Gössl, Küchenhoff:

Bayesian analysis of logistic regression with an unknown change point

Sonderforschungsbereich 386, Paper 148 (1999)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

Bayesian analysis of logistic regression with an unknown change point

Christoff Gössl, Helmut Küchenhoff
University of Munich, Institute of Statistics,
Akademiestraße 1, D–80799 München

Abstract

We discuss Bayesian estimation of a logistic regression model with an unknown threshold limiting value (TLV). In these models it is assumed that there is no effect of a covariate on the response under a certain unknown TLV. The estimation of these models with a focus on the TLV in a Bayesian context by Markov chain Monte Carlo (MCMC) methods is considered. We extend the model by accounting for measurement error in the covariate. The Bayesian solution is compared with the likelihood solution proposed by Küchenhoff and Carroll (1997) using a data set concerning the relationship between dust concentration in the working place and the occurrence of chronic bronchitis.

Keywords: threshold limiting value (TLV), segmented regression, measurement error, MCMC

1 Introduction

In toxicology, environmental and occupational epidemiology the assessment of threshold limiting values (TLVs) is an important task. In a dose-response relationship the TLV is the dose of the toxin or a substance under which there is no influence on the response. In many applications there is a controversy about the existence of such a TLV from a substantive point of view. In empirical studies evidence for the existence of a TLV and its estimation is often difficult, since distinguishing between no effect and a small effect can only be done by huge data sets. There are different models and methods for assessing a TLV, see e.g. Küchenhoff and Ulm (1997). In this paper we concentrate on a fully parametric logistic regression model proposed by Ulm (1991). In this model, which is a segmented regression model, the TLV is treated as an unknown parameter, which can be estimated assuming its existence. The interval estimates of the TLV give some evidence about its existence, since a TLV which is smaller than the smallest observed dose is equivalent to a non existing TLV. While the theoretical and practical problems in maximum likelihood estimation and the frequentist treatment of this model has been discussed by Küchenhoff and Wellisch (1997), we use a Bayesian approach. In this context no differentiability assumptions are necessary and it can be implemented with Markov chain Monte Carlo (MCMC) methods. We apply our methods to a study concerning the relationship between dust concentration in the working place and the occurrence of chronic bronchitis. In this study the exposure can only be measured with substantial measurement error. Therefore we also show how to incorporate this measurement error in our model. Since there are different approaches and possibilities concerning the MCMC algorithm and the assumption of the distribution of the regressor variable, we give a detailed discussion of the bronchitis example. The results are compared with those of a frequentist approach.

The paper is organized as follows. In Section 2 we present the model and a Bayesian solution of the problem of estimating the limiting value of a logistic threshold model. We propose a way to calculate the estimates by means of MCMC methods.
The modeling and the handling of measurement error in the dose covariate of our model is treated in Section 3.
In Section 4 we apply our methods to analyze in detail an occupational study regarding the assessing of a TLV for dust concentration in the working place. Further, our methods are compared with the different approaches as investigated by Küchenhoff and Carroll (1997).

2 A Bayesian Approach to the Logistic Threshold Model

In the following we focus our analysis on the logistic threshold model proposed by Ulm (1991):

$$P(Y = 1|X = x, Z = z) = G(z'\beta_{k-} + \beta_k(x - \tau)_+), \tag{1}$$

$$\text{where} \quad G(t) = (1 + \exp(-t))^{-1},$$

$$\beta \in \mathbb{R}^k, \quad \beta_{k-} = (\beta_1, \ldots, \beta_{k-1})' \text{ and } (x - \tau)_+ = \max(0, x - \tau).$$

Here, Y denotes the response variable, X is the dose variable, Z refers to further covariates. The unknown model parameters are β and the TLV τ. As can be seen from (1) there is no influence of X on Y, if X is smaller than τ, which exactly reflects the concept of a TLV.

