Heteroscedastic Nonlinear Regression Models with Random Effects

Frank Gilberg and Wolfgang Urfer

Department of Statistics

University of Dortmund

Vogelpothsweg 87

D-44221 Dortmund, Germany

Abstract. We discuss an extension of the nonlinear random effects model from Lindstrom and Bates (1990) by adding a flexible transformation to both sides of the model (see Carroll and Ruppert (1988)) and describe a procedure for parameter estimation. This method combines pseudo maximum likelihood estimators for the transform-both-sides and weighting model and maximum likelihood (or restricted maximum likelihood) estimators for the linear mixed effects models. A validation of this new method is performed by analyzing a simulated set of enzyme kinetic data published by Jones (1993).

Key words: Nonlinear regression; Transformation and weighting; Pseudo likelihood estimation; Repeated measures data; Random effects; Enzyme kinetics

1. Introduction

In many nonlinear regression problems the classical assumptions about the underlying error structure are violated because the data contain nonconstant variances, outliers or a skewed error distribution. For detecting such deviations Carroll and Ruppert (1988) introduced a very flexible class of models which are called weighted transform-both-sides models. The flexibility of these models allows to integrate possible error structures in the model, each determined by additional parameters, which can be estimated from the data. For calculating estimators for the parameters it is assumed that weighting and transformation yields approximately independent and normal distributed errors. Then it is possible to derive likelihood functions and to calculate maximum likelihood estimators for the parameters.

In this report we will focus our discussion on nonlinear models with random effects which are very useful for analysing repeated measures data. Lindstrom and Bates (1990) showed the application of such models and introduced algorithms for parameter estimation but they did not involve heteroscedasticity of the error structure. This problem is described for generalized linear mixed models by Breslow and Clayton (1992) who presented very interesting examples for estimating general variance/covariance structures. Estimation of the regression and covariance parameters based on maximum likelihood methods by applying EM-algorithms.

In this report we combine the two strategies to a nonlinear weighted transform-both-sides model with random effects for repeated measures data. This is an extension of the published versions of mixed effects models from Lindstrom and Bates (1990) or Jones (1993) because a possible transformation is involved in the EM-algorithm and is updated every iteration step.

In section 2.1 we describe some features of the weighted transform-both-sides model. A procedure for estimating the parameters is the pseudo maximum likelihood method which is discussed in section 2.2. After an overview of random effects models in section 3.1 and maximum likelihood methods for parameter estimation in section 3.2 we present the new model and an EM-algorithm for parameter estimation in section 4. The application of the new model to a simulated set of enzyme kinetic data published by Jones (1993) is described in section 5 where we compare several possible models and discuss the advantages of our new method.

2. Fixed effects models

2.1 The transform-both-sides and weighting model

The nonlinear regression model

$$y_j = f(x_j, \beta) + \varepsilon_j \tag{1}$$

($j = 1, \ldots, n$), based on the assumptions that the expected value of the response y_j is equal to the specified nonlinear model function $f(x_j, \beta)$ with the predictor x_j and the unknown parameter β. The errors ε_j, $j = 1, \ldots, n$, are independent random variables, which have constant variances σ^2 and therefore come from the same distribution.

Usually β is estimated by least square methods, which yields optimal results if the errors are normally distributed. If there are deviations from the classical error distribution it is necessary to prove the validity of the assumptions so that the least squares procedure does not lead to biased parameter estimates.

For detecting such unknown violations Carroll and Ruppert (1988) introduced a very flexible class of regression models. They are called weighted transform-both-sides model and could be expressed as

$$y_j^{(\lambda)} = f(x_j, \beta)^{(\lambda)} + g_j(\theta)\varepsilon_j \tag{2}$$

(j = 1, . . . , n), where $g_j(\theta)$ is a general variance function which depends on the predictor and/or the response in a way that it is fixed with additional parameters. In this model the same transformation is applied to both sides of the model equation, so that it yields approximative constant variances and/or normally distributed errors. For this intention Carroll and Ruppert (1988) used the Box-Cox power transformation (see Box and Cox (1964)) which is defined by

$$h(z,\lambda) = z^{(\lambda)} = \begin{cases} (z^{\lambda}-1)/\lambda & \text{if } \lambda \neq 0 \\ \log(z) & \text{if } \lambda = 0 \end{cases}.$$

The estimation of the parameter vector (β, λ, θ) takes place with the maximum likelihood method under the assumption that transformation and weighting yields approximative normal distributed errors (see Section 2.2).

