Fitting Threshold Models using the SAS® Procedures NLIN and NLMIXED
Todd Coffey, Washington State University, Pullman, WA

ABSTRACT
Threshold regression models have applications in diverse fields, including toxicology, cell biology, ecology, infectious disease, epidemiology, finance, and econometrics. The name of these models is derived from a location in the data in which the response changes abruptly and a new pattern emerges. The statistical model for data with this pattern requires two distinct equations, connected at the threshold. Often, the initial part of the model is a baseline response or background effect and the subsequent part after the threshold estimate includes a linear slope or nonlinear function. When the threshold, or common join point between the distinct models on either side of it is unknown, the statistical model is nonlinear in the parameters, regardless of the functional form of the piecewise models. In general, nonlinear models can be challenging to fit due to the lack of a closed-form estimate for parameters. Threshold models can provide additional difficulty due to the continuity and smoothness constraints at the threshold. SAS® has two procedures that are commonly used to fit threshold models. PROC NLIN fits nonlinear models and is used when observations in the dataset are independent. PROC NLMIXED fits nonlinear mixed models, and when observations come from the same subject, can be used to model the within-subject correlation using random coefficients. In this paper, we discuss the similarities and differences between these two SAS/STAT® procedures. Then, using two different datasets, one from an experiment on the toxicological effects of pesticides and another from a veterinary surgical study, we demonstrate how to fit threshold models, with or without random coefficients. We compare and contrast parameter estimates, standard errors, and other results from the algorithms in PROC NLIN and PROC NLMIXED.

INTRODUCTION
Threshold regression models have two different functional forms, connected at the threshold or join point, which separates the two patterns. The threshold can be thought of as the point where the response begins to change, or that exceeding the threshold will cause a dramatic change in the measured response. When the threshold is unknown, it is estimated as a parameter and the subsequent model is nonlinear in the parameters, regardless of the functional forms of the models on either side of it. These models are used in many fields, particularly those where an abrupt change in the response is a natural consequence of exceeding the limits of one or more variables. Examples from the literature include toxicology (Schwartz, Gennings, and Chinchili, 1995; Coffey and Gennings, 2007), cell biology (Hung 2012), ecology (Samia and Chan, 2011), infectious disease (Samia, Chan, and Stenseth, 2007), epidemiology (Pastor-Barriuso R, Guallar E, and Coresh J, 2003), finance (Pesaran and Pick, 2007), and econometrics (Lee and Seo, 2008).

Nonlinear regression models are more difficult to fit than those with only a linear slope because no closed-form equation exists for the parameters. Threshold models add to the difficulty because of the continuity and smoothness constraints at the threshold estimate. However, the NLIN procedure SAS® has been used successfully for decades to fit these types of models when the observations are independent. When observations are not independent, perhaps due to dependence in time, space, or other factors, within-subject correlation should be accounted for in the estimate of standard errors. PROC NLIN cannot estimate this correlation, but it can be accomplished using a relatively newer procedure, PROC NLMIXED. NLMIXED fits nonlinear mixed effects models, accommodating both fixed and random effects that are nonlinear in the parameters.

In this paper, we discuss the similarities and differences between these two SAS/STAT® procedures. Then, using two different datasets, one from an experiment on the toxicological effects of pesticides and another from a veterinary surgical study, we demonstrate how to fit threshold models, with or without random coefficients. We compare and contrast parameter estimates, standard errors, and other results from the algorithms in PROC NLIN and PROC NLMIXED.

MOTIVATING DATA
The first motivating dataset comes from a toxicological experiment on pesticides. The population of the United States (US) is exposed daily to pesticide residues on fruit. The US Environmental Protection Agency conducted a study to determine the neurological effects of a mixture of five organophosphorus pesticides commonly found on fruit. The study is described in detail in Moser et al. (2005). In brief, rats were dosed with either a vehicle control or one of six doses of the pesticide mixture. At the time of peak behavioral effects, rats were placed into a chamber where the animal’s gait was observed and a score calculated. Figure 1 shows the gait score expressed as the percentage of the control (0 mg/kg) for each of the rats. Decreasing scores represent a worse outcome. There is considerable variability in the scores of each dose, which is due in part to the subjectivity of the measurement. Of special interest is a relatively similar response for doses of 0 and 5 mg/kg. For this latter dose, there appears to be no deleterious
effect. However, doses greater than or equal to 30 mg/kg result in lower average gait scores, indicating a potential behavioral effect caused by the pesticide. Toxicologists are interested in estimating the dose threshold for which increased doses result in untoward effects.