In contrast to the classical frequentist inference, the parameters of a Bayesian model are not supposed to be fix but at random. For each of them exists a probability function, which reflects the prior knowledge of their value, the so-called priors. Now it is possible, according to the theorem of Bayes, to determine

in combination with the likelihood function of the data a so-called posterior of the parameters. This posterior distribution includes all knowledge relating to the parameters once from the prior and on the other hand from the likelihood.
The theorem of Bayes in its simplest form runs:

$$p(\theta|data) = \frac{p(\theta, data)}{p(data)}.$$

Here, *data* denotes the observed and θ the unknown parameters and latent variables. The numerator is the product of the likelihood and the priors. Note that in contrast to likelihood analysis, no further assumptions on $p(\theta, data)$ like differentiability in θ are needed for the analysis.
From this posterior the Bayesian point and interval estimates are derived. The median or the mean of the partial densities are, depending on the used loss-function, appropriate estimates for the parameters. In this paper, we use the mean of the posterior as point estimate and probability intervals, which can be regarded as Bayesian equivalents to the classical confidence-sets.

Thus, to derive the Bayesian posterior for our logistic threshold model, we have to determine the conditional likelihood function and the prior distributions.
The conditional likelihood of the i.i.d. sample (y_i, z_i, x_i) $i = 1, \ldots, n$ is according to (1) given by

$$[\mathbf{Y}|\mathbf{Z}, \mathbf{X}, \beta, \tau] = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \qquad (2)$$
$$\text{where} \qquad \pi_i = G(\beta_{k-} z_i + \beta_k (x_i - \tau)_+).$$

As usual "[]" refers to the density (or probability) of the corresponding random variables. We assume that the threshold is in the range of our observed data $\mathbf{X}$ and use a uniform prior in the range of the observed data for the threshold τ. For β a flat prior is assumed.

Now the posterior density of the parameters is

$$[\beta, \tau | \mathbf{Y}, \mathbf{Z}, \mathbf{X}] = \frac{[\mathbf{Y}|\mathbf{Z}, \mathbf{X}, \beta, \tau][\beta][\tau]}{\int [\mathbf{Y}|\mathbf{Z}, \mathbf{X}, \beta, \tau][\beta][\tau] d\mu(\beta, \tau)}. \tag{3}$$

Based on this density the above mentioned estimates are to be calculated. Although the determination of the numerator is easy, in most practical cases it is not possible to evaluate the denominator in an analytic way. For the threshold model, we solve this problem by means of MCMC methods. These methods allow to take a sample from a density only known up to a normalizing constant, which is in our particular problem the denominator. Then, on the basis of this sample of the posterior, the Bayesian estimates can simply be calculated, see e.g. Gilks, Richardson, and Spiegelhalter (1996).

For the logistic threshold model we use a two step Metropolis Hastings (MH) algorithm with multivariate random walk proposals in each step. We sample the parameter β_k and the threshold τ in one step and the parameter vector β_{k-} in step two. The full conditionals are straightforward. Since the densities cannot be analytically determined, it is not possible to apply the Gibbs–Sampler. We take two steps because of the strong dependence between the threshold τ and β_k. The covariance matrices of the proposals are tuned according to test runs to acceptance rates from 0.3 to 0.4. The starting values are chosen over-dispersed at random and the burn-in phase has to be determined by comparing several runs and then discarded. Due to high autocorrelation and slow convergence in the MH-output it is often necessary to thin out the simulated chain by taking only every k-th observation into the sample. We choose k such that the autocorrelation decreases to a sufficiently low level. The total extend of the runs depends on the convergence of the Markov chains and can be determined by comparing the point estimates of several runs. Figure (1) shows a trajectory of such a run and the belonging histogram with kernel density estimate for a simulated dataset. For results on real data we refer to Section 4.

Figure 1.

3 Errors in Variables

In many practical regression problems the regressors can only be measured with measurement error. Here, we want to propose a solution for incorporating measurement error of the variable X in our Bayesian model. A general introduction to the measurement error problem in threshold models is given by Küchenhoff and Carroll (1997), see also Carroll, Ruppert, Stefanski (1995).