For enzyme kinetic data Ruppert, Cressie and Carroll (1989) illustrated the application of the weighted transform-both-sides model to the Michaelis-Menten equation. This function is used in many biological and biochemical situations to characterize the relationship between the velocitiy v_j of an enzyme catalyzed reaction and the corresponding concentration s_j of substrate for the j-th experiment (j = 1, . . . , n). The authors discussed in detail the model

$$v_j^{(\lambda)} = \left(\frac{V_{max} s_i}{K_m + s_j} \right)^{(\lambda)} + s_j^{\theta} \varepsilon_j$$

(j = 1, . . . , n) with the variance function $g_j(\theta) = s_j^{\theta}$. This power function takes into consideration a possible influence of the substrate concentration to the variability of the observed velocity. The model works very well when analyzing real data but it is necessary to make sure that too many variance parameters for finding the underlying error structure may bias the results.

2.2 The pseudo maximum likelihood method for parameter estimation

Now we describe the pseudo maximum likelihood method which yields estimators for all parameters simultanously. The procedure can be implemented easily with standard software routines for nonlinear optimization (see Giltinan and Ruppert (1989)). The method based on the definition of a special residual sum of squares which can be minimized with respect to all parameters and yields maximum likelihood estimators.

The derivation of the pseudo regression model is expressed under the assumption that in model (2) transformation and weighting yields approximative normal distributed errors with independent and homogeanous variances (Carroll and Ruppert (1988)). Then the log-likelihood function can be written without constant terms as

$$l(\lambda,\theta,\beta,\sigma) = \sum_{j=1}^{n} \left\{(\lambda-1)\log(y_j) - \log\left[\sigma g_j(\theta)\right]\right\} - \sum_{j=1}^{n} r_j^2(\beta,\lambda,\theta) \Big/ \left(2\sigma^2\right)$$

where $r_j(\beta,\lambda,\theta) = \left[y_j^{(\lambda)} - f\left(x_j,\beta\right)^{(\lambda)}\right] \Big/ g_j(\theta)$.

If we maximize the log-likelihood with respect to all parameters directly the estimate of the variance σ^2 may be negative. For this reason we replace the variance for given β, λ and θ by its maximum likelihood estimator

$$\hat{\sigma}^2 = n^{-1} \sum_{j=1}^{n} \left\{\left[y_j^{(\lambda)} - f\left(x_j,\beta\right)^{(\lambda)}\right] \Big/ g_j(\theta)\right\}^2$$

and changed to the concentrated log-likelihood

$$l(\lambda,\theta,\beta) = \sum_{j=1}^{n} \left\{(\lambda-1)\log(y_j) - \log\left[\hat{\sigma}(\lambda,\theta,\beta) g_j(\theta)\right]\right\} - n/2 .$$

After some additional algebraic calculation we get the relationship

$$l(\lambda,\beta,\theta) = -(n/2)\log\left(n^{-1} \sum_{j=1}^{n} \left[\dot{g}(\theta) r_j(\lambda,\beta,\theta) / \dot{y}^{\lambda-1}\right]^2 \right) - n/2$$

where $\dot{g}(\theta) = \prod_{j=1}^{n} \left(g_j(\theta)\right)^{1/n}$ and $\dot{y} = \prod_{j=1}^{n} \left(y_j\right)^{1/n}$ are geometric means.

This form of the log-likelihood and the calculation of the maximum likelihood estimators could be further simplified if we minimize the sum of squares

$$\sum_{j=1}^{n} \left[\dot{g}(\theta) r_j(\lambda,\beta,\theta) / \dot{y}^{\lambda-1}\right]^2$$

instead of maximizing the concentrated log-likelihood function.

