62.

Yuan, Y. and Yin, G. (2011). Dose-response curve estimation: A semiparametric mixture approach. *Biometrics*, 67:1543–1554.

8 Appendix: technical details

8.1 Proof of the results in Section 3

Proof of (3.5) Let $\hat{\theta}_{H_1}, \hat{\sigma}^2_{H_1}$ denote the solution of the optimization problem defined in (3.4). By the Karush-Kuhn-Tucker conditions [see Nocedal and Wright (2006)] there exist a constant $\lambda \geq 0$, such that $\hat{\theta}_{H_1}, \hat{\sigma}^2_{H_1}$ are a solution of the system

$$\frac{n}{2}\,\frac{1}{\sigma^2} - \frac{1}{2\sigma^4}\,\|Y - X\theta\|^2 = 0, \qquad \frac{X^T(Y - X\theta)}{\sigma^2} + \lambda\,\tilde{c} = 0,$$

where $\tilde{c} = X^T A^T c$. If $c^T AX\hat{\theta} = \tilde{c}^T\hat{\theta} > 0$, $\lambda = 0$ and we obtain $\hat{\theta}_{H_1} = \hat{\theta}$, where $\hat{\theta} = (X^TX)^{-1}X^TY$ is the common least squares estimate. Otherwise we have $\lambda > 0$ and the solution is given by $\hat{\theta}_{H_1} = \hat{\theta}_{H_0}$ where $\hat{\theta}_{H_0} = (X^TX)^{-1}X^TY - \hat{\lambda}(X^TX)^{-1}X^TA^Tc$ and $\hat{\lambda}$ is defined by (3.6).

Proof of Theorem 3.1. Recall the representation of the matrix X^T in (3.2). We will show at the end of this proof that in the case $d = N$ the range of the matrix X is of the form

$$\text{range}(X) = \{(\lambda_1 1^T_{n_1}, \lambda_2 1^T_{n_2}, \ldots, \lambda_N 1^T_{n_N})^T \mid \lambda_1, \lambda_2, \ldots, \lambda_N \in \mathbb{R}\} \subset \mathbb{R}^n. \tag{8.1}$$

Consequently, the matrix corresponding to the projection of $\mathbb{R}^n$ onto range(X) is block diagonal, i.e.

$$P_X = X(X^TX)^{-1}X^T = \text{diag}\Big(\frac{1}{n_1}1_{n_1}1^T_{n_1}, \ldots, \frac{1}{n_N}1_{n_N}1^T_{n_N}\Big) \in \mathbb{R}^{n\times n}.$$

Recalling the definition of the matrix A in (3.3), this implies for the numerator in (3.8)

$$\begin{aligned} \tilde{c}^T\hat{\theta} &= c^TAX(X^TX)^{-1}X^TY = c^TAP_XY \\ &= c^TA(\overline{Y}_1, \ldots, \overline{Y}_1, \ldots, \overline{Y}_N, \ldots, \overline{Y}_N)^T = c^T\overline{Y} \end{aligned}$$

and for the denominator

$$\tilde{c}^T(X^TX)^{-1}\tilde{c} \;=\; c^TAX(X^TX)^{-1}X^TA^Tc = c^T\mathrm{diag}\Big(\frac{1}{n_1},\ldots,\frac{1}{n_N}\Big)c = \sum_{i=1}^N \frac{c_i^2}{n_i}.$$

Similarly, we have for the remaining term

$$\|Y - X\hat{\theta}\|_2^2 = Y^T(I_n - P_X)Y = \sum_{i=1}^N\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2$$

and the assertion of Theorem 3.1 follows.
The proof can be completed by showing (8.1). Let $\otimes$ denote the Kronecker product. From the definition of the matrix X in (3.2) we have

$$\begin{aligned}\mathrm{range}(X) &= \{(1_{n_1}^T \otimes f(x_1),\ldots,1_{n_N}^T \otimes f(x_N))^T\theta \mid \theta\in\mathbb{R}^d\}\\ &= \{(1_{n_1}^T f^T(x_1)\theta,\ldots,1_{n_N}^T f^T(x_N)\theta)^T \mid \theta\in\mathbb{R}^d\}\end{aligned}$$

and the inclusion "$\subset$" in (8.1) is obvious. For the converse inclusion observe that in the case $d = N$ $\{(f^T(x_1)\theta,\ldots,f^T(x_N)\theta)^T \mid \theta\in\mathbb{R}^d\} = \mathbb{R}^d$ by the linear independence of the vectors $f(x_1),\ldots,f(x_N)$.