![Figure 1. Gait Score (% of Control) from 80 Rats Treated with a Vehicle Control or One of Six Doses of a Pesticide Mixture.](image)

The second motivating dataset comes from canines receiving neurological evaluation at the Washington State University Veterinary Teaching Hospital (unpublished data). Elevated levels of creatinine kinase (CK) and lactate dehydrogenase (LD) taken from the cerebrospinal fluid (CSF) are known to be associated with neurological disease. During sampling of the CSF, the surgeon’s needle can sometimes unintentionally draw blood, which contains CK and LD. This iatrogenic blood contamination can lead to diagnoses that are more severe than warranted due simply to misinformation. In this experiment, CSF was taken from canines and then pooled to form six independent samples. Each sample was then serially diluted 8 times in blood. The amount of LD and CK and their isoenzymes was then measured. Figure 2 displays the levels of total CK (expressed in log units) for each sample measured neat (no dilution) and with each of the 8 hemodilutions. The figure shows that samples had varying levels of background total CK. Between 3 and 6 dilutions, each sample experienced a sustained increase in the total CK level due to the dilution with blood containing CK. The objective was to determine the threshold beyond which total CK levels increased above background so that clinicians can understand the limit of blood contamination before improper diagnoses are made.
STATISTICAL METHODS

Threshold regression models are part of the general class of nonlinear models. The nonlinear model can be written for a single response $y_i$, vector of predictor variables $x_i$, and vector of model parameters $\theta$ with $i=1, 2, \ldots, N$ observations as $y_i = f(x_i, \theta) + \epsilon_i$. Generically, the mean response of a threshold model for a single parameter can be written as shown in Equation 1. Here, $\delta$ is called a threshold parameter because the model for values less than the threshold parameter differs from the model greater than or equal to it. If the threshold parameter is unknown, the resulting model is nonlinear regardless of the form of the two different parts of the model. Often, the part of the model denoted by $f_1$ is a background response or intercept and $f_2$ is of polynomial or nonlinear form.

\[
\begin{align*}
  f(x_i, \theta) &= \begin{cases} 
    f_1(x_i, \theta) & x_i < \delta \\
    f_2(x_i, \theta) & x_i \geq \delta 
  \end{cases} 
\end{align*}
\]  

Equation 1

TOXICOLOGICAL DATASET

For the toxicological dataset, the nonlinear threshold model can be written as shown in Equation 2. The parameter $\alpha$ is the lowest response, $\alpha+\gamma$ is the maximum response, $\theta$ is a slope parameter, $\delta$ is the threshold parameter, and $x_i$ is the dose. Note that the part of this model greater than the threshold is nonlinear, even before the addition of $\delta$.

\[
\begin{align*}
  f(x_i, \theta) &= \begin{cases} 
    \alpha + \gamma & x_i < \delta \\
    \alpha + \gamma e^{\theta(x_i-\delta)} & x_i \geq \delta 
  \end{cases} 
\end{align*}
\]  

Equation 2

SURGICAL DATASET

For the surgical dataset, if we assume that the residuals are independent and normally distributed with a mean of 0 and common variance, we can write the model as shown in Equation 3. This nonlinear model has a background response parameter of $\alpha$, and includes a linear slope parameter ($\beta$) after the threshold $\delta$.

\[
\begin{align*}
  f(x_i, \theta) &= \begin{cases} 
    \alpha & x_i < \delta \\
    \alpha + \beta(x_i - \delta) & x_i \geq \delta 
  \end{cases} 
\end{align*}
\]  

Equation 3
Due to the presence of within-sample correlation due to multiple dilutions of the same sample, a nonlinear mixed model with a random coefficient for the background response is more appropriate and is shown in Equation 4. In this model, α is the background measurement level, β is a linear slope parameter, δ is the threshold parameter, \( x_i \) is the dilution number for the \( i \)th sample \((i=0, 1, 2, ..., 8)\), and \( u_0 \) is the random background coefficient. The random coefficient allows the background level to differ for each subject and induces correlation among all dilutions of the same sample. The statistical assumptions made for this model are that the random background level is considered to be independent from the residual variance and that both are normally distributed with a mean of 0. The variance of the random effect can be written as \( \sigma^2_u \) while the residual variance is written as \( \sigma^2 \). The total variance for any measurement is \( \sigma^2_u + \sigma^2 \), the covariance between any two measurements from the same dilution series is \( \sigma^2_u \), and the resulting correlation is equal to \( \frac{\sigma^2_u}{\sigma^2_u + \sigma^2} \).