We assume an additive measurement error model, i.e. instead of X the variable $W = X + U$ is observed, where U is the measurement error which is independent of Y, X, Z and is normally distributed with $E(U) = 0$ and $V(U) = \sigma_u^2$.

Now we split our model in three parts:

$$\begin{aligned} \text{main model} \qquad & [Y \mid X, Z, \zeta] \\ \text{error model} \qquad & [W \mid X, \eta] \\ \text{covariable model} \qquad & [X \mid Z, \lambda], \end{aligned}$$

where ζ, η and λ are the model parameters.
Using the independence assumption of U and (Y, W, X) the likelihood of the whole model can be written as

$$\begin{aligned} [\mathbf{Y}, \mathbf{W} | \mathbf{Z}, \theta] &= \prod_{i=1}^{n} \int [y_i | x, z_i, \zeta][w_i | x, \eta][x | z_i, \lambda] d\mu(x), \qquad (4) \\ \text{where} \qquad \theta &= (\zeta, \eta, \lambda). \end{aligned}$$

For our Bayesian approach the underlying model is the threshold model (1), as measurement error model we define $W \mid X \sim N(X, \sigma_u^2)$ and finally we assume X to be independent from Z and to have a normal distribution. This approach is extended in Section 4 to a finite mixture of normal distributions which is a flexible model for X with few parameters.

With respect to the priors, we propose as in Section 2 that no additional information is available and therefore define again noninformatives for β and τ. For the

parameters μ_x and σ_x^2 of the covariable model we assume, similar to the noninformatives, a normal distribution with mean 0 and a very large variance s^2 and a highly dispersed inverse-gamma distribution with parameters 1 and 0.005 so that its expectation equals infinity. The use of proper priors is here advisable as to avoid improper posteriors, see e. g. Besag et al. (1995). We further assume that the error variance σ_u^2 is known. For the posterior of the unknown parameters we get by suitable conditional independence assumptions

$$\left[\beta, \tau, \sigma_u^2, \mu_x, \sigma_x^2 | \mathbf{Y}, \mathbf{Z}, \mathbf{W}\right] \propto \left[\mathbf{Y}, \mathbf{W} | \mathbf{Z}, \beta, \tau, \sigma_u^2, \mu_x, \sigma_x^2\right] [\beta][\tau][\sigma_u^2][\mu_x, \sigma_x^2].$$

As in the previous section, it is not possible to derive the posterior analytically. Thus, we have again to apply the MH-algorithm. In addition we have to cope with the fact that the integral of formula (4) is in general not evaluable. In this paper, we want to solve the latter by means of MCMC methods, too. For another way of getting analytically the integral, which works with an approximation of the logit- by the probit-model, see Carroll, Ruppert, Stefanski (1995). For the foundations of the following method we refer to Richardson (1996). The idea is to add the unknown variable X to the parameters with a prior according to the covariable model and sample it from its full conditional. The partial densities of the other parameters will not be affected and the integration of formula (4) is implicitly carried out by the algorithm.

Now we describe the problem more formally. As mentioned above we assume X to be independent from Z and for simplicity distributed according to a single normal distribution $N(\mu_x, \sigma_x^2)$. We regard the latent variables X_i as parameters and decompose the likelihood as follows

$$[\mathbf{Y}, \mathbf{W} | \mathbf{Z}, \mathbf{X}, \beta, \tau, \sigma_u^2,] = [\mathbf{Y} | \mathbf{Z}, \mathbf{X}, \beta, \tau][\mathbf{W} | \mathbf{X}, \sigma_u^2]. \tag{5}$$

With the above priors and $X \sim N(\mu_x, \sigma_x^2)$, the posterior takes the form:

$$\begin{aligned} [\beta, \tau, \sigma_u^2, \mu_x, \sigma_x^2, \mathbf{X} | \mathbf{Y}, \mathbf{Z}, \mathbf{W}] &\propto [\mathbf{Y}, \mathbf{W} | \mathbf{Z}, \mathbf{X}, \beta, \tau, \sigma_u^2][\mathbf{X} | \mu_x, \sigma_x^2][\beta][\tau][\mu_x, \sigma_x^2], \\ &\propto [\mathbf{Y} | \mathbf{Z}, \mathbf{X}, \beta, \tau][\mathbf{W} | \mathbf{X}, \sigma_u^2][\mathbf{X} | \mu_x, \sigma_x^2][\beta][\tau][\mu_x, \sigma_x^2]. \end{aligned}$$