Proof of Theorem 3.2. The power functions of the LR test (3.11) and the contrast test (3.12) are given by

$$\Phi\Big(\frac{c^T\mu}{\sigma(\tilde{c}(X^TX)^{-1}\tilde{c})^{1/2}} - u_{1-\alpha}\Big) \quad\text{and}\quad \Phi\Big(\frac{c^T\mu}{\sigma\sqrt{\sum_{i=1}^N c_i^2/n_i}} - u_{1-\alpha}\Big),$$

respectively. Let $P_X = X(X^TX)^{-1}X^T$ denote the projection matrix onto range(X) and note that all eigenvalues of P_X are given by 0 and 1. This yields

$$\tilde{c}^T(X^TX)^{-1}\tilde{c} = \|P_XA^Tc\|_2^2 \le \|A^Tc\|_2^2 = c^TAA^Tc = \sum_{i=1}^N\frac{c_i^2}{n_i}$$

and the inequality in (3.13) is now obvious. Moreover, the inequality is strict if and only if $A^Tc \notin \mathrm{range}(X)$, which proves the second assertion of Theorem 3.2.

8.2 Likelihood ratio tests under loss of identifiability.

In this section we present some arguments why the likelihood ratio test statistic converges weakly to the distribution of the random variable defined in (4.1), which justify the use of

its quantiles in the corresponding test. Our approach follows closely the paper of Liu and Shao (2003), but in contrast to these authors we consider regression models corresponding to the case of independent but not identically distributed data. Because most arguments are very similar to the reasoning given by these authors, we will only give a brief sketch of the derivation of the asymptotic properties of the likelihood ratio test. We use a slightly more general terminology which covers the nonlinear regression model considered in the previous sections as a special case.

Let $Y_1, \ldots, Y_n$ denote independent random variables depending on fixed covariates $x_1, \ldots, x_n$, respectively, with densities $f(y_i|x_i, \nu)$, where ν denotes a $(d+1)$-dimensional vector of parameters. We assume that the measure $\frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ converges weakly to a non-degenerate measure ξ and are interested in the hypothesis $H_0 : \nu \in \Xi_1$ versus $H_1 : \nu \in \Xi_2 \backslash \Xi_1$, where $\Xi_1 \subset \Xi_2 \subset \mathbb{R}^{d+1}$ denote parameter spaces. We allow that under the null hypothesis some parameters of the model are not identifiable. More precisely, if the null hypothesis is satisfied and $f_0(y|x)$ denotes the "true" density of the random variable Y (at experimental condition x), then there exist parameters $\nu \neq \tilde{\nu}$ in Ξ_1, such that

$$f_0(y|x) = f(y|x, \nu) = f(y|x, \tilde{\nu}).$$

For $i = 1, 2$ let $\Xi_{i0} = \{\nu \in \Xi_i \mid f(y|x, \nu) = f_0(y|x)\}$ denote the parameters in the set Ξ_i corresponding to the "true" density $f_0(y|x)$. We assume that for any $x \in \mathcal{X}$ the likelihood ratio $\ell_\nu(y|x) = \frac{f(y|x,\nu)}{f_0(y|x)}$ is square integrable with respect to $f_0(y|x)$. Following Liu and Shao (2003) we now define the quantities

$$\mathcal{S}_\nu(y|x) = \frac{\ell_\nu(y|x) - 1}{D(\nu)}, \tag{8.2}$$

where

$$D^2(\nu) = \int_{\mathcal{X}} (l_\nu(y|x) - 1)^2 f_0(y|x) dy d\xi(x). \tag{8.3}$$

For $\varepsilon > 0$ consider the sets $\Xi_{i\varepsilon} = \{\nu \in \Xi_i \mid 0 < D(\nu) < \varepsilon\}$ and define (for $i = 1, 2$) $\mathcal{F}_{i\varepsilon} = \{\mathcal{S}_\nu \mid \nu \in \Xi_{i\varepsilon}\}$, where the function $\mathcal{S}_\nu$ is defined in (8.2). Assume that for some $\varepsilon > 0$ the following conditions hold

(A1) The functions $\nu \mapsto D(\nu)$ and $\nu \mapsto \int_{\mathcal{X}} \int (\sqrt{l_\nu(y|x)} - 1)^2 f_0(y|x) dy d\xi(x)$ are bounded and continuous on Ξ_2.

