ABSTRACT
Multilevel models have been used to analyze nested data to take into account the dependence structure that may be present. This is accomplished by including random coefficients in the model. In this work, we consider the multilevel framework adapted to common situations in Survival Analysis. Two approaches are discussed: one based on a parametric model (Weibull), and other on the semi parametric Cox proportional hazards model. Assuming that the random coefficients are normally distributed, we discuss how to use GLIMMIX macro to fit both models. We also present a macro to optimize the procedure and show how to use it in a real data example, based on an experiment with rats.

INTRODUCTION
Multilevel models were developed to analyze nested data [4]. In survival analysis, there are many ways in which the data may be nested: when the event of interest is recurrent, repeated measurements in the same experimental unit or simply when the data can be naturally grouped. In Section 4 we discuss a real data example of an experiment with rats undertaken in the Biology department of the University of São Paulo, Brazil. In this experiment, several measurements were taken in each rat; in multilevel terminology, we have units grouped in different levels, i.e., the measurements taken in a given rat are considered units of the first level and each animal is a unit of the second level. This example is a two level study, however multilevel models allow that units can be grouped in more general hierarchy.

Standard models are not suitable for nested data because the independence assumption is not generally true. In order to take into account the dependence underlying the observations, random coefficients are included in the specification of the model, which is the main difference between multilevel models and the usual ones.

The estimation procedures for multilevel generalized linear models are based on the penalized quasi likelihood and marginal quasi likelihood, both of which are already implemented in the GLIMMIX macro. For multilevel models in survival analysis, we implemented two macros which allow the use of GLIMMIX macro for the estimation of parameters in some proportional hazard models.

SPECIFICATION OF A MULTILEVEL MODEL IN SURVIVAL ANALYSIS
We consider a two level model. However, extension to a model with more levels is straightforward. We denote by Ti the time up to the occurrence of the event of interest for the i-th observation of the first level in the j-th unit of the second level. We consider that there are J units on the second level and Ij units on the first level belonging to the j-th unit of the second level. It is observed the random variable Ti unless the observation is censored by a nonnegative random variable Ci. As usually, we define the random quantities

\[ Y_{ij} = \min (T_{ij}, C_{ij}) \]
\[ \delta_{ij} = I(T_{ij} \leq C_{ij}) \]

For the second level, we assume that observations in the same cluster are dependent, whereas for different clusters they are independent.

In this work, we focus on proportional hazards models, in which the hazard function can be expressed in the form

\[ \alpha_0(t) = \alpha_0(t) \exp(\beta x_{ij}) \]

where \( \alpha_0(t) \) is a known function that does not involve \( t \). Two cases are considered: \( \alpha_0(t) \) is an arbitrary function (Cox proportional hazards model) and \( \alpha_0(t) \) is specified (exponential and Weibull models).

In the multilevel model, the linear predictor may be written in the form \( \eta_{ij} = \mathbf{x}_{ij} \beta + \mathbf{z}_{ij} \theta \), where \( \mathbf{z}_{ij} \) is a known matrix of covariates and \( \mathbf{b} \) is a vector of random variables normally distributed with means equal to zero. Notice that the vector \( \mathbf{z}_{ij} \) is the same for all units in the first level belonging to the same cluster. We also assume that the elements of the vector \( \mathbf{b} \) are independent, so that the covariance matrix is diagonal. We denote this matrix by \( \mathbf{D}(\theta) \), where \( \theta \) is a vector containing the unknown parameters of the covariance matrix.

PROPORTIONAL HAZARDS PARAMETRIC MODELS
The models considered here are proportional hazards models in which the hazard function can be expressed as

\[ \alpha_0(t | \mathbf{x}_{ij}) = \alpha_0(t) \exp \left( \sum_{j=1}^{J} \beta_j x_{ij} \right) \]

and

\[ \alpha_{ij}(t | \mathbf{x}_{ij}) = \frac{f_{ij}(t | \mathbf{x}_{ij})}{S_{ij}(t | \mathbf{x}_{ij})} = \frac{f_{ij}(t | \mathbf{x}_{ij})}{1 - S_{ij}(t | \mathbf{x}_{ij})} \]

it follows that the conditional probability density function of the failure time can be expressed as

\[ f_{ij}(t | \mathbf{x}_{ij}) = \alpha_0(t) \exp (\eta_{ij} - \alpha_0(t) \exp (\eta_{ij})), \]

where \( \Lambda_0(t) \) is the baseline cumulative hazard function.