\[
f(x_i, \theta) = \begin{cases} 
\alpha + u_0 & x_i < \delta \\
(\alpha + u_0) + \beta \cdot (x_i - \delta) & x_i \geq \delta 
\end{cases}
\tag{Equation 4}
\]

**COMPARISON OF SAS® PROCEDURES**

Both PROC NLIN and PROC NLMIXED fit nonlinear models. These two procedures have many commonalities: both use iterative fitting methods, provide a grid of starting values for parameters, and automatically calculate analytic derivatives. They also differ on important features. PROC NLIN uses nonlinear least squares while NLMIXED uses maximum likelihood. Only NLMIXED can induce within-subject correlation through the addition of random effects. NLMIXED also requires that the residual variance be estimated as a parameter, which is not an option in NLIN. Although it won’t be discussed in this paper, NLMIXED has the ability to model data that comes from several distributions besides the Gaussian, or normal.

<table>
<thead>
<tr>
<th>Feature</th>
<th>PROC NLIN</th>
<th>PROC NLMIXED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use iterative fitting methods</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Parameters can be given a grid of starting values</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Analytic derivatives are calculated automatically</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Estimation of parameters</td>
<td>Nonlinear least squares</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td>Enabled for random effects (within-subject correlation)</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>The residual variance is an estimated parameter</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Allows for response variable to follow distributions other than Normal</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Table 1. Comparison of SAS® Procedures

**USING AND UNDERSTANDING SAS® CODE**

**TOXICOLOGICAL DATASET**

The SAS® code used to fit Equation 2 to the toxicological dataset with PROC NLIN is shown at the end of this section. The parameters of this nonlinear model are estimated using nonlinear least squares. NLIN includes four different algorithms that minimize the sum of squared errors (SSE). Because of the nonlinearity in the parameters, no closed-form solution exists. Further details of each algorithm can be found in the SAS/STAT User’s Guide.

The four algorithms that can be used to minimize the SSE are the Gauss-Newton (denoted by METHOD=gauss), Marquardt, Newton, and gradient. Of these, the first three are more robust than the last. The Gauss-Newton method approximates the nonlinear model using a first-order Taylor series and is the default. The Marquardt algorithm is a compromise between the Gauss-Newton and gradient methods. It imposes a ridging penalty that varies between 0 and positive infinity and performs well with highly correlated parameter estimates. A penalty of 0 provides identical results to the Gauss-Newton method while a penalty of infinity gives the same solution as the gradient method. A penalty between the extremes provides a compromise. The Newton method approximates the SSE directly using the
second derivatives. The gradient method converges very slowly and is recommended only when the initial values are poor.

The parameters are given starting values in the PARMS statement as shown below. A grid can be specified by adding multiple starting values for each parameter using the TO and BY options. Additional programming statements that define the statistical model or other constants can be added. In this case, the mu= line specifies the statistical model of Equation 2. The MODEL statement is required and takes the form of response=model, where the only allowed distribution for the response is normal. The equation for the model (mu=) could have been added directly to this line. An output dataset with predicted values, confidence intervals, and other diagnostic and inferential statistics can be obtained with the OUTPUT statement. NLIN can also estimate parameters using weighted least squares with weights contained in a variable listed in the WEIGHT statement and can constrain parameter estimates using the BOUNDS statement (not shown).

```
proc nlin method=gauss (*or marquardt, newton, gradient);
   parms alpha=45 to 55 by 5 theta=-0.05 to -0.01 by 0.01 gamma=45 to 55 by 5
delta=10 to 30 by 10;
   mu=(alpha+gamma)*(dose<delta) + (alpha+gamma*exp(theta*(dose-
delta)))*(dose>=delta);
   model per_ctrl=mu;
   output out=out p=pred;
run;
```

SURGICAL DATASET

The SAS® code used to fit Equation 4 to the surgical dataset with PROC NLMIXED is shown at the end of this section. Nonlinear mixed models are more difficult to fit than nonlinear models with uncorrelated observations because they require maximizing a marginal likelihood function in addition to an optimization method. This marginal likelihood function must be integrated over the random coefficients. NLMIXED includes four different choices to approximate the marginal likelihood. To obtain parameter estimates, several different optimization techniques are included. Three different methods can be used to compare different models, including AICC (Akaike’s Information Criteria with finite-sample correction).