(A2) $$\max_{1 \leq i \leq n} \sup_{\nu \in \Xi_{2\varepsilon}} |\mathcal{S}_\nu(Y_i|x_i)| = o_P(n^{1/2}).$$

(A3) $$\sup_{S \in \mathcal{F}_{2\varepsilon}} \Big| \frac{1}{n} \sum_{j=1}^n \mathcal{S}_\nu^2(Y_j|x_j) - \int_{\mathcal{X}} \int f(y|x) \mathcal{S}_\nu^2(y|x) dy d\xi(x) \Big| \xrightarrow{P} 0.$$

(A4) $$\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(\mathcal{S}_\nu(Y_j|x_j)-\int_{\mathcal{X}}\int f(y|x)\mathcal{S}_\nu(y|x)dyd\xi(x)\Big)\Big)_{S\in\mathcal{F}_{2\varepsilon}}\xrightarrow{w}(\mathcal{G}_\mathcal{S})_{S\in\mathcal{F}_{2\varepsilon}}$$

where $\xrightarrow{w}$ denotes weak convergence in the sense of Pollard (1984) and $\mathcal{G}$ denotes a separable, centered Gaussian process with covariance kernel

$$\mathrm{Cov}(\mathcal{G}_{\mathcal{S}_1},\mathcal{G}_{\mathcal{S}_2})=\int_{\mathcal{X}}\int \mathcal{S}_1(y|x)\mathcal{S}_2(y|x)f_0(y|x)\,dyd\xi(x).$$

Under these regularity conditions, it follows that with probability converging to 1 the maximum likelihood estimate satisfies $\hat{\nu}_n\in\Xi_{i\varepsilon}\cup\Xi_{i0}$ $(i=1,2)$ and, as a consequence, the the likelihood ratio test statistics

$$L_n=2\log\frac{\sup\Big\{\prod_{i=1}^{n}f(Y_i|x_i,\nu)\mid\nu\in\Xi_2\Big\}}{\sup\Big\{\prod_{i=1}^{n}f(Y_i|x_i,\nu)\mid\nu\in\Xi_1\Big\}},\ L_{n\varepsilon}=2\log\frac{\sup\Big\{\prod_{i=1}^{n}f(Y_i|x_i,\nu)\mid\nu\in\Xi_{2\varepsilon}\Big\}}{\sup\Big\{\prod_{i=1}^{n}f(Y_i|x_i,\nu)\mid\nu\in\Xi_{1\varepsilon}\Big\}} \tag{8.4}$$

have the same asymptotic properties. For $i=1,2$ define the set of functions

$$\mathcal{F}_i=\Big\{\mathcal{S}\in L^2:\exists\left\{\nu^{(m)}\right\}\in\Xi_{i\varepsilon}:\lim_{m\to\infty}D^2(\nu^{(m)})=0,\ \lim_{m\to\infty}||\mathcal{S}_{\nu^{(m)}}-\mathcal{S}||_\xi=0\Big\}, \tag{8.5}$$

where for a function $g(y|x)$

$$\|g\|_\xi=\Big(\int_{\mathcal{X}}\int g^2(Y|x)f_0(y|x)dyd\xi(x)\Big)^{1/2} \tag{8.6}$$

denotes the "norm" with respect to the design ξ. Assume additionally that

(A5) The classes of functions $\mathcal{F}_i$ $(i=1,2)$ satisfy the conditions in Definition 2.4 of Liu and Shao (2003).

