Developments in Statistics
Andrej Mrvar and Anuška Ferligoj (Editors)
Metodološki zvezki, 17, Ljubljana: FDV, 2002

A Sensitivity Analysis to Link Specification for Random Effects Models: A Case Study in Burnout Research

Lara Lusa[1], Patrizia Rozbowsky[2], and Dario Gregori[2]

Abstract

For discrete dependent variables hierarchical generalized linear models can be used to establish the relationship between response and covariates, introducing random effects for the groups.

The link function, that relates the linear predictor to the expected value, just as in the generalized linear models, is usually assumed to be known and fixed. In some cases it might be useful to improve the modeling flexibility, allowing the link to be a member of a class indexed by one or more unknown parameters, that can be estimated. Some families of link functions have been introduced, which for certain choices of the parameters reduce to some of the well known link functions, as *logit*, *probit* and *complementary log-log*. While the performance of these families has been investigated for the generalized linear models, not much work has been done for hierarchical generalized linear models. Here we consider, in particular, the global performance of the model within a family of link functions, possible changes in the set of statistically significant parameters for fixed and random effects, as well as problems related to comparison of clusters. All considerations will be based on a re-analysis of a data-set related to the study of burnout syndrome among teachers at various levels of instruction.

[1] Unit of Medical Statistics and Biometry, Istituto Nazionale per lo Studio e la Cura dei Tumori, Milano, Italy; e-mail: lara.lusa@istitutotumori.mi.it.

[2] Department of Economics and Statistics, University of Trieste, Italy.

The authors wish to thank an anonymous referee and the Editors for their comments and suggestions, which greatly improved the paper. Responsibility remains, of course, to the authors.

1 Introduction

Very often in data analysis there is the need to take the multilevel structure into account. When the sampling mechanism is of multistage type rather than simple random, the clustering of data leads to correlation between units belonging to the same cluster, and this fact has to be accounted for in the statistical analysis. In some cases, the correlation structure itself is of interest and studying it helps understanding the differences between clusters. Some well known examples where multilevel structure is present are studies where responses to treatment of patients treated in different hospitals are observed, subjects are observed repeatedly, observations are taken on pupils belonging to different classes and schools.

Statistical methods to take within group correlation into account have been proposed and are nowadays widely used in literature; correlation can be treated either as a nuisance (Liang and Zeger, 1986) or as a phenomenon to be investigated (Searle et al., 1992).

For discrete dependent variables, hierarchical generalized linear models (HGLM) (Goldstein, 1995; Lee and Nelder, 1996) can be used to establish the relationship between response and covariates, introducing in the linear predictor some coefficients that are random. Just as in generalized linear models (GLM) (McCullagh and Nelder, 1989), a link function relates the linear predictor to the expected value and its form is usually assumed to be known and fixed.

In some cases it might be useful to improve the modeling flexibility, allowing the link function to be a member of a class indexed by one or more unknown parameters to be estimated. While the performance of some families of link functions has been investigated for the generalized linear models (Prentice, 1976; Pregibon, 1980; Stukel, 1988; Czado, 1997), not much work has been done for hierarchical generalized linear models (Taylor et al., 1996; Oberg and Davidian, 2000).

In this work, we reanalyze a data-set related to the level of burnout among teachers of the school district of Trieste, addressing some questions regarding the sensitivity of hierarchical generalized linear models to link specification. In particular we will look to the global performance of the model within some families of link functions, possible changes in the set of statistically significant parameters for fixed and random effects and in the precision of their estimates, changes in the fit of the model, as well as problems related to comparison of institutions.

2 The case study: Assessment of level of burnout among teachers

A study aimed to assess the level of burnout among teachers belonging to the school district of Trieste was conducted. More than four hundred questionnaires were collected in thirty nine institutions.

Burnout can be defined as a type of occupational stress which involves particularly "helping professions": it occurs when subjects perceive a significant gap between expectations of successful professional performance and an observed, far less satisfying, reality (Friedman, 2000). The increasing interest in studying the burnout is due to the need of recognition, prevention and remediation of this syndrome in workplace. In the last few years, many researches have been conducted in order to measure the level of burnout in specific professional categories (medical doctors, nurses, teachers, policemen, etc.).

A lot of empirical researches linked the construct of teachers' burnout to other variables, suggesting a theoretical perspective which supports the idea that burnout is a complex phenomenon which cannot be explained, or cannot be associated only with personal characteristics or individual variables, which is the result of external systemic factors such as bureaucracy, poor administrative support and difficult working conditions (Grosch and Olsen, 1992).

In our research we assessed the level of burnout using Italian validated version of the Maslach Burnout Inventory (MBI) (Maslach, 1992), a multidimensional construct which allows the analysis of the perceived burnout in terms of three subscales: the emotional exhaustion, the depersonalization and the reduced personal accomplishment. Data (the so-called BOT data-set) were collected during a three months period; teachers were recruited using a cluster stratified sampling: at the first stage a random sample of eleven institutions was taken from each educational level (four of them refused to participate to the study), then a certain number of teachers from each institution were interviewed; the size of the sample taken within each institution depended on the number of teachers employed in it. Moreover, demographic and professional data regarding teachers were recorded (sex, age, years spent in teaching, years spent in the current institution and educational degree).