This problem can be solved with standard nonlinear optimization routines (e.g. PROC NLIN from SAS) by using the 'pseudo-model' method where a dummy-variable D_j equal to zero replaces the response and a model

$$D_j = \left[\dot{g}(\theta)/\dot{y}^{\lambda-1}\right] r_j(\lambda, \beta, \theta)$$

can fit to the data. The main purpose of this method is that the residuals are the same as in model (2), but now the response does not depend on parameters.

3. Mixed effects models

3.1 Nonlinear models with random effects

A general nonlinear model for characterizing common covariance structures for repeatedly observed individuals was introduced by Lindstrom and Bates (1990). If y_i is a vector of observations, x_i the vector of the predictor and ε_i the vector of the stochastic error which have the same length n_i for each individual, i = 1, . . . , M, a model with random effects can be determined by the equation

$$y_i = f(\phi_i, x_i) + \varepsilon_i \,.$$

In this model ϕ_i is a parameter who vary from individual to individual and can be expressed as

$$\phi_i = A_i \beta + B_i b_i$$

where A_i and B_i are known design matrices with dimensions $n_i \times p$ and $n_i \times q$ and β is a $p \times 1$-vector of fixed effects. The $q \times 1$-vector b_i of random effects of the i-th individual and the error vector ε_i are independent normal distributed random variables with covariance matrices D_i and Λ_i. The distribution of the random effects is equal for all individuals so that the matrices D_i are identical for i = 1, . . . , M.

Now the response vector has for each individual the same distribution

$$y_i | b_i \sim N\left(f(\phi_i, x_i), \sigma^2 \Lambda_i\right)$$

where σ^2 describes the variance component of the total variance with respect to the error ε_i. The covariance matrix Λ_i depends on the individuals only through its dimension ($n_i \times n_i$) and is normaly set to an identity matrix.

If we write the M models for each individual together into one general model we get

$$y = \begin{bmatrix} y_1 \\ : \\ y_M \end{bmatrix}, \quad \phi = \begin{bmatrix} \phi_1 \\ : \\ \phi_M \end{bmatrix}, \quad f(\phi, x) = \begin{bmatrix} f(\phi_1, x_1) \\ : \\ f(\phi_M, x_M) \end{bmatrix}$$

$\tilde{D}$ = diag (D, D, . . . , D) and Λ = diag (Λ_1, . . . , Λ_M).

Now the common model is

$$y = f(\phi, x) + \varepsilon$$

where the parameter is $\phi = A\,\beta + B\,b$ with B = diag(B_1, . . . , B_M), $b = [b_1 \ldots b_M]^T$ and $A = [A_1 \ldots A_M]^T$ and the response vector is then conditionally distributed like

$$y|b \sim N(f(\phi, x), \sigma^2 \Lambda)$$

with $b \sim N(0, \sigma^2 \tilde{D})$ and $\varepsilon \sim N(0, \sigma^2 \Lambda)$.

In the general model with random effects we are interested in estimating the vector β of fixed effects and the vector b of random effects and we have to make inference about the variance components of the covariance matrices of the distributions of b and y. The estimation takes place with maximum likelihood methods.

3.2 Maximum likelihood estimation

For the derivation of likelihood functions it is necessary to linearize the model

$$y = f(A\beta + Bb, x) + \varepsilon$$

with respect to the fixed and random effects, which yields an approximate model

$$y = f(A\hat{\beta} + B\hat{b}, x) + \hat{Z}(b - \hat{b}) + \hat{X}(\beta - \hat{\beta}) + \varepsilon$$

with design matrices $\hat{Z} = \left. \frac{\partial f(\phi, x)}{\partial b^T} \right|_{\hat{\beta}, \hat{b}}$ and $\hat{X} = \left. \frac{\partial f(\phi, x)}{\partial \beta^T} \right|_{\hat{\beta}, \hat{b}}$.

By rearranging the linearized model we get the equation

$$w = \hat{X}\beta + \hat{Z}b + \varepsilon$$

where $w = y - f\left(A\hat{\beta} + B\hat{b}, x\right) + \hat{Z}\hat{b} + \hat{X}\hat{\beta}$.