For the MH-algorithm, we use, as in Section 2, two Metropolis steps with random walk proposals for the parameters β and τ. Furthermore, we add a Metropolis-step for X. Thus, full conditionals are given by

$$\begin{aligned}
\left[\beta_{k-}|\mathbf{Y},\mathbf{Z},\mathbf{W},\mathbf{X},\beta_k,\tau,\sigma_u^2,\mu_x,\sigma_x^2\right] &\propto [\mathbf{Y}|\mathbf{Z},\mathbf{X},\beta,\tau], \\
\left[\beta_k,\tau|\mathbf{Y},\mathbf{Z},\mathbf{W},\mathbf{X},\beta_{k-},\sigma_u^2,\mu_x,\sigma_x^2\right] &\propto [\mathbf{Y}|\mathbf{Z},\mathbf{X},\beta,\tau][\tau], \\
\left[x_i|y_i,z_i,w_i,\beta,\tau,\sigma_u^2,\mu_x,\sigma_x^2\right] &\propto [y_i|z_i,x_i,\beta,\tau]\left[w_i|x_i,\sigma_u^2\right] \\
&\quad \left[x_i|\mu_x,\sigma_x^2\right], \quad i=1,\ldots,n.
\end{aligned}$$

Due to our choice of normal and inverse-gamma distributions, respectively, for the parameters μ_x and σ_x^2 we can derive their full conditionals as

$$\begin{aligned}
(\mu_x|\mathbf{X},\sigma_x^2) &\sim N\left(\frac{s^2\sum x_i}{s^2n+\sigma_x^2}, \frac{s^2\sigma_x^2}{s^2n+\sigma_x^2}\right), \\
(\sigma_x^2|\mathbf{X},\mu) &\sim IG\left(\frac{n}{2}+1, \frac{1}{2}\sum(x_i-\mu_x)^2+0.005\right).
\end{aligned}$$

Concerning the details of the implementation of the algorithm such as fixing the starting values we refer to Section 2. An application of the algorithm is reported in the following section.

4 Bronchitis Study

In several occupational studies conducted by the German research foundation (DFG) the relationship between average dust concentration in the working place and the occurrence of a chronic bronchitis reaction (Y) has been investigated. The disease was measured based on medical examinations like a questionnaire about symptoms, chest x-rays and lung function analysis.

Further covariates were smoking (SMK) and duration of exposure (DUR). We use the data of 1.256 Munich workers which were also analyzed by Küchenhoff and Carroll (1997). It should be mentioned that, we use the quantity

$X = \log(1 + \text{dust-concentration})$in our calculations.

$$P(Y = 1) = G(\beta_1 + \beta_2 \text{SMK} + \beta_3 \text{DUR} + \beta_4 (X - \tau)_+) \tag{6}$$

Because of the concentration measurements were gained by averaging over several single measurements scattered over the period and raw estimates for earlier periods, it seems to be appropriate not only to apply the simple threshold model but also the model of Section 3.

For the simple model without measurement error the algorithm of Section 2 can directly be used. Only the proposals of the Markov chain have to be modified in the described manner. Apart from that we derived our estimates from one chain with 50000 iterations, where we took due to the high autocorrelations every 50th observation in our sample. With respect to the burn-in, we discarded the first 500 iterations. Table 1 shows the estimates and those of Küchenhoff and Carroll (1997) gained by the classical methods. While the estimators are nearly identical, the estimators for the variance are higher for the classical methods. For example the estimated variance for the threshold τ was 0.41 compared to 0.28 in the Bayes analysis.

Table 1 .