According to Rozbowsky et al. (2000), the clinical/pathological subject situation was defined as a presence of high level of emotional exhaustion and depersonalization at the same time.

The observed prevalence of burnout was higher for teachers belonging to junior schools, where about thirty percent of teachers were classified as being in a clinical status. For elementary, high school and university level the prevalence was slightly below the average (0.226). A wide heterogeneity in proportions of clinical cases of burnout was observed between institutions. Female teachers presented a higher level of burnout (0.242) than male teachers (0.185); the latter represented about a quarter of the sample. Teachers aged between 40 and 51 had an above average level of burnout, while younger and older teachers were below average and the same was reported for teacher with more than 26 years of teaching and for those which had spent less than 4 years in the current institution. The complete summary of the observed prevalence of burnout is reported in Table 1 and in Table 2.

Table 1: Observed prevalence of clinical burnout and sample size.

Level	Institution	Clinical cases	Sample size
Elementary		0.218	101
	Collodi/Giotti/Pertini (Se1)	0.125	8
	Dardi/Manna (Se2)	0.368	19
	Lovisato (Se3)	0.200	10
	Morpurgo (Se4)	0.364	11
	Saba (Se5)	0.118	17
	Sauro (Se6)	0.333	9
	Suvich (Se7)	0.111	27
Junior		0.283	99
	Addobbati (Sm1)	0.375	8
	Bergamas (Sm2)	0.000	6
	Brunner (Sm3)	0.500	6
	Codermatz (Sm4)	0.556	9
	Corsi (Sm5)	0.200	5
	Dante (Sm6)	0.200	10
	Julia (Sm7)	0.556	9
	Manzoni (Sm8)	0.182	11
	Sauro (Sm9)	0.125	8
	Stuparich (Sm10)	0.143	14
	Svevo (Sm11)	0.308	13
High		0.211	109
	Carducci (Ss1)	0.000	6
	Volta (Ss2)	0.000	6
	Carli (Ss3)	0.348	23
	Da Vinci (Ss4)	0.000	8
	Fabiani (Ss5)	0.444	9
	Deledda (Ss6)	0.000	11
	Dante (Ss7)	0.000	5
	Petrarca (Ss8)	0.375	8
	Petrarca 2 (Ss9)	0.100	10
	Galilei (Ss10)	0.400	10
	Oberdan (Ss11)	0.231	13
University		0.194	103
	Economy (F1)	0.100	10
	Pharmacy (F2)	0.100	10
	Law (F3)	0.000	10
	Engineering (F4)	0.300	10
	Literature and Philosophy (F5)	0.111	9
	Mathematics (F6)	0.200	10
	Psychology (F7)	0.273	11
	Foreign Literature and Languages (F8)	0.300	10
	Educational Science (F9)	0.250	8
	Political Sciences (F10)	0.267	15
General		0.226	412

Table 2: Prevalence of clinical burnout and sample proportions for categories of teachers.

		Clinical cases	Sample composition
Gender	Male	0.185	0.29
	Female	0.242	0.71
Age	21–40	0.151	0.28
	41–46	0.300	0.24
	47–51	0.255	0.21
	52–70	0.217	0.27
Number of teaching years	0–11	0.233	0.25
	12–20	0.250	0.31
	21–26	0.250	0.20
	27–43	0.189	0.24
Number of teaching years in current institution	0–4	0.182	0.26
	5–9	0.269	0.25
	21–26	0.242	0.24
	27–43	0.212	0.25
Degree	Bachelor	0.312	0.75
	Other	0.188	0.25

3 Statistical analysis

About one fourth of the 412 sampled subjects had at least a missing value over the 22 recorded items. Missing data were imputed using the observed median value for the specific institution for each of the missing items.

The nesting structure of data had to be taken into account; teachers came from 39 different institutions, and these could be considered as a random sample from the population of teaching institutions of the school district of Trieste. Belonging the institutions to four different levels of education, made us consider the possibility of introducing an additional level of nesting in the structure of data. Anyhow, these levels could not be considered a random sample from a population and no variable at level of education was available, which would justify the use of random effects (Snijders and Bosker, 1999). We tested the presence of systematic differences between levels of education carrying out a chi-squared test to assess the absence of heterogeneity in proportions of presence of burnout, which yielded $\chi^2_3 = 2.60$, corresponding to a p-value of 0.46, therefore not rejecting the hypothesis of homogeneity between levels of education.

Some evidence of heterogeneity was found between different institutions, using the test of heterogeneity of proportions of Commenges and Jacqmin (1994). A value of $z = 1.33$, with a p-value< 0.1 was obtained. It seemed therefore reasonable to assume a multilevel structure where the institutions represented clusters.

The estimated true variance between institutions' dependent proportions of presence of burnout is $\hat{\tau}^2 = 0.009$, corresponding to a standard deviation $\hat{\tau} = 0.095$, which is relatively large compared to the average probability of being in a clinical status of burnout, $p = 0.226$.