The four approximations to the marginal likelihood function that can be specified in NLMIXED are the adaptive Gauss-Hermitage quadrature, first-order, Hardy quadrature, and adaptive importance sampling. These are specified using METHOD= in the PROC NLMIXED statement. Pinheiro and Bates (1995) provide evidence that adaptive Gauss-Hermitage quadrature has many advantages for most situations and it is the default method. The first-order method which uses a Taylor series approximation is well-known but can only be used when the response variable is normally distributed. Importance sampling uses Monte Carlo integration. Hardy quadrature uses an adaptive trapezoidal rule and is more restrictive because it can only be used with a single random effect. In addition to the approximation methods, several different choices for optimization are available and can be specified using TECH= in the NLMIXED statement. Because it provides a good balance between speed and stability, quasi-Newton optimization is the default (TECH=QUANEW). Details of these optimization methods can be found in the SAS/STAT User’s Guide. As with PROC NLIN, the parameters are given starting values using the PARMS statement. In this case, the variance parameters s2e and s2u0 were not given starting values, so they default to 1, but values could be specified. The BOUNDS statement forces the residual variance ε to be positive and the random background coefficient u0 to be 0 or greater. For this dataset, these commands are not necessary because both variance components are estimable, but the statement prevents a solution with negative variance components. As with NLIN, programming statements such as that shown in the line beginning with 'mod=' can be added. The MODEL statement is also required but it takes a much different form than that in NLIN. The model is specified using the response, followed by a tilde, and then the conditional distribution given the random effects. In this case, the log values of total CK are modeled using a normal distribution with mean described by the function called ‘mod’ and variance denoted by the residual variance, s2e. Several other distributions, including binomial, Poisson, and a general log likelihood function, are available. The random background coefficient is added using the random statement. Again, the random effect has a distribution; in this case we specify it as a normal distribution with a mean of 0 and variance of s2u0, which is just the name for the variance component attributed to the random background coefficient. As opposed to the model statement, the distribution of the random effects must be normal. We also must add the SUBJECT= command to tell the algorithm which levels of the variable are correlated. The PREDICT statement is useful in creating a wide variety of estimated functions which can be put into an output dataset. In this case, the code simply asks to get the predictions for individual values—which include the random effects—at each point in the dataset and include them into a new dataset called outpred_lck_total.
proc nlmixed data=all method=firo; (*or gauss, hardy, or isamp) tech=quanew;
    parms alpha=0 to 0.2 by 0.2 beta=0.1 to 0.3 by 0.1 delta=3 to 6 by 1;
    bounds s2e>0, s2u0>=0;
    mod=(alpha+u0)*(dilution<delta) + ((alpha+u0) + beta*(dilution-delta))*(dilution>=delta);  
    model lck_total~normal(mod,s2e);
    random u0 ~ normal(0,s2u0) subject=series;
    predict mod out=outpred_lck_total;
run;

RESULTS

TOXICOLOGICAL DATASET

The parameter estimates (Table 2) and standard errors (Table 3) for Equation 2 using all four fitting methods in NLIN are shown below. Notice that the parameter estimates for the Gauss-Newton, Marquardt, and Newton methods are identical. The gradient method has considerably different estimates. One way of assessing model fit is by looking at the resulting SSE. Because its minimized SSE is larger than the other methods, the parameter estimates are considered to be less accurate than the others. Although fitting threshold models can be challenging, as evidenced by the different results from the gradient method, this dataset shows consistent parameter estimates from three methods. The standard errors for the Gauss-Newton and Marquardt methods are also identical. This equivalency seems to indicate that a ridging penalty was not imposed by the Marquardt method so it defaulted to the Gauss-Newton method. The standard errors for the Newton method are all lower than the previous two.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Gauss-Newton</th>
<th>Marquardt</th>
<th>Newton</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>46.8</td>
<td>46.8</td>
<td>46.8</td>
<td>50.0</td>
</tr>
<tr>
<td>γ</td>
<td>53.1</td>
<td>53.1</td>
<td>53.1</td>
<td>50.0</td>
</tr>
<tr>
<td>θ</td>
<td>-0.033</td>
<td>-0.033</td>
<td>-0.033</td>
<td>-0.042</td>
</tr>
<tr>
<td>δ</td>
<td>17.1</td>
<td>17.1</td>
<td>17.1</td>
<td>20.0</td>
</tr>
<tr>
<td>SSE</td>
<td>16232.6</td>
<td>16232.6</td>
<td>16232.6</td>
<td>16532.8</td>
</tr>
</tbody>
</table>