Now it is possible to derive log-likelihood functions for maximum likelihood or restricted maximum likelihood estimation of the parameters under the assumption that the unconditional distribution of the response could be expressed as

$$y \sim N\left(f\left(A\hat{\beta} + B\hat{b}, x\right) - \hat{Z}\hat{b}, \sigma^2 V \right)$$

where $V = \Lambda + \hat{Z}\tilde{D}\hat{Z}^T$.

This procedure will be the basis of the development and evaluation of a new method to detect heterogenous error structures in nonlinear random effects models. The new approach will be discussed in the next section.

4. The transform-both-sides and weighting model with random effects

The nonlinear weighted transform-both-sides model and the nonlinear mixed effects model are for themselves very flexible and applicable tools for examining data with heterogenous variance-covariance structures. In this report we combine the two strategies to a nonlinear weighted transform-both-sides model with random effects for repeated measures data. This extension leads to a model

$$y_i^{(\lambda)} = f\left(x_i, \phi_i\right)^{(\lambda)} + g_i(\theta)\varepsilon_i \qquad (i = 1, \ldots, n)$$

The new aspect of this attempt is established by the fact that the transformation of the data is not fixed before performing the statistical analyses as proposed by Lindstrom and Bates (1990). In our procedure parameter estimation takes place with a modified EM-algorithm by updating the rule for the transformation at each iteration.

The new algorithm based on the iteration scheme of Lindstrom and Bates (1990) from section 3.2 and consists of the following steps:

1. In order to calculate initial estimates for the vector β of the fixed effects and the parameter λ for transformation and θ for weighting we used the weighted transform-both-sides model (2) from section 2.1.

2. For the estimation of the fixed and random effects and the variance components of the covariance matrice Λ of the random effects we approximate the nonlinear model by linearization with respect to fixed and random effects. The parameters for transformation and weighting are fixed in this step.

3. The estimated random effects from step 2 will be used as fixed parameters during the next iteration in the weighted transform-both-sides model. Step 2 and 3 will be reiterated until the residual sum of squares shows no identifiable difference.

The linearization of the model is similar to the method described in section 3.2.

Now we get the approximation

$$y_i^{(\lambda)} = f\left(A\hat{\beta} + B\hat{b}, x_i\right)^{(\lambda)} + \hat{Z}_i\left(b_i - \hat{b}_i\right) + \hat{X}_i\left(\beta - \hat{\beta}\right) + g_i(\theta)\varepsilon_i$$

with $\hat{Z}_i = \left.\frac{\partial f_i^{(\lambda)}}{\partial b_i^T}\right|_{\hat{\beta},\hat{b}}$ and $\hat{X}_i = \left.\frac{\partial f_i^{(\lambda)}}{\partial \beta^T}\right|_{\hat{\beta},\hat{b}}$.

The conversion of this equation leads to the linear random effects model

$$w_i = \hat{X}_i\beta + \hat{Z}_i b_i + g_i(\theta)\varepsilon_i$$

with the calculated response $w_i = y_i^{(\lambda)} - f\left(A\hat{\beta} + B\hat{b}, x_i\right)^{(\lambda)} + \hat{Z}_i\hat{b}_i + \hat{X}_i\hat{\beta}$ which will be used in the second step of the iteration for maximum likelihood or restricted maximum likelihood estimation.

5. Example

For a validation of the new model and the modified algorithm for parameter estimation we used a simulated data set from Jones (1993). He analyzed the Michaelis-Menten enzyme kinetic function

$$v = \frac{V_{max} s}{K_m + s} \tag{3}$$

where v is the reaction velocity, V_{max} the maximum velocity of the reaction, s the concentration of substrate and K_m the half-saturation constant. The vital role of this kinetic function in the model-based inference of toxicokinetics is showed by Urfer and Becka (1996).