3.1 A multilevel logit regression model

The dichotomous variable measuring the presence or absence of clinical burnout can be considered as the result of an underlying continuous variable that measures the level of burnout. Assuming a logistic distribution for the underlying variable, a logit link was chosen and a stepwise forward procedure adopted to select fixed effects to be included in the model.

In a univariate analysis, none of the level-one available variables, namely referred to demographic and professional characteristics of teachers, resulted statistically significant at a level lower than $p = 0.22$ (corresponding to gender), when included one by one in the model. Eventually we selected a model including the fixed effects of gender and total number of years spent teaching. We decided to include non-significant fixed effects in the model since we wanted to estimate the contextual effect of institutions on the presence of clinical burnout, while controlling for the effect of gender and number of years spent teaching, which are known effects from previous studies (Capel, 1987). In fact the use of restrictive levels of rejection in the fixed effects part of the model, as 0.01 or 0.05, has been often shown to fail to identify variables that are known to be important (Mickey and Greenland, 1989).

The fitted model has thus the following form:

$$\operatorname{logit}\left(P(Y_{ij} = 1|b_i)\right) = \eta = \beta_0 + \text{Gender}_j\beta_1 + \text{TeachingYears}_j\beta_2 + b_i \qquad (3.1)$$

where $P(Y_{ij} = 1|b_i) = \pi_{ij}$ is the conditional probability of being in a clinical status of burnout for j-th subject, belonging to the i-th institution, η is the linear predictor and b_i is the random intercept, which is supposed to have zero mean and variance τ_0^2 to be estimated. The summary for the selected model is reported in Table 3.

3.2 Sensitivity analysis to link specification on the BOT data-set

The choice of a logit link function for the BOT data-set, although driven from some considerations related to the nature of the observed variable, is somehow arbitrary (McCullagh and Nelder, 1989).

Link mis-specification may affect some important issues in the study of burnout, which include changes in estimated probabilities of being in clinical status of burnout,

Table 3: Summary of the logit model including gender and number of teaching years.

Fixed Effect	Parameter	S.E.
Gender	0.173	0.144
Teaching Years	-0.003	0.013
Random Effect		
Level-two variance	0.421	0.215
Log Likelihood	-223.6719	

the set of fixed effects that can be considered to influence the level of burnout and issues related to comparison of institutions' effect.

We chose four well known parametric families of link functions and assessed their performance and the changes that they implied on the BOT data. The families that were considered are the following:

- Aranda-Ordaz Symmetric Family (Aranda-Ordaz, 1981)

$$\eta = \frac{2}{\lambda} \frac{\pi^\lambda - (1-\pi)^\lambda}{\pi^\lambda + (1-\pi)^\lambda}$$

  which reduces to the logit link when $\lambda \to 0$ and to the identity link for $\lambda = 1$. The probit link is approximated by $\lambda = 0.3955$ (Aranda-Ordaz, 1981).

- Aranda-Ordaz Asymmetric Family (Aranda-Ordaz, 1981)

$$\eta = \log\Big(\frac{(1-\pi)^{-\lambda} - 1}{\lambda}\Big)$$

  which corresponds to the logit link for $\lambda = 1$ and to the complementary log log link for $\lambda \to 0$.

- The Stukel Family (Stukel, 1988), for which only the left tail modification has been considered, because more appropriate in case of probabilities π below 0.5 ($\eta < 0$)

$$h_\lambda(\eta) = \text{logit}(\pi)$$

  with

$$h_\lambda(\eta) = \begin{cases} -\lambda^{-1}(\exp(\lambda|\eta|) - 1) & \text{if } \lambda > 0 \text{ and } \eta < 0 \\ \eta & \text{if } \lambda = 0 \text{ or } \eta > 0 \\ -\lambda^{-1}\log(1 - \lambda|\eta|) & \text{if } \lambda < 0 \text{ and } \eta < 0 \end{cases}$$

  which reduces to the logit link for $\lambda \to 0$

- Czado family (Czado, 1994)