The likelihood function for censored data, considering the variables \( \delta_{ij} \) already defined, is given by

\[ L(\beta | \mathbf{b}) = \prod_{j=1}^{J} \prod_{i=1}^{I} \left( f(t_{ij} | \mathbf{x}_{ij}) S(t_{ij} | \mathbf{x}_{ij}) \right)^{\delta_{ij}} \]

where \( J \) is the number of units in the second level and \( I_j \) the number of first level units belonging to the \( j \)-th unit on the second level.

The log-likelihood is easily obtained, substituting \( f(t_{ij} | \mathbf{x}_{ij}) \) and \( S(t_{ij} | \mathbf{x}_{ij}) \) by their expressions, leading us to

\[ L(\beta | \mathbf{b}) = \sum_{i=1}^{I_j} \left[ \delta_{ij} \left( \log (\alpha_0(t_{ij})) + \eta_{ij} - \alpha_0(t_{ij}) \exp (\eta_{ij}) \right) \right] \]

Adding and subtracting \( \log (\Lambda_0(t_{ij})) \) in the above expression, and denoting \( \mu_{ij} = \log (\Lambda_0(t_{ij}) \exp (\eta_{ij})) \), it follows the log-likelihood can be rewritten as
\[ l(\beta \mid b) = \sum_{i,j} \left[ \delta_{ij} \log \mu_{ij} - \mu_{ij} \right] + \sum_{i,j} \delta_{ij} \log \left( \frac{\alpha(t_{ij})}{\Lambda_0(t_{ij})} \right) \]  

(3)

The second term in the right hand side does not involve the unknown vector of parameters \( \beta \). The first term is the one of interest when we are estimating \( \beta \) and it can be compared to a Poisson log-likelihood. Recall that the log likelihood of independent Poisson random variables \( w_k \), with \( E(w_k) = \mu_k \), is

\[ l = \sum_{i,j} \left( w_{ij} \log \mu_{ij} - \mu_{ij} \right) - \sum_{i,j} \log w_{ij} + \sum_{i,j} \left( w_{ij} \log \mu_{ij} - \mu_{ij} \right) + k. \]  

(4)

Comparing the first term in the right hand side of (3) and expression (4), we see that the log-likelihood (3) can be viewed as a Poisson log-likelihood. Therefore, it is possible to use estimating methodologies developed for generalized linear models to obtain the numeric estimates of some parametric models in survival analysis.

For the exponential model, we have \( \alpha(t_{ij} \mid x_{ij}) = e^{\beta x_{ij}} \) and the probability density function is

\[ f(t_{ij}) = \exp(\eta_{ij} - t_{ij} \lambda_{ij}). \]

Therefore \( \log \mu_{ij} = \log t_{ij} + \log \lambda_{ij} \). One must be careful and notice that here is an offset \( \log t_{ij} \), which is a covariate with known coefficient. The parameters of the exponential model can be easily obtained using generalized linear model with the Poisson distribution.

For the Weibull distribution, the hazard function is

\[ \alpha(t_{ij} \mid x_{ij}) = \lambda_{ij} t_{ij}^{\rho}_{ij}, \]

so that

\[ \alpha_0(t_{ij}) = \lambda_{ij} t_{ij}^{\rho}_{ij} \]

and

\[ \Lambda_0(t) = \lambda_{ij} t_{ij}^{\rho}_{ij}. \]

In this case, the second term of the right side of (3) does not involve the vector \( \beta \), but \( \alpha_0(t_{ij})/\Lambda_0(t_{ij}) = \lambda_{ij} t_{ij}^{\rho}_{ij} \) depends on the unknown parameter \( \lambda \), which also must be estimated. Because of that, the estimating procedure must be modified accordingly.

In the Weibull model, the log-likelihood is given by

\[ l = \sum_{i,j} \delta_{ij} (\log \lambda - \log t_{ij} + \sum_{i,j} (\delta_{ij} \log \mu_{ij} - \mu_{ij}). \]

where \( \mu_{ij} = \lambda_{ij} t_{ij}^{\rho}_{ij} \).

The suggestion given by [5] is to find the estimating equation for \( \lambda \) and add a new step to the estimating procedure. Taking the derivative of \( l \) and making \( \partial l/\partial \lambda = 0 \), we have

\[ \frac{\partial l}{\partial \lambda} = \sum_{i,j} \delta_{ij} \frac{\lambda_{ij} t_{ij}^{\rho}_{ij}}{\lambda} + \sum_{i,j} (\delta_{ij} - \mu_{ij}) \log t_{ij} = 0, \]

and, hence,

\[ \hat{\lambda} = \frac{\sum_{i,j} (\delta_{ij} - \mu_{ij}) \log t_{ij}}{r}, \]

where \( r \) is the number of non-censored observations.