Because the parameters V_{max} and K_m in the model must be positive he used the variable transformation

$$\beta_1 = \ln(V_{max}) \text{ and } \beta_2 = \ln(K_m)$$

for making sure that the optimization does not lead to negative estimators if the starting values are not good enough. This parameterization leads to a random effects version of model (3)

$$v_{ij} = \frac{s_{ij}\exp(\beta_1 + \gamma_{i1})}{s_{ij} + \exp(\beta_1 + \gamma_{i2})} + \varepsilon_{ij} \tag{4}$$

where the subscipt i = 1, . . . , M represents the individual experiment and j = 1, . . . , n_i are the observations for individual i.

For the simulation Jones used the following assumptions:

The values for the fixed effects were $\beta_1 = 0$ and $\beta_2 = 0$ (or $V_{max} = 1$ and $K_m = 1$) and the distribution of the random effects was the bivariate normal

$$\begin{bmatrix} \gamma_{i1} \\ \gamma_{i2} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.04 & 0.02 \\ 0.02 & 0.04 \end{bmatrix}\right)$$

with a correlation 0.5.

For the error term of the simulated observations he used the constant coefficient of variation variance model with the variance function $g(\theta) = 0.1 f(x, \beta)$ which implied a value of 0.01 for the overall variance σ^2. The whole procedure of the simulation is fully described by Jones (1993). The resulting observations and random effects were shown in the two tables on pages 148 and 149.

As Jones pointed out the distribution of the response y in model (6) is no longer Gaussian because of the reparameterization with respect to the fixed effects and the definition of the random effects. Nevertheless he assumed an approximate normal distribution for the response because in a real situation the distribution is unknown.

For the verification of this statement we will have a look at the empirical distributions of the simulated random effects. For the values of the actual simulated random effects we yield the empirical distribution

$$\begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} \sim N\left(\begin{bmatrix} 0.0037 \\ 0.0133 \end{bmatrix}, \begin{bmatrix} 0.0504 & 0.0173 \\ 0.0173 & 0.0346 \end{bmatrix}\right)$$

with a correlation coefficient r = 0.4145.

A comparison with the real distribution which was used by Jones to generate the random effects shows an overestimation of the expectation of the random effects. The covariance matrix showed also a bias in such a way that the variances of γ_{i1} and γ_{i2} were 25% higher respectivly 13% lower as in the underlying distribution. The covariance and the correlation were biased downwards by 13% and nearly 20%.

For the values of the simulated random effects after linearization of model (4) we yield the empirical distribution

$$\begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} \sim N\left(\begin{bmatrix} 0.0118 \\ 0.0302 \end{bmatrix}, \begin{bmatrix} 0.0516 & 0.0251 \\ 0.0251 & 0.0349 \end{bmatrix} \right)$$

with a correlation coefficient r = 0.5962.

The comparison with the real distribution of the random effects shows the same trend as the previous empirical distribution with reparameterization of the fixed and random effects. But now the bias for the expectation is three times higher as before. The bias for the variances of γ_{i1} and γ_{i2} are nearly the same as before but the covariance and the correlation are now biased upwards by nearly 25% and 20%.

This comparison shows that the assumption of Jones about the approximative equality of the two distributions is not fullfilled completely. But the question is now whether a statistical procedure is able to yield a comprehensive estimate of the covariance matrix D of the random effects with the variance components d_{11}, d_{12} and d_{22}. For getting an answer to this problem we examined several models with flexible variance/covariance structures.

As a first attempt we carried out an ordinary least squares fit of the parameters with the unweighted model

$$y_{ij} = f(\phi_i, x_{ij}) + \varepsilon_{ij}.$$

Model 1: Ordinary least squares (λ=1, θ=0)

$\hat{\sigma}^2 = 0.00296$	$\hat{d}_{11} = 0.05483$	$\hat{d}_{12} = 0.05053$	$\hat{d}_{22} = 0.05180$
	Lower 95% CL	Estimate ± SE	Upper 95% CL
β_1	-0.116	-0.031 ± 0.0436	0.054
β_2	-0.152	-0.050 ± 0.0520	0.052

In this case the estimates for V_{max} and Var(γ_{ij}) were relative near to the expected values from the simulation. This fact was not surprising because there was a high number of experiments and many observations for high substrate concentrations made it possible to fit the curves in this area very well. On the other side the estimated values for K_m and Var(γ_{i2}) were more biased because of the absence of an appropriate modeling of the variance structure in this model. The overall variance component σ^2 was clearly underestimated .