It then follows by similar arguments as given in the proof of Theorem 3.3 of Liu and Shao (2003) that for $n\to\infty$ the likelihood ratio statistic $L_{n\varepsilon}$ in (8.4) converges weakly to the random variable

$$L=\sup_{\mathcal{S}\in\mathcal{F}_2}(\mathcal{G}_\mathcal{S}\vee 0)^2-\sup_{\mathcal{S}\in\mathcal{F}_1}(\mathcal{G}_\mathcal{S}\vee 0)^2, \tag{8.7}$$

where $\mathcal{G}_\mathcal{S}$ denotes the centered Gaussian from assumption (A4). Roughly speaking, the asymptotic distribution in (8.7) is derived by considering the two testing problems

$$H_{0i}:f(y|x,\nu)=f_0(y|x)\qquad\text{versus}\qquad H_{1i}:\ \nu\in\Xi_{i\varepsilon}$$

$(i=1,2)$. We would also like to point out that in general the sets $\mathcal{F}_1$ and $\mathcal{F}_2$ defined in (8.5), and with them the limiting distribution in (8.7), may depend on the unknown data-generating density f_0 which enters in the distance D and the norm $\|\cdot\|_\xi$ defined in (8.3) and (8.6), respectively. However, there exist many cases where this dependence does not appear in the limiting distribution (8.7).

REMARK 8.1 Consider the situation described in model (2.1), where there exists a fixed finite number of different design conditions $x_1, ..., x_N$. In this case the observations $Y_1, ..., Y_n$ can be regrouped in a finite collection of groups of i.i.d. data. Assumptions (A1) - (A4) can be further simplified. In particular, in this case (A4) implies (A3) by Lemma 2.10.4 in van der Vaart and Wellner (1996). The weak convergence in (A4) follows if the classes of functions $\{y \mapsto S_\nu(y|x_i) | \nu \in \Xi_{2\varepsilon}\}$, $i = 1, ..., N$ are Donsker with separable limiting processes (see e.g. van der Vaart and Wellner (1996)). Moreover, condition (A2) follows if for each experimental condition x_i the random variable $\sup_{\nu \in \Xi_{2\varepsilon}} |\mathcal{S}_\nu(Y_i|x_i)|$ is square integrable.

REMARK 8.2 For location scale regression models of the form (4.1) the situation simplifies further. More precisely, observe that in those models the parameter can be decomposed as $\nu = (\nu_1, \theta_{(2)})$ with $\nu_1 = (\vartheta_0, \vartheta_1, \sigma)^T$. In particular, the null hypothesis is equivalent to $\vartheta_1 = 0$, and in this case the value of $\theta_{(2)}$ has no impact on the likelihood. In order to describe the limits in (8.5), let ϑ_0^*, σ^* denote the "true" parameters and define $\nu_1^* = (\vartheta_0^*, 0, \sigma^*)^T$ (note that $(\sigma^*)^2$ is the "true" variance). By similar arguments as on page 826 in Liu and Shao (2003) it suffices to consider limits $(\nu_1^{(n)}, \theta_{(2)}) \to (\nu_1^*, \theta_{(2)})$. Because the errors in (2.1) are centered normal distributed we obtain

$$\frac{\partial \ell_\nu}{\partial \nu_1}\Big|_{\nu=(\nu_1^*,\theta_{(2)})} = \frac{1}{\sigma^*}\Big(z(y|x), z(y|x)\tilde{\eta}(x, \theta_{(2)}), z^2(y|x) - 1\Big)^T, \tag{8.8}$$

where $z(y|x) = \frac{y-\vartheta_0^*}{\sigma^*}$. Note that under the null hypothesis we have $z(Y|x) \sim \mathcal{N}(0,1)$, and the distribution of $\frac{\partial \ell_\nu}{\partial \nu_1}|_{\nu=(\nu_1^*,\theta_{(2)})}$ does not depend on the parameter ϑ_0^*. Therefore we obtain by a Taylor expansion $\ell_\nu - 1 = (\nu_1 - \nu_1^*)^T \frac{\partial \ell_\nu}{\partial \nu_1}\Big|_{\nu=(\nu_1^*,\theta_{(2)})} + o(|\nu_1 - \nu_1^*|)$. It now follows by similar arguments as in Liu and Shao (2003) that the sets $\mathcal{F}_1$ and $\mathcal{F}_2$ in (8.7) are given by

$$\mathcal{F}_i = \Big\{ \frac{\beta^T \frac{\partial \ell_\nu}{\partial \nu_1} |_{\nu=(\nu_1^*,\theta_{(2)})}}{\|\beta^T \frac{\partial \ell_\nu}{\partial \nu_1} |_{\nu=(\nu_1^*,\theta_{(2)})} \|_\xi} \;\Big|\; \beta \in \mathcal{B}_i,\ \theta_{(2)} \in \Psi_i \Big\} \tag{8.9}$$