$$h_\lambda(\eta) = \text{logit}(\pi)$$

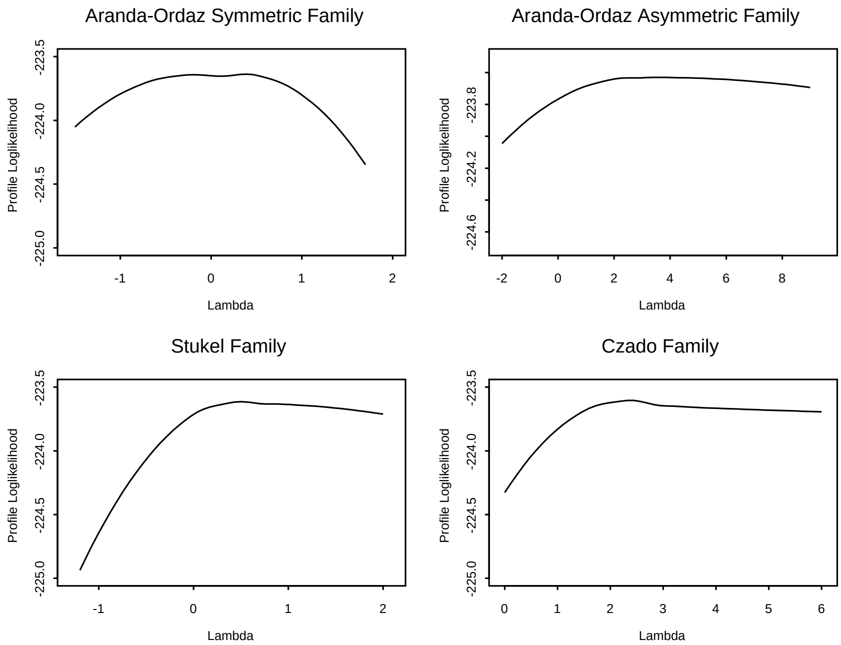

Figure 1: Profile loglikelihood for the four parametric families.

with

$$h_\lambda(\eta,\lambda) = \begin{cases} \eta & \text{if } \eta \geq 0 \\ -\frac{(-\eta+1)^\lambda - 1}{\lambda} & \text{if } \eta < 0 \end{cases}$$

which reduces to the logit link for $\lambda = 1$.

We fitted hierarchical generalized linear models with these link functions using `nlme()` function in the S-Plus package. Maximum likelihood estimates (MLE) were obtained computing MLE for all the components of the model except for the link parameter λ, using a fixed link function $F(.,\lambda_0)$ for an array of values of λ_0. The joint MLE was extracted from the loglikelihood profile plot for the link, as justified in Czado (1997).

Fixed effect of gender and number of years spent teaching were included in all considered models. Institutions were used as cluster.

Obtained profile log likelihoods resulted rather flat (Figure 1); except for Aranda-Ordaz Symmetric Family, profile log likelihood tended to remain constant after having obtained their maximum value, for increasing values of λ. Therefore little information was given about the optimal value of λ for these families and many values could be considered equally consistent. For Aranda-Ordaz Symmetric family, even though the profile log likelihood remained quite flat in the neighborhood of its maximum, a less wide range of compatible values for the link parameter λ was found.

Table 4: Maximum likelihood link estimates, loglikelihoods and LR Statistics (with respect to the logit model) for the BOT data.

ML Model	$\hat{\lambda}$	Loglikelihood	LRatio(p-value)
logit		-223.6719	
Aranda-Ordaz Symmetric	0.6	-223.6546	0.003 (0.86)
Aranda-Ordaz Asymmetric	3.6	-223.6324	0.078 (0.78)
Stukel	0.4	-223.6377	0.068 (0.79)
Czado	1.8	-223.6316	0.080 (0.78)

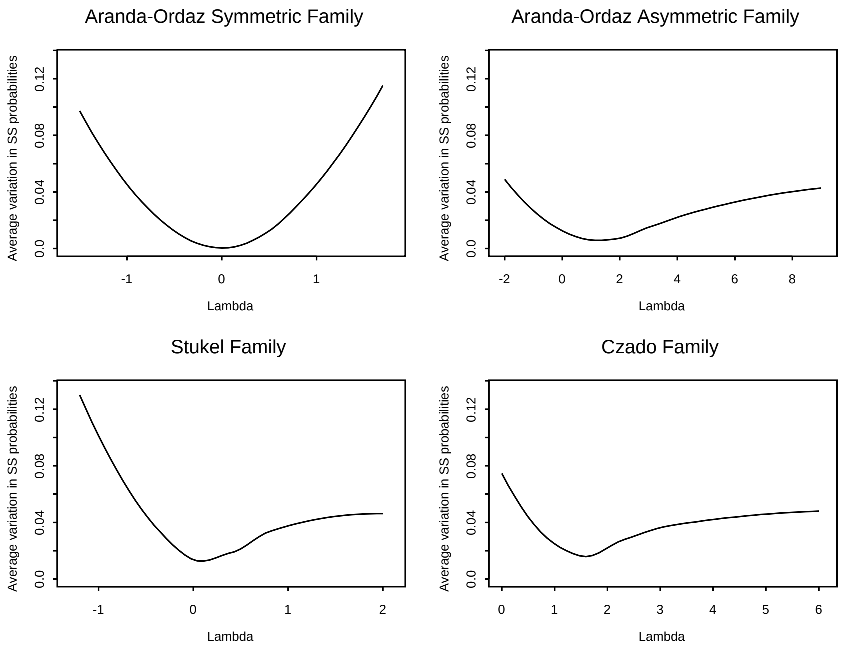

Figure 2: Mean percentage variation in estimated Subject Specific probabilities for the four parametric families with respect to the logit model.

A likelihood ratio test was performed to compare the model fitted with the logit link with models obtained from the four selected families: for all families the differences in likelihood ratio test strongly supported the null hypothesis of no difference with the logit model, therefore giving no evidence of link mis-specification (Table 4). In any case, the differences in the maximum values of all the profile log likelihoods were extremely small, indicating that the choice of the link function does not play an important role in this context.

The variations in estimated probabilities (both Population Averaged and Subject Specific) of being in a clinical status of burnout with respect to the logit model

Table 5: Mean and maximum percentage variations in estimated probabilities between logit model and maximum likelihood models for the four parametric families.