In order to take into account the measurement error more complex modifications of the presented model have to be done. The first is, to regard the measured dust-concentration as the error-exposed surrogate W. The true concentration X is unknown.

Following Küchenhoff and Carroll (1997), we model the distribution of the unknown variable X by a mixture of two normal distributions. Assuming an additive measurement error $W = X + U$ where $U \sim N(0, \sigma_u^2)$ then W is also a mixture of normals with the same number of mixing distributions as X. So taking two mixing distributions is justified by the empirical distribution of W. Consequently,

for $\lambda \in [0;1]$ we assume X to be distributed according to

$$X \sim MixN(\mu_{x1}, \sigma^2_{x2}, \mu_{x2}, \sigma^2_{x2}, \lambda), \quad \text{with}$$
$$[x|\mu_{x1}, \sigma^2_{x2}, \mu_{x2}, \sigma^2_{x2}, \lambda] = \lambda\sigma^{-1}_{x1}\phi\left(\frac{x-\mu_{x1}}{\sigma_{x1}}\right) + (1-\lambda)\sigma^{-1}_{x2}\phi\left(\frac{x-\mu_{x2}}{\sigma_{x2}}\right).$$

The covariable model is given above, the underlying model, the logistic threshold model, is straightforward and we suppose an additive measurement error model, where $W|X = x$ has a normal distribution, i.e. $(W|X = x) \sim N(x, \sigma^2_u)$. So, the likelihood is complete and is given by (4).

Thus, the prior distributions are chosen like in the previous section, we only have to take into account that the covariable model has more parameters than in the last section. Again, we assume the case of no prior information. Accordingly we define the priors for the additional parameters μ_{x2}, σ^2_{x2} for which we use the same dispersed normal and gamma distribution as for μ_{x1}, σ^2_{x1}, and the prior for λ where we use a uniform distribution on $[0;1]$. Because the integral of the likelihood is not evaluable, we also have to consider the different prior of the unknown variable X according to our covariable model.
With respect to a practical solution, we assume the variance of the measurement error as known and take the value proposed in Küchenhoff and Carroll (1997), $\sigma^2_u = 0.187^2$. Because of problems with model identification other assumptions, like setting another prior distribution on σ^2_u did not work in our model.

Therefore the posterior is of the form:

$$[\beta, \tau, \sigma^2_u, \mu_{x1}, \mu_{x2}, \sigma^2_{x1}, \sigma^2_{x2}, \lambda, \mathbf{X}|\mathbf{Y}, \mathbf{Z}, \mathbf{W}] \propto$$
$$[\mathbf{Y}|\mathbf{Z}, \mathbf{X}, \beta, \tau][\mathbf{W}|\mathbf{X}, \sigma^2_u][\mathbf{X}|\mu_{x1}, \mu_{x2}, \sigma^2_{x1}, \sigma^2_{x2}, \lambda][\beta][\tau][\mu_{x1}][\mu_{x2}][\sigma^2_{x1}][\sigma^2_{x2}][\lambda].$$

As far as the derivation of the above formula is quite simple, the adapting of the MH-algorithm is a far more sophisticated task. The main problem results from the determination of the original distribution of the mixture for the variable X, which is necessary for defining the full conditionals of the according parameters.

Here, we use a method which is given in detail in Robert (1996) and works in principle by defining an indicator-variable m_i, which for all observations of X states from which distribution of the mixture it comes.

In general, suppose for X a mixture of m normal distributions is given, with parameters $\mu_1, \ldots, \mu_m, \sigma_1^2, \ldots, \sigma_m^2, \lambda_1, \ldots, \lambda_m$: $x_i \sim \sum_{j=1}^{m} \lambda_j N(\mu_j, \sigma_j^2)$. Then weights $g_{ij} = \lambda_j \sigma_j^{-1} \phi\left(\frac{x_i - \mu_j}{\sigma_j}\right)$ can be calculated that correspond to the contributions of the several mixing distributions to the density of x_i. Here $\phi(x)$ denotes the standard normal density. Thus, defining a discrete random variable M_i on the set of distributions $\{1, \ldots, m\}$ with the standardized weights $p_{ij} = \frac{g_{ij}}{\sum_j g_{ij}}$ as probabilities, yields a variable that allocates each element of the mixture a probability of having generated the observation x_i. Consequently, realizations of this distribution assign every observation one particular underlying distribution.