Table 2. Parameter Estimates for Equation 2 using Four Different Optimization Algorithms in NLIN

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Gauss-Newton</th>
<th>Marquardt</th>
<th>Newton</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>3.01</td>
<td>3.01</td>
<td>2.85</td>
<td>2.74</td>
</tr>
<tr>
<td>γ</td>
<td>4.44</td>
<td>4.44</td>
<td>4.34</td>
<td>4.29</td>
</tr>
<tr>
<td>θ</td>
<td>0.013</td>
<td>0.013</td>
<td>0.011</td>
<td>0.018</td>
</tr>
<tr>
<td>δ</td>
<td>7.16</td>
<td>7.16</td>
<td>6.43</td>
<td>5.84</td>
</tr>
<tr>
<td>SSE</td>
<td>16232.6</td>
<td>16232.6</td>
<td>16232.6</td>
<td>16532.8</td>
</tr>
</tbody>
</table>

Table 3. Standard Errors for Parameter Estimates in Equation 2 using Four Different Optimization Algorithms in NLIN

SURGICAL DATASET

To contrast the two procedures, the surgical dataset was fit using both PROC NLIN and PROC NLMIXED in three different approaches. First, each sample was fit independently using NLIN. Then, all six samples were fit simultaneously using NLIN. Finally, all samples were fit simultaneously using NLMIXED and incorporated the within-sample correlation.

Table 4 contains the parameter estimates for Equation 3 when each sample was fit independently using PROC NLIN. The advantage of fitting each series independently is that a customized fit is achieved for each sample. The disadvantage is that three parameters are needed for each sample, thus the overall model used 18 parameters. Notice that the threshold parameter for these six samples ranged from 2.97 to 5.82. Table 5 contains the empirical
results (labeled as ‘Averaged Estimates from PROC NLIN’), or in other words, averages of the parameter estimates from Table 4, and the empirical standard errors and 95% confidence intervals.

<table>
<thead>
<tr>
<th>Series</th>
<th>Parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>0.57</td>
<td>0.70</td>
<td>1.11</td>
<td>0.70</td>
<td>0.95</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>0.15</td>
<td>0.13</td>
<td>0.12</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>δ</td>
<td>3.87</td>
<td>3.69</td>
<td>5.82</td>
<td>3.43</td>
<td>4.25</td>
<td>2.97</td>
</tr>
</tbody>
</table>

**Table 4. Parameter Estimates for Equation 3 by Each Series using NLIN**

Table 5 also contains the results for the two other approaches: the parameter estimates for Equation 3 fit simultaneously for all six samples using PROC NLIN (labeled as ‘Model Estimates from PROC NLIN’)—assuming all observations are independent—and the estimates from PROC NLMIXED (using the default options), which incorporates the within-sample correlation. Notice that the parameter estimates for the background response and slope are similar for all three analyses. However, the threshold parameter is considerably larger with NLMIXED. Its 95% confidence interval includes the estimates from the first two approaches so there is not a statistical difference. Interestingly, the standard errors for the threshold and slope parameters are both noticeably lower using NLMIXED but are in the middle of the estimates for the background response of the two analyses using NLIN. It is not surprising that the first approach has the highest standard error for α because each sample was fit independently. The difference between the second two approaches is the inclusion of the within-sample correlation.

<table>
<thead>
<tr>
<th>Averaged Estimates from PROC NLIN</th>
<th>Model Estimates from PROC NLIN</th>
<th>Model Estimates from PROC NLMIXED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Mean</td>
<td>Std Err</td>
</tr>
<tr>
<td>α</td>
<td>0.84</td>
<td>0.080</td>
</tr>
<tr>
<td>β</td>
<td>0.09</td>
<td>0.019</td>
</tr>
<tr>
<td>δ</td>
<td>4.00</td>
<td>0.40</td>
</tr>
</tbody>
</table>

**Table 5. Parameter Estimates, Standard Errors, and 95% Confidence Intervals for Equations 3 and 4 using Three Different Methods**