	Population Averaged		Subject Specific	
ML Model	Mean Δ p	Max Δ p	Mean Δ p	Max Δ p
Aranda-Ordaz Symmetric	3.49	9.68	4.04	10.97
Aranda-Ordaz Asymmetric	2.09	4.45	2.34	5.11
Stukel	3.91	9.36	4.52	11.83
Czado	3.29	8.49	3.68	10.20

were considered, for a range of values for the link parameter (Figure 2) and for the models corresponding to the MLE of λ (Table 5). It can be observed that on average the estimated probabilities from the ML models of the four families do not differ substantially from those estimated with the logit model (for Population Averaged probabilities, between 2% and 4% of mean percentage variation and about 10% of maximum percentage variation), obtaining therefore an estimated model with a fit quite similar to that obtained with the logit model. It can be noted that Subject Specific estimates of probabilities are always above the respective Population Averaged ones.

Point estimates of fixed and random effects and their precision are deemed to differ from those obtained from the logit model due to different hypothesis on the distribution of the underlying unobservable variable that are implied from the choice of the link function. The estimates for the fixed effects were substantially less precise for all families except for Stukel; their confidence intervals were about 1.3 times wider for the ML model obtained from Aranda-Ordaz Symmetric Family, twice wider for Aranda-Ordaz Asymmetric Family family, and about three times wider for Czado family. A graphical representation of the estimates for gender effect with its confidence intervals is given in Figure 3 for a range of values of λ. In Figure 4 the same is done for the level-two variance.

A problem that is often of interest in practice is related to comparison of institutions; in particular, in this framework, we are interested in assessing whether there are any institutions that have a significantly different impact from other institutions on the level of burnout of their teachers, once that their demographic and professional characteristics have been taken into account.

It is of common use in psychological research to compare estimated random effects to assess a ranking between the clusters or to assess if there is evidence of any significant difference between them.

The crude comparison of the estimated random effects cannot give any evidence of differences among institutions. Here, two criteria to test for the differences between institutions are considered; according to the first, two institutions have a

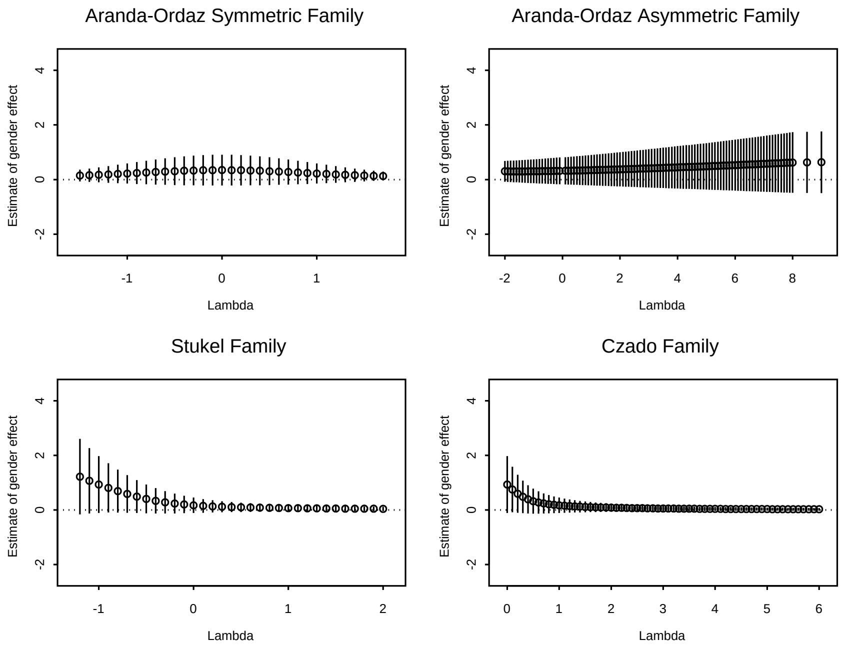

Figure 3: Gender effect (95% confidence intervals) for the four parametric families of links.

significantly different impact on the burnout of their teachers if the confidence intervals of their estimated random effects do not overlap. The width of the confidence intervals is adjusted as suggested by Goldstein and Healy Goldstein and Healy (1995) to take multiple comparisons into account. The second approach, less restrictive, considers two institutions to perform differently if the point estimate of the random effect does not fall into the confidence interval of the other institution. Institutions were compared within the level of education to which they belonged, not being sensible a comparison among institutions belonging to different levels.

Among elementary schools, using the first criterion and the logit link function, a school significantly different from a group of four schools, that have the lowest impact on the burnout of teachers, can be identified. Another group of two schools can be isolated, which has a significantly worse impact than the best two. Using the less stringent criterion two significantly different groups of four and three schools, ranked at the extremes, are identifiable (Table 6). Similar conclusions can be drawn for junior schools (Table 7), where the groups of significantly different institutions become wider using the second criterion. At the high school level (Table 8) three groups of schools can be considered significantly different in their effect on the level of burnout. Very little evidence of significant differences among university faculties was found (Table 9).