After having sampled an origin for every observation, the means, variances and mixing parameters of these distributions can be updated in familiar Gibbs- or MH-steps, according to the priors and the observations that come from this distribution. In our case, we use as above conjugate normal and inverse gamma priors for the means and variances and a non-informative prior for the mixing parameter so that two Gibbs- and one MH-step can be applied.

Thus, the full conditionals for our particular algorithm with the mixture of two normal distributions are:

$$
\begin{aligned}
&[\beta_{4-} | \mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{X}, \beta_4, \tau, \sigma_u^2, \mu_{x1}, \sigma_{x1}^2, \mu_{x2}, \sigma_{x2}^2, \lambda] \propto [\mathbf{Y} | \mathbf{Z}, \mathbf{X}, \beta, \tau], \\
&\left[\beta_4, \tau | \mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{X}, \beta_{4-}, \sigma_u^2, \mu_{x1}, \sigma_{x1}^2, \mu_{x2}, \sigma_{x2}^2, \lambda\right] \propto [\mathbf{Y} | \mathbf{Z}, \mathbf{X}, \beta, \tau]\,[\tau], \\
&\left[\lambda | \mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{X}, \beta, \tau, \sigma_u^2, \mu_{x1}, \sigma_{x1}^2, \mu_{x2}, \sigma_{x2}^2\right] \propto \left[\mathbf{X} | \mu_{x1}, \sigma_{x1}^2, \mu_{x2}, \sigma_{x2}^2, \lambda\right][\lambda], \\
&\qquad \left[x_i | y_i, z_i, w_i, \beta, \tau, \sigma_u^2, \mu_{x1}, \sigma_{x1}^2, \mu_{x2}, \sigma_{x2}^2, \lambda\right] \\
&\qquad\qquad \propto [y_i | z_i, x_i, \beta, \tau]\,[w_i | x_i, \sigma_u^2][\mathbf{X} | \mu_{x1}, \sigma_{x1}^2, \mu_{x2}, \sigma_{x2}^2, \lambda],
\end{aligned}
$$

for the MH-steps and for Gibbs-sampling we get

$$
\begin{aligned}
&(\mu_j|\mathbf{X}, \sigma_j^2) \sim N(\frac{10\sum x_i^j}{10n_j + \sigma_j^2}, \frac{10\sigma_j^2}{10n_j + \sigma_j^2}),\\
&(\sigma_j^2|\mathbf{X}, \mu_j) \sim IG(\frac{n_j}{2} + 1, -\frac{1}{2}\sum (x_i^j - \mu_j)^2 + 0.005), \quad j = 1, 2,\\
&(u_i|\mathbf{X}, \mu_{x1}, \mu_{x2}, \sigma_{x1}^2, \sigma_{x2}^2, \lambda) = \begin{cases} P(u_i = 1) &= p_{i1} \\ P(u_i = 2) &= 1 - p_{i1} \end{cases},
\end{aligned}
$$

where p_{i1} is defined as above, n_j denotes the number of observations x_i coming from distribution $u_i = j$ and x_i^j are their values themselves.

Again, we derived our estimates from one chain of the length of 50000 iterations where we took every 50th observation into our sample. In contrast to the simple model the autocorrelations then still had a value of about 0.6 to 0.7. But with regard to the computation time we accepted this high level and the resulting biases. The discarded burn-in extended again to 500 iterations. Table 2 shows the Bayes-estimates. For comparison the estimate for the threshold derived by Küchenhoff and Carroll (1997) under the assumption of a mixture of normal distributions for X, where $X \sim MixN(0.52, 0.144^2, 1.93, 0.106^2, 0.61)$, and a variance for the error model of $\sigma_u^2 = 0.187^2$, takes a value of 1.76.