The estimating procedure in the Weibull model starts with \( \lambda = 1 \), i.e., exponential distribution. We get the values of \( \beta \) and a new value of \( \lambda \) is obtained using (5). With this new value, we get a new vector \( \beta \), and this procedure is repeated until convergence.

**COX PROPORTIONAL HAZARDS MODEL**

In this section we consider here the Cox model in which the hazard function can be written as

\[ \alpha(\mid x) = \alpha_0(t) \exp(\gamma) \]

where \( x \) is known vector of covariates and \( \eta \) is the linear predictor. In the Cox model, the function \( \alpha_0(t) \) is arbitrary. In order to estimate the parameter vector of the model, Cox [3] proposed the partial likelihood, which, assuming there are no ties, is given by

\[ L = \prod_{i \in F} \frac{\exp(\eta_{iy(i)})}{\sum_{R(i)} \exp(\eta_{iy(i)})}. \]

where \( R(t) \) is the set of individuals at risk and \( F \) is the set of individuals whose failure time was observed (not censored).

In multilevel models, random coefficients are added in the linear predictor, i.e., \( \eta_{ik} = x_i \beta + z_i \beta \). As we did in the last section, we also work in this section with the conditional distribution given the vector of random coefficients \( \beta \). In the Cox model, it is also possible to obtain the partial maximum likelihood estimator of the vector of parameters using generalized linear models, with Poisson distribution. To do so, we must create new variables and construct the corresponding likelihood function.

Let \( t_k, k = 1, \ldots, L \) be the times in which a failure has occurred. For each \( t_k \), we create the following variables for the observations belonging to \( R(t_k) \):

\[ y_{ij(k)} = \begin{cases} 1 & \text{if the observation failed at } t_k, \\ 0 & \text{otherwise} \end{cases} \]

(7)

where \( j \) refers to units on the first level, \( j \) to units on the second level and \( k \) is referring to the failure time. Note that we create \( n_k \) new variables for each observation, where \( n_k \) is the number of risks set that this observation belongs to.

Suppose now that \( y_{ij(k)} \) is distributed as a Poisson variable with mean \( \mu_{ij(k)} = \exp(\alpha_i + \eta_{ij(k)}) \). Also, suppose that these variables are all independent. The likelihood for the failure time \( t_k \) may be easily calculated as

\[ l_k = \prod_{i \in R(i)} \exp(-\sum_{R(i)} \mu_{ij(k)}) = \exp \left( \sum_{R(i)} \exp(\eta_{ij(k)}) \right) \]

According to the results of the Poisson distribution, we have that

\[ \sum_{R(i)} \mu_{ij(k)} = \text{various on the second level} \]

and

\[ \exp \left( \sum_{R(i)} \exp(\eta_{ij(k)}) \right). \]

The maximum value of the likelihood function is then

\[ L^* = \prod_{i \in R(i)} \sum_{R(i)} \exp(\eta_{ij(k)}) \exp(1) = \prod_{i \in R(i)} \exp(\eta_{ij(k)}). \]

This expression is proportional to (6) and, because of that, the value of \( \beta \) which maximizes expression (6) must coincide with the one which maximizes the expression of \( L^* \) above. Therefore, methods developed for generalized linear models can be used in order to get the partial maximum likelihood estimator of \( \beta \). Also, when there are ties, it can be shown that the likelihood constructed using the Poisson variables created here is
proportional to the Breslow's approximation to the partial likelihood.

In both parametric and semi parametric approaches, standard errors for the parameter estimates can be obtained using
generalized linear models, provided that the observed information is used.

**ESTIMATION IN MULTILEVEL MODELS**

In the previous two sections it was shown that estimation of
proportional hazards models can be done using generalized linear models. We use this fact to estimate the parameters of interest in multilevel models, i.e., models with random effects.