Some differences in the ranking order was found fitting models with parametric

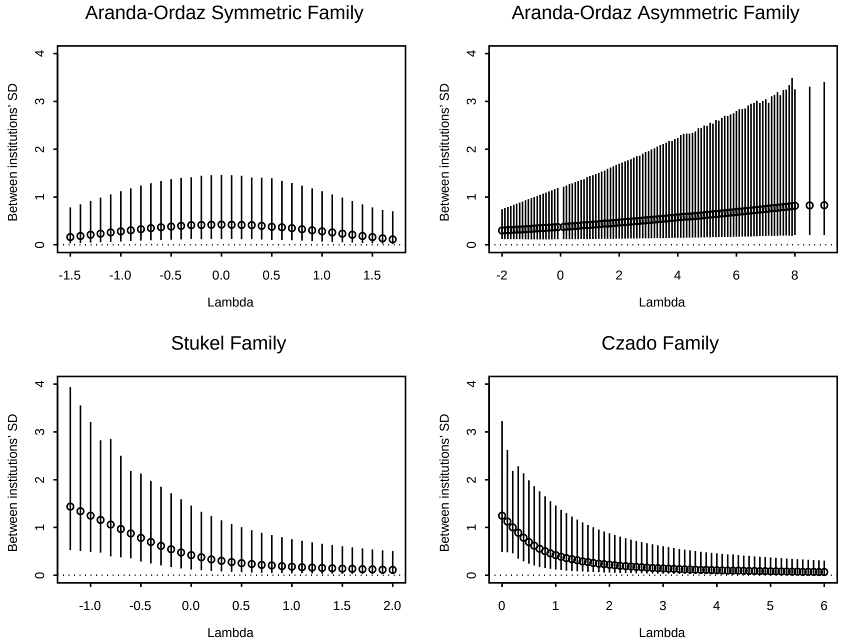

Figure 4: Between-institutions' standard deviation (95% confidence intervals) for the four parametric families of links.

Table 6: Institutional comparison among elementary schools. Significant differences between institutions according to first ($\star\star$) and second criterion ($\star$) for all maximum likelihood models (see Section 3.2, page 8); $\star_{-C}$: satisfied for all models except for Czado family; $\star_C$: satisfied only for Czado family.

Institution	Se7	Se5	Se1	Se3	Se6	Se4	Se2
Se7				$\star$	$\star\star$	$\star\star$	$\star\star$
Se5				$\star_{-C}$	$\star\star$	$\star\star$	$\star\star$
Se1					$\star$	$\star$	$\star\star$
Se3					$\star_C$	$\star$	$\star\star$
Se6							$\star_{-C}$
Se4							
Se2							

families of links. The results obtained with Aranda-Ordaz and Stukel families did not show much differences from the original ranking obtained through the logit model. For the model fitted using the Czado family some results differed from those obtained using the other link families. Some of the institutions that were not found to be significantly different with the other link functions satisfied both

Table 7: Institutional comparison among junior schools. Significant differences between institutions according to first ($\star\star$) and second criterion ($\star$) for all maximum likelihood models (see Section 3.2, page 8); $\star_{-C}$: satisfied for all models except for Czado family; $\star_C$: satisfied only for Czado family; $\star\star_{AS,C}$: satisfied only for Czado and Aranda-Ordaz Symmetric Family.

Institution	Sm 2	Sm 10	Sm 9	Sm 8	Sm 6	Sm 5	Sm 11	Sm 1	Sm 3	Sm 7	Sm 4
Sm2							$\star$	$\star$	$\star\star_C\star$	$\star\star$	$\star\star$
Sm10							$\star\star_C\star$	$\star\star_{AS,C}\star$	$\star\star$	$\star\star$	$\star\star$
Sm9							$\star$	$\star$	$\star\star_C\star$	$\star\star$	$\star\star$
Sm8							$\star$	$\star$	$\star\star_C\star$	$\star\star$	$\star\star$
Sm6							$\star_{-C}$	$\star$	$\star$	$\star\star$	$\star\star$
Sm5								$\star_C$	$\star$	$\star$	$\star\star$
Sm11									$\star_C$	$\star\star_C\star$	$\star\star_C\star$
Sm1										$\star$	$\star\star_C\star$
Sm3											$\star$
Sm7											$\star_C$
Sm4											

Table 8: Institutional comparison among high schools. Significant differences between institutions according to first ($\star\star$) and second criterion ($\star$) for all maximum likelihood models (see Section 3.2, page 8); $\star_{-C}$: satisfied for all models except for Czado family; $\star_C$: satisfied only for Czado family; $\star\star_{AS,C}$: satisfied only for Czado and Aranda-Ordaz Symmetric Family.