Table 2 .

Further, the classical methods with the same assumptions give different estimates for the parameter τ depending on the method, see Küchenhoff and Carroll (1997). While the likelihood estimator is 1.76, the simex-estimator (Cook and Stefanksi, 1994) which does not use the information about the distribution of X gives a result of 1.40, which is close to our result. An interesting point is that variance estimation is higher for the Bayesian approach than it is for all classical methods where the s.e. varies from 0.12 to 0.23. A reason could be that in the Bayesian approach all sources of variability are modeled automatically, while in the classical approach this is not the case.

As mentioned above β_4 and τ are sampled in the same MH-step, because of their high correlation. The estimate for this correlation in the simulated dataset of Section 2 was 0.82, for the dust-dataset it was 0.71. In Figure 2 a scatter-plot and a two-dimensional kernel density estimate of the posterior's sample of β_4 and τ for the simple model in the bronchitis study also shows this high correlation.

Figure 2.

5 Discussion

We have shown that the Bayesian analysis for the complicated model of a segmented regression with errors in the regressors can be done by MCMC methods. We use a mixture of normals for the distribution of the regressor variable, which is a flexible parametric model. In this setting a classical analysis is very difficult, both theoretically and from a practical point of view. Another important argument for Bayesian analysis for finding threshold limiting values in epidemiology is the possibility of including knowledge from other studies or from substantive considerations by selecting a suitable prior distribution for the TLV.

A possible extension of our model would be to drop the assumption of an existing threshold, but to find out whether there is a threshold by data analysis. Another point is to treat the number of normals in the mixture distribution as a further parameter. For these problems methods proposed by Green (1995) have to be included into the MCMC-algorithm.

Acknowledgements

C. Gössl's research was supported by the SFB 386 form the German research foundation (DFG). We would like to thank Leonhard Knorr-Held for useful discussions and valuable comments.

References

[1] **Besag, J.E., Green, P.J., Higdon, D. & Mengersen, K. (1995).** *Bayesian computation and stochastic systems (with discussion).* Statistical Science, **10**, 3–66.

[2] **Carroll, R.J., Ruppert, D. & Stefanski, L.A. (1995):** *Nonlinear measurement error models.* Chapman & Hall, New York.

[3] **Cook, J.R. & Stefanski, L.A. (1994):** *Simulation extrapolation estimation in parametric measurement error models.* Journal of the American Statistical association, **89**, 1314–1328.

[4] **Gilks, W.R., Richardson, S. & Spiegelhalter, D.J. (1996):** *Markov Chain Monte Carlo in Practice.* Chapman & Hall, London.

[5] **Green, P. (1995)**: *Reversible Jump. MCMC computation and Bayesian model determination.* Biometrika, **82**, 711–732.

[6] **Küchenhoff, H., Carroll, R. J. (1997)** *Segmented Regression with errors in predictors: Semi-parametric and parametric methods.*, Statistics in Medicine, **16**, 169–188.

[7] **Küchenhoff, H., Ulm, K. (1997):** *Comparison of statistical methods for assessing threshold limiting values in Occupational Epidemiology.*, Computational Statistics **12:** 249–264.

[8] **Küchenhoff, H., Wellisch, U. (1997):** *Asymptotics for generalized linear segmented regression models with unknown breakpoint*, Discussion Paper Nr. 83, SFB 386, München.

[9] **Richardson, S. (1996):** *Measurement error.* in Gilks, Richardson, Spiegelhalter (eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 401–418.

[10] **Robert, C.P. (1996):** *Mixtures of distributions: inference and estimation*, in Gilks, Richardson, Spiegelhalter (eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, London, pp. 441–464.

[11] **Ulm, K. (1991):** *A statistical method for assessing a threshold in epidemiological studies*, Statistics in Medicine **10:** 341–349.