$(i = 1, 2)$, where the sets $\mathcal{B}_1$ and $\mathcal{B}_2$ are defined by $\mathcal{B}_1 = \Big\{(\beta_0, 0, \sigma)^T \in \mathbb{R}^3 \mid \beta_0^2 + \sigma^2 = 1\Big\}$ and $\mathcal{B}_2\Big\{(\beta_0, \beta_1, \sigma)^T \in \mathbb{R}^3 \mid \beta_0^2 + \beta_1^2 + \sigma^2 = 1;\ \beta_1 \geq 0\Big\}$, respectively. Combining these results with (8.8) and (8.9) yields (note that the unknown parameter σ^* is canceling)

$$\begin{aligned}
\mathcal{F}_1 &= \Big\{ \frac{\beta_0 z(y|x) + \sigma(z^2(y|x) - 1)}{\|\beta_0 z + \sigma(z^2 - 1)\|_\xi} \;\Big|\; \beta_0^2 + \sigma^2 = 1 \Big\}, \\
\mathcal{F}_2 &= \Big\{ \frac{\beta_0 z(y|x) + \beta_1 z(y|x)\tilde{\eta}(x, \theta_{(2)}) + \sigma(z^2(y|x) - 1)}{\|\beta_0 z + \beta_1 z\tilde{\eta} + \sigma(z^2 - 1)\|_\xi} \;\Big|\; \beta_0^2 + \beta_1^2 + \sigma^2 = 1,\ \beta_1 \geq 0,\ \theta_{(2)} \in \Psi \Big\},
\end{aligned}$$

where we note the fact (recall the definition of (8.6)) $\|\beta_0 z + \sigma(z^2 - 1)\|_\xi^2 = \beta_0^2 + 2\sigma^2 = 1 + \sigma^2$ and $\|\beta_0 z + \beta_1 z\tilde{\eta} + \sigma(z^2 - 1)\|_\xi^2 \int_{\mathcal{X}} (\beta_0 + \beta_1 \tilde{\eta}(x, \theta_{(2)}))^2 d\xi(x) + 2\sigma^2$. Note that $\mathcal{F}_1 \subset \mathcal{F}_2$, and

thus it suffices to consider one Gaussian process $\{\mathcal{G}_{\mathcal{S}}\}_{\mathcal{S}\in\mathcal{F}_2}$ indexed by the functions if $\mathcal{F}_2$. Its covariance kernel is given by

$$\mathrm{Cov}(\mathcal{G}_{\mathcal{S}_1},\mathcal{G}_{\mathcal{S}_2}) \;=\; \frac{\int_{\mathcal{X}}\prod_{i=1}^{2}(\beta_0^{(i)}+\beta_1^{(i)}\tilde{\eta}(x,\theta_{(2)}))d\xi(x)+2\sigma^{(1)}\sigma^{(2)}}{\prod_{i=1}^{2}(\int_{\mathcal{X}}(\beta_0^{(i)}+\beta_1^{(i)}\tilde{\eta}(x,\theta_{(2)}))^2d\xi(x)+2(\sigma^{(i)})^2)^{1/2}} \quad \text{if} \quad \mathcal{S}_1,\mathcal{S}_2\in\mathcal{F}_2$$

where $(\beta_0^{(i)},\beta_1^{(i)},\sigma^{(i)})$ denote the parameters corresponding to the function $\mathcal{S}$. It is now easy to see that this process has the same distributional properties as the stochastic process $\{\mathcal{W}_{\mathcal{S}}\}_{\mathcal{S}\in\mathcal{M}_2}$ defined in (4.5).