Institution	Ss 6	Ss 4	Ss 1	Ss 2	Ss 7	Ss 9	Ss 11	Ss 8	Ss 10	Ss 5	Ss 3
Ss6							$\star$	$\star\star$	$\star\star$	$\star\star$	$\star\star$
Ss4							$\star_{-C}$	$\star\star$	$\star\star$	$\star\star$	$\star\star$
Ss1							$\star_{-C}$	$\star\star_C\star$	$\star\star_{AS,C}\star$	$\star\star_{-C}\star$	$\star\star$
Ss2							$\star_{-C}$	$\star\star_C\star$	$\star$	$\star\star_{-C}\star$	$\star\star$
Ss7							$\star_{-C}$	$\star\star_C\star$	$\star\star_C\star$	$\star\star$	$\star\star$
Ss9							$\star_{-C}$	$\star\star_C\star$	$\star\star_{-C}\star$	$\star\star_{-C}\star$	$\star\star$
Ss11								$\star\star_C$	$\star$	$\star$	$\star\star_C\star$
Ss8											$\star_C$
Ss10											$\star\star_C$
Ss5											$\star_C$
Ss3											

Table 9: Ranking comparison among university faculties. Institutional comparison among junior schools. Significant differences between institutions according to first ($\star\star$) and second criterion ($\star$) for all maximum likelihood models (see Section 3.2, page 8); $\star_{-C}$: satisfied for all models except for Czado family; $\star_{C}$: satisfied only for Czado family.

Institution	F3	F1	F2	F5	F6	F9	F7	F8	F10	F4
F3					$\star_{-C}$	$\star$	$\star$	$\star$	$\star\star$	$\star\star_{-C}\star$
F1						$\star_{-C}$	$\star$	$\star$	$\star\star_{C}\star$	$\star$
F2						$\star_{-C}$	$\star$	$\star$	$\star\star_{C}\star$	$\star$
F5							$\star$	$\star$	$\star\star_{C}\star$	$\star$
F6							$\star_{C}$	$\star_{C}$	$\star$	$\star$
F9									$\star_{C}$	
F7										
F8										
F1										
F4										

proposed criteria for the model fitted with the Czado link (Tables 7-9), letting the analyst drawing different conclusions about the influence of institutions on the level of burnout.

4 Final Remarks

In the framework of generalized linear models and of hierarchical generalized linear models, the link function plays an important role, specifying the link between the random and systematic components of the model, i.e. between the expected value and the linear predictor. The choice of the link function has consequences on the fit of the model as well as on the interpretation of parameters included in the model, and is usually made without having much information about its appropriateness for the data that are being analyzed. In fact, in practice usually only the link functions for which computer software is available are considered, and the choice between models fitted with different link functions is made using some goodness of fit test. Canonical links are frequently preferred, for the simplicity of calculations implied in the estimation and because they lead desirable statistical properties for the model (McCullagh and Nelder, 1989).

For binary data it was shown that mis-specification of the link function leads to bias and decreased precision of estimates of success probability and regression parameters (Czado, 1989). Therefore in some cases it might be useful to let the link function to be chosen estimating it from data. The most common approach allows the link to be a member of a parametric class indexed by one or more unknown parameters, that are estimated in the same fashion as the other unknown parameters of the model.

Issues related to estimation of the link function have been widely analyzed in the GLM framework and in particular for binary data, where a number of parametric families have been proposed, which embed the logit and other well known link functions, see for instance (Prentice, 1976; Pregibon, 1980; Stukel, 1988; Czado, 1997). In non-linear random effects some attention has been very recently devoted to analyzing suitable transformation of the response variables separately (Taylor et al., 1996) or not (Oberg and Davidian, 2000) from the transformations of the predictors and, as far as we know, no investigation has been carried out on the choice of link function in such models. Nevertheless, the effects of a mis-specified link function are, at least in principle, not trivial. A link function is commonly associated with a specific underlying distribution: mis-specifying it has a consequence an over (under) estimation of the variance or the tail distribution.

While estimating the form of the link function usually improves the fit of the model when compared to canonical links, there are some drawbacks associated with it. Data might be over-fitted, leading to flat likelihoods and numerical problems in the estimating procedure. Letting the estimate of the link function be data-driven within a parametric family of link functions, implies difficulties in the interpretation of parameters and therefore in understanding the influence of covariates. Moreover, estimation of the link has been shown to increase the variance of the regression parameters and predicted probabilities (Czado, 1997), unless link and regression parameters are not correlated, orthogonal in the sense of Cox and Reid (1987). Additional problems arise in the HGLM framework, where orthogonality should hold also for the parameters related to random effects that are present in the model and might be difficult to obtain.

A problem that has been established in GLM framework, is that p-values associated with commonly used goodness of fit tests do not quantify appropriately changes in the fit of the model and in some specific quantities, that might be of interest for the experimenter (Czado and Munk, 2000). Therefore, it could be interesting to determine which are the differences that arise when fitting a model with a non-canonical link function instead of a using a fixed link function, looking beyond standard tests to compare models.

From the point of view of burnout research, there is a high interest in analyzing burnout in terms of effects of contextual and individual variables on subject level. All aspects of explaining and interpreting such things are in fact affected from link function choice:

- the set of individual (fixed effects) that influence level of burnout;
- prediction of clinical burnout given personal and institutional characteristics;
- evaluation of institutional effects on burnout and their comparison.