Then we applied the true model of the simulation

$$y_{ij} = f(\phi_i, x_{ij}) + f(\phi_i, x_{ij})\varepsilon_{ij}$$

by using the correct variance function for weighting.

Model 2: weighting with correct variance function (Jones (1993))

$\hat{\sigma}^2 = 0.00911$	$\hat{d}_{11} = 0.05696$	$\hat{d}_{12} = 0.03024$	$\hat{d}_{22} = 0.04023$
	Lower 95% CL	Estimate ± SE	Upper 95% CL
β_1	-0.111	-0.023 ± 0.0450	0.062
β_2	-0.093	-0.012 ± 0.0412	0.069

It was very surprising that the estimation for all the interesting parameters was not very well because all the variance components of Λ were all slightly overestimated. The reason for this fact may be the bias in the simulation procedure so that this weighted model is no more the correct one.

A third attempt of fitting the data based on transformation. We choose the log-transformation and applied it to both sides of the model which yields the relationship

$$\log(y_{ij}) = \log(f(\phi_i, x_{ij})) + \varepsilon_{ij}\,.$$

A comparison with the empirical distribution and the results of model 2 showed a much better fit. All the estimated variance components were nearly the same as in the simulation. First of all the estimation of Var(γ_{i2}) and Cov(γ_{i1}, γ_{i2}) was more accurate as in model 2.

Model 3: log-Transformation (λ=0, θ=0)

$\hat{\sigma}^2 = 0.00989$	$\hat{d}_{11} = 0.05525$	$\hat{d}_{12} = 0.02664$	$\hat{d}_{22} = 0.03654$
	Lower 95% CL	Estimate ± SE	Upper 95% CL
β_1	-0.117	-0.030 ± 0.0444	0.057
β_2	-0.089	-0.011 ± 0.0397	0.067

In the fourth model we tried another weighting function. The power function $g(\theta) = x^{\theta}$ based on the intention to model a possible influence of the substrate concentration on the response which yields the model equation

$$y_{ij} = f(\phi_i, x_{ij}) + x_{ij}^{\theta}\varepsilon_{ij}.$$

Obviously the application of this weighted model to the data was not very sucessful. All the variance components of Λ and also the overall variance σ^2 were highly underestimated. So this model seems to be unadequate to fit the simulated data very well.

Model 4: weighting with variance function $g(\theta)=x^{\theta}$ (λ=1, θ=0.45)

$\hat{\sigma}^2 = 0.00146$	$\hat{d}_{11} = 0.03839$	$\hat{d}_{12} = 0.02348$	$\hat{d}_{22} = 0.02872$
	Lower 95% CL	Estimate ± SE	Upper 95% CL
β_1	-0.097	-0.023 ± 0.0379	0.051
β_2	-0.090	-0.014 ± 0.0388	0.062

Because of the bad results from the fit with model 4 we extended the model by combining it with model 3. Then we get the weighted model with fixed log-transformation

$$\log(y_{ij}) = \log(f(\phi_i, x_{ij})) + x_{ij}^{\theta}\varepsilon_{ij}.$$

Model 5: weighting with variance function $g(\theta)=x^{\theta}$ and log-transformation (λ=0, θ=-0.04)

$\hat{\sigma}^2 = 0.00983$	$\hat{d}_{11} = 0.05620$	$\hat{d}_{12} = 0.02768$	$\hat{d}_{22} = 0.03661$
	Lower 95% CL	Estimate ± SE	Upper 95% CL
β_1	-0.117	-0.030 ± 0.0445	0.057
β_2	-0.108	-0.011 ± 0.0397	0.067

At first the estimation of the variance parameter $\hat{\theta} = -0.04$ indicated that the weighting function was not appropriate as we have seen for model 4. Therefore the use of model 5 showed a great advantage of models with more flexible variance/covariance structures. The log-transformation was obviously able to correct the inhomogenity in the simulated data set alone (see model 3). So the weighting is not necessary and by estimating the variance parameter through $\hat{\theta} = -0.04$ showed that model 5 was overparameterized (see also model 7). Nevertheless the estimates of all variance components of the random effects model were as good as in model 3.