In the generalized linear model framework, we suppose that the response variable has a distribution that belongs to the exponential family. Denote by \( f_y(y|\mathbf{b}) \) the conditional probability density function of \( y|\mathbf{b} \) and \( f_\theta(\mathbf{b}) \) the probability density function of the random vector \( \mathbf{b} \). The marginal likelihood may be calculated by

\[
L(\mathbf{b}, \theta) = \int f_y(y|\mathbf{b}) f_\theta(\mathbf{b}) d\mathbf{b}.
\]

In particular, if the distribution of the response variables is Poisson and assuming that \( \mathbf{b} \) is normally distributed, we have

\[
L(\mathbf{b}, \theta) \propto \exp \left\{ \sum_{i,j} y_{ij} \log \mu_{ij} - \mu_{ij} - \frac{1}{2} \mathbf{b}^T \mathbf{D}^{-1} \mathbf{b} \right\} f_\theta(\mathbf{b}),
\]

where \( \mu_{ij}^\beta \) depends on the random vector \( \mathbf{b} \) and the integral (9) cannot be analytically calculated. Therefore, it is necessary to use some method of approximation of integrals. We used two different approaches, which were named Penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL). Both of these approaches are developed for normally distributed random coefficients. We do not describe these methods here and an heuristic derivation of them can be found in [2].

**ESTIMATION USING SAS SOFTWARE**

The PQL and MQL approach for estimating the parameters in generalized linear mixed models has already been implemented in SAS software (GLIMMIX macro); however the procedure described here has not been implemented in SAS software yet. Using the approach of equivalence of likelihoods described in the previous section, two macros were developed, one for the Weibull model and other for the Cox model. The estimation of the parameters in the proportional hazards model can be done using the macros we developed in conjunction with the GLIMMIX macro.

For the exponential case, the GLIMMIX can be used directly with the appropriate specification of the model and it is not necessary to develop any macro. The Poisson variables in this case are the censoring variables, which already exist in the original dataset. It is important to notice that the censoring variable must be appropriately defined, with the value 1 for failure and the value 0 for censored observations.

For the Weibull model, we also use the censoring variable as the Poisson variable, which already exists in the original dataset. However, it was necessary to develop the macro in order to include the new step in the estimating procedure for the estimation of the shape parameter \( \lambda \). The macro uses the GLIMMIX iteratively, according to the following steps:

1. Estimation of the parameters \( \beta \) under the exponential model, using GLIMMIX;
2. Estimation of the parameter \( \lambda \) using (5);
3. New values of the parameters \( \beta \) are obtained using the value of \( \lambda \) in the last step.

Steps 2 and 3 are repeated until convergence is reached. The last output of the GLIMMIX, which is an output of the mixed procedure, contains the estimates of the parameters of interest \( \beta \) and also for the parameters in the covariance matrix \( \mathbf{D}(\theta) \) for the random coefficients.

The macro developed for the Cox model creates variables in the same way it was described in section 2.1. The macro creates a new dataset which contains the Poisson variables described earlier. The output dataset is created in a way that GLIMMIX can be easily employed with the appropriate specification of the model. With GLIMMIX, we get the estimates of \( \beta \) and \( \theta \) (covariance parameters of the random coefficients).

**APPLICATION TO REAL DATA**

This methodology was applied to data from a study undertaken in the Biology department of University of São Paulo. The objective of the study was to verify if some areas of the brain of rats are involved with spatial memory. The experiment was carried out as follows.

The rats were divided into three groups. All rats in the first group had small brain lesions surgically made. The second group of rats also underwent surgery, but the lesions were in a different area of the brain. Finally, the third group underwent a similar surgery, however without any lesions (control group).

After the rats had recovered from the surgery, the experiment begun. The rats were put in a pool in which there was a hidden platform. Each animal had to swim until reaching the platform. This procedure was repeated during 14 days, with two trials per day, in a total of 28 trials for each rat. The response variable was the time until the rat reached the platform. If a rat could not find the platform by itself in two minutes, it was taken to the platform by the researchers and the observation was considered as censored. The main purpose of the investigators was to evaluate the learning process and to compare the three groups.

It is important to notice that we have 28 observations for each rat and, because of that, it is not appropriate to assume the observations are independent. This is the reason why the multilevel approach was chosen in this study. Each time observed is a unit of first level and each rat is a unit of the second level. The plot of the mean time (including the censored times) against the trial by group suggests that the rats learn where the platform is and the time spent until they reach it tends to decrease.
For this study, we used the Cox's semi parametric model with the effect of the day, trial (first or second in the day) and group as covariates. We considered the intercept as random, i.e., we included random coefficients additively in the linear predictor. In order to do so, we changed the parameterization of the Poisson variables means in section 2.2: $\mu_{ijk} = \exp(\alpha_i + \alpha_j + \beta_k)$, with $\alpha_0=0$. The results of the multilevel Cox's semi parametric model, using the PQL approach, are shown in Table 1.