Fitting models selecting link functions that differed from our initial choice of a logit link, implied different hypothesis on the distribution of level-one residuals, that were not assumed to have a logistic standardized distribution any longer. Modifying the hypothesis on the underlying distribution of the unobserved response variable implies that point estimates of fixed effects included in the model change, as well as variance parameters of the random part. All the estimated models supported the results of non significance of the fixed effects of gender and number of years spent teaching, as well as the presence of significant difference between institutions.

Estimated probabilities obtained from the models fitted using the parametric links did not differ substantially from those of the logit model. The average variation was between 2% and 4% for Population Averaged probabilities, and just slightly higher for Subject Specific ones. The maximum variation of about 10% was obtained with the Aranda-Ordaz Symmetric family for Population Averaged probabilities and of about 12% with Stukel family for Subject Specific ones. Therefore from the point of view of the fit of the model we can conclude that differences are not substantial.

One of the families of link functions gave results that were significantly different from those obtained with the other families when it came to compare the effect of institutions on the level of burnout of their teachers. It has to be recalled that models obtained using different link functions were not significantly different according to the likelihood ratio test. It is therefore important to notice the sensitivity of procedures related to institutional comparison to link specification, *i.e.* that different conclusions can be drawn when comparing institutions using a family of link functions instead of another even if the fitted models do not seem to differ significantly when compared using a testing procedure.

Theoretical work needs to be done in the HGLM framework to establish asymptotic properties of estimates obtained using a parametric link function and to determine appropriate testing procedures. A numerical estimating procedure that allows the joint estimation of all the unknown parameters of the model would be needed.

References

[1] Aranda-Ordaz, F.J. (1981): On two families of transformations to additivity for binary response data. *Biometrika*, **68**, 357–364.

[2] Capel, S.A. (1987): The incidence of and influences on stress and burnout in secondary school teachers. *British Journal of Educational Psychology*, **57**, 279–288.

[3] Commenges, D. and Jacqmin, H. (1994): The intraclass correlation coefficient: distribution-free definition and test. *Biometrics*, **50**, 517–526.

[4] Cox, D.R. and Reid, N. (1987): Parameter orthogonality and approximate conditional inference. *Journal of Royal Statistical Society, Series B*, **49**, 1–39.

[5] Czado, C. (1989): *Link Misspecification and Data Selected Transformations in Binary Regression Models*. School of Operarions Research and Industrial Engineering, Cornell University, Ithaca, NY.

[6] Czado, C. (1994): Parametric link modification of both tails in binary regression. *Statistical Papers*, **35**, 189–201.

[7] Czado, C. (1997): On selecting parametric link transformation families in generalized linear models. *Journal of Statistical Planning and Inference*, **61**, 125–139.

[8] Czado, C. and Munk, A. (2000): Noncanonical links in generalized linear models. When is the effort justified? Available at `http://www.math.yorku.ca/Who/Faculty/Czado/menu.html`.

[9] Friedman,I. A. (2000): Burnout in teachers: shattered dreams of impeccable professional performance. *Journal of Clinical Psychology*, **56**, 595–606.

[10] Goldstein, H. and Healy, M.J.R. (1995): The graphical presentation of a collection of means. *Journal of the Royal Statistical Society, Series A*, **158**, 175–177.

[11] Goldstein, H. (1995): *Multilevel Statistical Models*. London: Edward Arnold.

[12] Grosch, W.N. and Olsen, D. C. (2000): Clergy burnout: An integrative approach. *Journal of Clinical Psychology*, **56**, 619–632.

[13] Lee, Y and Nelder, J.A. (1996): Hierarchical generalised linear models (with discussion). *Journal of the Royal Statistical Society (Series B)*, **58**, 619-678.

[14] Liang, K.Y. and Zeger, S.L. (1986): Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–20.

[15] Maslach, C. (1992): *Burnout: The Cost of Caring*. New Jersey: Prentice Hall Press.

[16] McCullagh, P. and Nelder, J.A. (1989): *Generalized Linear Models*. London: Chapmann and Hall.

[17] Mickey, J. and Greenland, S. (1989): A study of the impact of confounder-selection criteria on effect estimation, *American Journal of Epidemiology*, **129**, 125–137.

[18] Oberg, A. and Davidian, M. (2000): Estimating data transformations in nonlinear mixed effects models. *Biometrics*, **56**, 65–72.

[19] Pregibon, D. (1980): Goodness of link tests for generalized linear models. *Applied Statistics*, **29**, 15–24.

[20] Prentice, R.L. (1976): Generalization of the probit and logit methods for dose response curves. *Biometrics*, **32**, 761–768.

[21] Rozbowsky, P., Semeraro, A., Cervai, S., Gregori, D., and Gabassi, G. (2000): The risk of burnout: a dynamic phase model. *Risk Decision and Policy* (in press).

[22] Searle, S.R., Casella, G., and McCulloch, C.E. (1992): *Variance Components*. New York: Wiley.

[23] Snijders, T. and Bosker, R. (1999): *Multilevel Analysis*. London: Sage.

[24] Stukel, T.A. (1988): Generalized Logistic Models. *Journal of American Statistical Association*, **83**, 426–431.

[25] Taylor, J.M.G., Cumberland, W.G., and Meng, X. (1996): Components of variance models with transformations. *Australian Journal of Statistics*, **38**, 183–191.