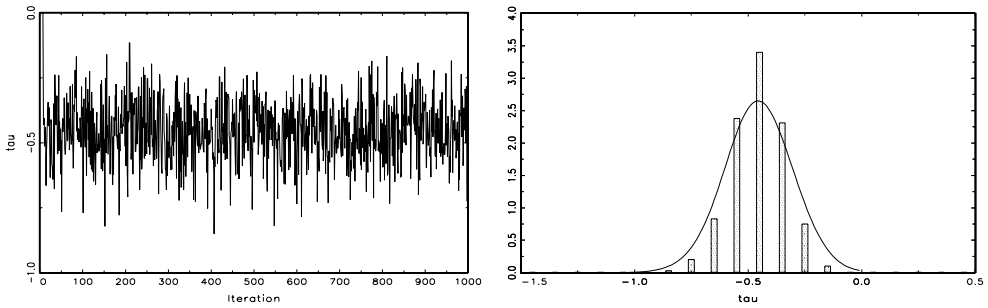

Figure 1: Trajectory and histogram with kernel density estimate of a MH-algorithm

Param.	ML-estim.	Bayes-estim.	Bayes-Var.
β_1	-3.00	-3.01	0.24
β_2	0.68	0.69	0.17
β_3	0.039	0.040	0.62
β_4	0.85	0.91	0.35
τ	1.27	1.27	0.28

Table 1: Likelihood- and Bayes-estimates of the simple model

Param.	β_1	β_2	β_3	β_4	τ	μ_{x1}	μ_{x2}	σ^2_{x1}	σ^2_{x2}	λ
Mean	-3.03	0.69	0.40	1.67	1.44	0.519	1.927	0.023	0.013	0.607
Var.	0.24	0.18	0.59	1.00	0.37	0.012	0.008	0.006	0.007	0.014

Table 2: Bayes estimates for the dust-dataset for the threshold model with errors in variables

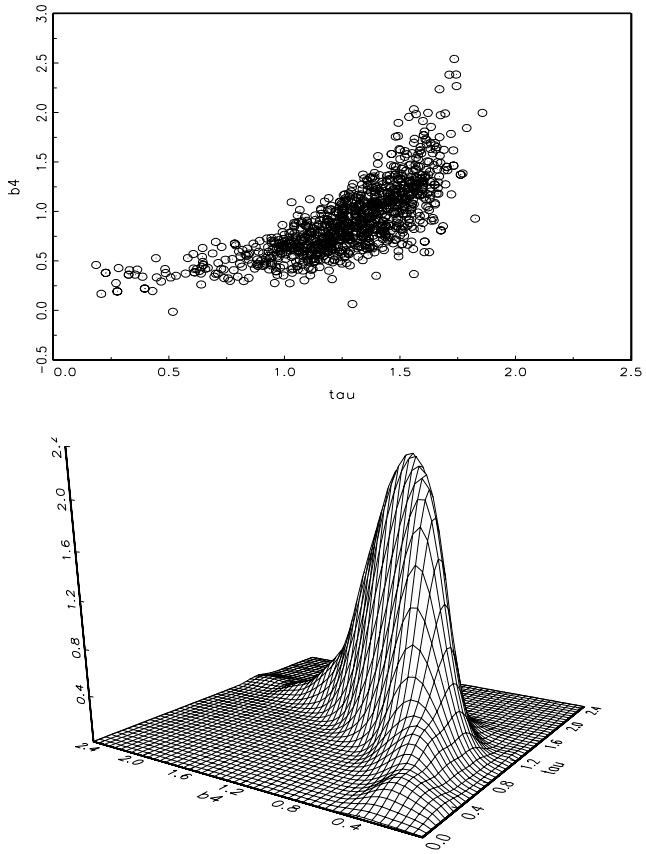

Figure 2: Scatter-Plot and two-dimensional kernel density estimation of the parameters β_4 and τ of the simple model of the Bronchitis study

Address for correspondence:

Christoff Gössl
Max-Planck-Institut für Psychiatrie
Kraepelinstr. 10
80804 München
Germany
Tel.: ++49 89 30 622 359
Fax: ++49 89 30 622 520
email: goessl@mpipsykl.mpg.de