The surgical dataset was also fit to Equation 4 using four different approximations to the marginal likelihood. The tables below show the parameter estimates (Table 6) and standard errors (Table 7) from the four different likelihood approximations available in PROC NLMIXED. The first order, Gauss-Hermitage, and importance sampling methods all produced similar if not identical parameter estimates and standard errors. Only the Hardy quadrature method had difficulty fitting the model to the data. A warning in the log indicated that the final Hessian matrix had at least one negative eigenvalue and that the second-order optimality criterion had been violated. Although this warning indicates that the results should not be interpreted, the threshold estimate was identical and the slope estimate similar to the other methods. Note also that the AICC was considerably larger, another indication that this method did not fit the data nearly as well as the others. When the data was refit using Hardy quadrature and the Nelder-Mead simplex, an optimization technique that does not require derivatives, the fit was improved and all standard errors were estimated. More research could be done on the optimal combination of likelihood approximations and optimization techniques. The predicted and observed values from the Gauss method for Equation 2 and the first order method for Equation 4 are shown in Figures 3 and 4, respectively.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>First Order</th>
<th>Gauss-Hermitage</th>
<th>Hardy</th>
<th>Importance Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.85</td>
<td>0.85</td>
<td>0.26</td>
<td>0.85</td>
</tr>
<tr>
<td>β</td>
<td>0.095</td>
<td>0.095</td>
<td>0.094</td>
<td>0.095</td>
</tr>
<tr>
<td>δ</td>
<td>4.46</td>
<td>4.46</td>
<td>4.46</td>
<td>4.46</td>
</tr>
<tr>
<td>s2e</td>
<td>0.0065</td>
<td>0.0065</td>
<td>0.027</td>
<td>0.0065</td>
</tr>
<tr>
<td>s2u0</td>
<td>0.024</td>
<td>0.024</td>
<td>0*</td>
<td>0.024</td>
</tr>
<tr>
<td>AICC</td>
<td>-86.1</td>
<td>-86.1</td>
<td>-10.7</td>
<td>-86.9</td>
</tr>
</tbody>
</table>

*--the variance component could not be estimated because the estimate wanted to go below its lower bound

Table 6. Parameter Estimates for Equation 4 using Four Different Likelihood Approximations in NL MIXED

<table>
<thead>
<tr>
<th>Parameter</th>
<th>First Order</th>
<th>Gauss-Hermitage</th>
<th>Hardy</th>
<th>Importance Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.065</td>
<td>0.065</td>
<td>0.103</td>
<td>0.064</td>
</tr>
<tr>
<td>β</td>
<td>0.015</td>
<td>0.015</td>
<td>.</td>
<td>0.015</td>
</tr>
<tr>
<td>δ</td>
<td>0.396</td>
<td>0.396</td>
<td>.</td>
<td>0.396</td>
</tr>
<tr>
<td>s2e</td>
<td>0.001</td>
<td>0.001</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>s2u0</td>
<td>0.014</td>
<td>0.014</td>
<td>.</td>
<td>0.014</td>
</tr>
<tr>
<td>AICC</td>
<td>-86.1</td>
<td>-86.1</td>
<td>-10.7</td>
<td>-86.9</td>
</tr>
</tbody>
</table>

Table 7. Standard Errors for Parameter Estimates of Equation 4 using Four Different Likelihood Approximations in NL MIXED

Figure 3. Equation 2 Predicted Values and Observed Gait Scores
<Fitting Threshold Models using the SAS® Procedures NLIN and NLMIXED>, continued
Figure 4. Equation 4 Predicted Values and Observed Total CK Values
CONCLUSION

In this paper we highlighted similarities and differences between the SAS® procedures NLIN and NLMIXED in fitting nonlinear threshold models, demonstrated how to fit threshold models with (NLMIXED) and without (NLIN) random coefficients, and compared the results using different optimization methods and approximations to the marginal likelihood. One principal advantage of using NLMIXED is that random effects can be estimated to properly model within-subject correlation, which NLIN cannot do. In each procedure, 3 of the 4 methods gave similar results. The gradient method in NLIN and the Hardy quadrature approximation in NLMIXED produced different fits and results. More research is needed to understand the best optimization methods to use with each marginal likelihood approximation in NLMIXED.

REFERENCES


ACKNOWLEDGMENTS

The author thanks Annie Chen, DVM, and Jourdan Brune for providing the surgical dataset.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Todd Coffey  
Enterprise: Washington State University  
Address: PO Box 643113  
City, State ZIP: Pullman, WA 99164-3113  
E-mail: todd.coffey@wsu.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.