The application of the log-transformation worked very well but in a next step we tried the more flexible model

$$y_{ij}^{(\lambda)} = f(\phi_i, x_{ij})^{(\lambda)} + \varepsilon_{ij}$$

where the parameter λ of the transformation will be estimated from the data.

Model 6: Transform-both-sides (λ=0.1, θ=0)

$\hat{\sigma}^2 = 0.00799$	$\hat{d}_{11} = 0.05662$	$\hat{d}_{12} = 0.02841$	$\hat{d}_{22} = 0.03681$
	Lower 95% CL	Estimate ± SE	Upper 95% CL
β_1	-0.121	-0.033 ± 0.0447	0.055
β_2	-0.089	-0.011 ± 0.0399	0.067

One more time model 6 showed the same advantages as model 3 and model 5. The parameter for the transformation was estimated near zero. So we could demonstrate the flexibility of this approach because similar to model 5 the underlying error structure of the simulated observations could be estimated with this model. The estimation of the variance components had also the same quality as in model 3 or model 5.

Finally we used the most flexible weighted transform-both sides model

$$y_{ij}^{(\lambda)} = f(\phi_i, x_{ij})^{(\lambda)} + x_{ij}^{\theta}\varepsilon_{ij}.$$

The weighted transform-both sides model showed similar results as model 6. Only the overall variance component σ^2 was clearly underestimated. The estimated value for the transformation parameter λ showed a slight deviation to model 6 which could be affected in this case by the inappropriate weighting function. Nevertheless the weighted transform-both-sides model is the most flexible one of all models we have examined. So in a situation of practical data analysis the error structure is unknown and therefore this model is the model of choice.

Model 7: Transform-both-sides and weighting with variance function
$g(\theta)=x^{\theta}$ (λ=0.23, θ=0.07)

$\hat{\sigma}^2 = 0.00589$	$\hat{d}_{11} = 0.05673$	$\hat{d}_{12} = 0.02891$	$\hat{d}_{22} = 0.03704$
	Lower 95% CL	Estimate ± SE	Upper 95% CL
β_1	-0.117	-0.029 ± 0.0448	0.057
β_2	-0.089	-0.012 ± 0.0402	0.067

Acknowledgements. This work was motivated by the research project 'Complexity Reduction in Multivariate Data Structures', supported by the Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich 475.

6. References

Box, G. E. P. and Cox, D. R. (1964). An analyses of transformations. Journal of the Royal Statistical Society, Series B 26, 211-246

Breslow N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association 88, 9-25

Carroll, R. J. and Ruppert, D. (1998). Transformations and weighting in Regression. London: Chapman and Hall

Edler, L., Gilberg, F., Portier, C. and Urfer, W. (1997). Statistische Verfahren zur Auswertung von Enzyminduktionskinetiken in der Risikoabschätzung. Informatik, Biometrie und Epidemiologie in Medizin und Biologie 28 (4), 213-226

Gilberg, F. (1996). Nichtlineare Regressionsmodelle mit zufälligen Effekten zur Schätzung von Parametern einer Enzym- oder Rezeptor-Bindungs-Kinetik. Dissertation, Fachbereich Statistik, Universität Dortmund

Giltinan, D. M. and Ruppert, D. (1989). Fitting heteroscedastic regression models to individual pharmacokinetic data using standard statistical software. J. Pharmacokin. Biopharm. 17, 601-614

Jones, R. H. (1993). Longitudinal Data with Serial Correlation: A State-space Approach. London: Chapman and Hall

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. Biometrics 46, 673-687.

Ruppert, D., Cressie, N. and Carroll, R. J. (1989). A Transformation/Weighting Model for Estimating Michaelis-Menten Parameters. Biometrics 45, 637-656.

Urfer, W. and Becka, M. (1996): Exploratory and model based inference in toxicokinetics. In: B.J.T. Morgan (Ed.): Statistics in Toxicology, 198-216. Oxford University Press