The test regarding the covariance parameter is of great interest because if the hypothesis that this parameter is zero is not rejected, then the independence model can be used. We see in this data that this parameter is significantly different from zero, which shows that the observations are not independent. We also see that the parameter associated with group 1 of rats is significant, which shows that the times observed from rats in that group are different from the times associated with rats in group 3 (control group). The relative risk equals approximately 0.17, indicating that the chance that a rat from group 1 reaches the platform is 0.17 times the chance of a rat from group 3 reaching it, i.e., rats from group 1 take more time to reach the platform than rats from group 3. Therefore, this data indicate that the lesions in the brain of rats in group 1 do affect their spatial memory. The difference between groups 2 and 3 is not significant, showing that the lesions in the brain of rats in group 2 may not affect their spatial memory.

### Table 1: Results of Cox semiparametric multilevel model

<table>
<thead>
<tr>
<th>Param.</th>
<th>Estim.</th>
<th>Stand. error</th>
<th>Exp(β)</th>
<th>Wald test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>-2.482</td>
<td>0.2394</td>
<td>0.08</td>
<td>-10.37</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Day 2</td>
<td>-2.027</td>
<td>0.2190</td>
<td>0.13</td>
<td>-9.26</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Day 3</td>
<td>-1.563</td>
<td>0.1994</td>
<td>0.21</td>
<td>-7.84</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Day 4</td>
<td>-1.472</td>
<td>0.1979</td>
<td>0.23</td>
<td>-7.44</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Day 5</td>
<td>-1.022</td>
<td>0.1885</td>
<td>0.36</td>
<td>-5.42</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Day 6</td>
<td>-0.843</td>
<td>0.1852</td>
<td>0.43</td>
<td>-4.55</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Day 7</td>
<td>-0.789</td>
<td>0.1856</td>
<td>0.45</td>
<td>-4.25</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Day 8</td>
<td>-0.393</td>
<td>0.1827</td>
<td>0.67</td>
<td>-2.15</td>
<td>0.0313</td>
</tr>
<tr>
<td>Day 9</td>
<td>-0.465</td>
<td>0.1844</td>
<td>0.63</td>
<td>-2.52</td>
<td>0.0117</td>
</tr>
<tr>
<td>Day 10</td>
<td>-0.249</td>
<td>0.1823</td>
<td>0.78</td>
<td>-1.37</td>
<td>0.1722</td>
</tr>
<tr>
<td>Day 11</td>
<td>-0.185</td>
<td>0.1814</td>
<td>0.83</td>
<td>-1.02</td>
<td>0.3081</td>
</tr>
<tr>
<td>Day 12</td>
<td>0.007</td>
<td>0.1807</td>
<td>1.01</td>
<td>0.04</td>
<td>0.9698</td>
</tr>
<tr>
<td>Day 13</td>
<td>-0.258</td>
<td>0.1813</td>
<td>0.77</td>
<td>-1.42</td>
<td>0.1544</td>
</tr>
<tr>
<td>Day 14</td>
<td>0.000</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial 1</td>
<td>-0.234</td>
<td>0.0740</td>
<td>0.79</td>
<td>-3.16</td>
<td>0.0016</td>
</tr>
<tr>
<td>Trial 2</td>
<td>0.000</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>-0.433</td>
<td>0.1688</td>
<td>0.65</td>
<td>-2.56</td>
<td>0.0104</td>
</tr>
<tr>
<td>Group 2</td>
<td>-0.176</td>
<td>0.1610</td>
<td>0.84</td>
<td>-1.09</td>
<td>0.2745</td>
</tr>
<tr>
<td>Group 3</td>
<td>0.000</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Covar. Param. 0.099 0.0380 6.81 0.0045

### CONCLUSION

In this paper we consider multilevel models in survival analysis, which are models that include random effects in order to take into account the dependence among observations. The estimating procedure is based on an equivalence of the likelihood of Poisson variables appropriately created and the likelihood of some proportional hazards models considered. We developed two macros in SAS software that allows the use of GLIMMIX for estimation of parameters in survival models with a two levels structure. This approach was applied in real data example and showed that the dependence among observations cannot be ignored.

### REFERENCES


### CONTACT INFORMATION

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

Gisela Tunes da Silva, Ph.D. Student
University of São Paulo
Rua do Matão, 1010
São Paulo, SP, Brazil - 05508-090
tunes@ime.usp.br

Antonio Carlos P. Lima, Assistant Professor
University of São Paulo
Rua do Matão, 1010
São Paulo, SP, Brazil - 05508-090
acarlos@ime.usp.br
www.ime.usp.br/~acarlos

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.