

Multilevel Models and Robustness

Modelli Multilivello e Robustezza

Nadia Solaro¹, Silvia Salini², Pier Alda Ferrari²

¹ Department of Statistics, University of Milan-Bicocca

e-mail: nadia.solaro@unimib.it

² Department of Economics Business and Statistics, University of Milan

Riassunto: In questo lavoro il binomio robustezza–modelli multilivello viene affrontato secondo due diverse accezioni: la robustezza nei modelli multilivello e i modelli multilivello come procedura robusta. Nel primo caso vengono discusse le proprietà statistiche degli usuali stimatori dei parametri dei modelli multilivello quando è violata l'ipotesi di normalità delle componenti erratiche. Nel secondo, il modello multilivello è utilizzato come approccio robusto al problema della taratura statistica in presenza di dati raggruppati. Lo stesso viene poi comparato con modelli alternativi mediante uno studio di simulazione.

Keywords: ML and REML estimates, multilevel calibration estimator, multivariate calibration, multivariate exponential power distributions, sandwich estimator.

1. Introduction

In classical inference when a statistical model is adopted some specific assumptions are imposed. As it is well-known, they may concern the linearity of the model, normal distribution of the variables, independence of the observations and so on. Anyway, what may occur in many practical situations is that these assumptions are not completely satisfied: for instance, the distribution is different from the one assumed, some outliers are present, data belong to more than one group. This kind of misspecifications or perturbations in data may lead to relevant consequences on inference. It is then important to investigate the behaviour of the classical statistical procedures when the assumptions are not, or are only approximately, true. It is also advisable to set up methods which avoid erroneous conclusions in situations like those above mentioned.

With this purpose many studies have been carried out and new statistics and procedures that could be little affected by departures from assumptions have been developed. These statistics and procedures are usually referred to as “robust”.

In the literature different approaches towards the robustness problem are present. In particular, Huber (1981) and Hampel (Hampel *et al.*, 1986) capture the problem in a mathematical framework by defining some new concepts in order to describe the behaviour of statistics in the neighborhood of the assumed model. Specifically, the influence function is introduced and some “robustness measures” such as the gross error sensitivity and the local shift sensitivity are proposed in addition to the classical concept of efficiency. These new proposals represent now an organic set of tools that allow statisticians to compare different methods and to choose the one which is more convenient to adopt.

Another approach to robustness, which also addresses a more proper data analysis

framework, is oriented to develop statistical procedures that first detect patterns in multidimensional data and then fit the majority of them. Here the principal goal is to identify either outlying units or substructures in the data, such as the existence of groups, and their effects on fitted models and inferences. This methodology, which has been developed quite recently, is known as Forward Search (Atkinson and Riani, 2000; Atkinson, Riani and Cerioli, 2004).

Taking this into account, in this paper we focus on linear multilevel models, which describe the linear relation of a dependent variable on a set of explanatory variables when data are organized in groups according to a hierarchical structure (Goldstein, 1995; Snijders and Bosker, 1999). In particular, we treat the connection “robustness” and “multilevel models” according to a dual direction of analysis: the robustness in multilevel models and multilevel models for robustness.

Our paper is divided into two main parts. In the first one we examine the consequences on maximum likelihood estimators for multilevel model parameters when the assumption of normality of the random effects distribution is violated. The misspecification considered here is generated by the family of Multivariate Exponential Power distributions. What may be perceived is that maximum likelihood estimators for fixed effects and level-1 variance are not affected by MEP distributed random effects, while the most relevant impact are on estimators for level-2 variance components. For instance, the variability of these estimators strictly depends on the type of MEP assumed. Furthermore, model-based standard errors are far from being reliable estimates, so that corrections for them are absolutely necessary. In this context we consider the type of correction proposed by Verbake and Lesaffre (1997) in linear mixed-effects models for longitudinal data.

In the second part we deal with the problem of calibration when data are grouped. In this context we suggest using multilevel models. In fact, the goal is the detection of a significant group effect and then the development of a calibration estimator which is still reliable when units are clustered. In order to compare this proposal with other approaches, a simulation study is carried out under some experimental conditions, which are defined by combinations of a range of number of groups and group sizes.

The paper is organized as follows: in Section 2 the definition of linear multilevel models for a two-level hierarchical structure is introduced; in Section 3 a brief discussion on the meaning of robustness in multilevel model framework is reported; Section 4 is given over to a presentation on the behaviour of maximum likelihood estimators for the multilevel model parameters when random effects are distributed according to a member in the family of Multivariate Exponential Power distributions; in Section 5 a possible correction for the standard errors is illustrated; Section 6 deals with the linear multivariate calibration model and recalls some contributes on the robustness in calibration; in Section 7 the linear multilevel model is specified in the calibration framework. A first simulation study is then carried out. Finally, Section 8 is devoted to conclusions and future endeavours.

2. Linear Multilevel Models

Whenever a linear multilevel model has to be built for a two-level hierarchical data structure, consisting of J groups (level-2 units) and n_j units within group (level-1 units),

Laird and Ware's matrix formulation may be applied (Laird and Ware, 1982):

$$\mathbf{Y}_j = \mathbf{Z}_j\boldsymbol{\gamma} + \mathbf{X}_j\mathbf{U}_j + \boldsymbol{\varepsilon}_j, \quad (1)$$

where $\mathbf{Y}_j$ is a r.v. (random vector) of quantitative Y_{ij} variables; $\mathbf{Z}_j$ and $\mathbf{X}_j$ are the so called design matrices; $\boldsymbol{\gamma}$ is a vector of p fixed effects; $\mathbf{U}_j$ is a r.v. of q level-2 errors, also termed random effects; finally, $\boldsymbol{\varepsilon}_j$ is a r.v. of level-1 errors, ($i = 1, \dots, n_j; j = 1, \dots, J$). The errors at each level have zero mean. Their variance-covariance matrices are given, respectively, by: $\mathbf{V}(\mathbf{U}_j) = \mathbf{T} = [\tau_{hm}]_{h,m=0,\dots,q-1}$ and $\mathbf{V}(\boldsymbol{\varepsilon}_j) = \sigma^2\mathbf{I}_{(n_j)}$; moreover, $\text{Cov}(\mathbf{U}_j, \boldsymbol{\varepsilon}_j) = \mathbf{0}$ for all j . The elements τ_{hm} of $\mathbf{T}$ and σ^2 are, respectively, level-2 and level-1 variance components. From these assumptions on the errors it follows for the conditional model: $\mathbf{E}(\mathbf{Y}_j|\mathbf{U}_j = \mathbf{u}_j) = \mathbf{Z}_j\boldsymbol{\gamma} + \mathbf{X}_j\mathbf{u}_j$ and $\mathbf{V}(\mathbf{Y}_j|\mathbf{U}_j = \mathbf{u}_j) = \sigma^2\mathbf{I}_{(n_j)}$, whereas for the marginal model: $\mathbf{E}(\mathbf{Y}_j) = \mathbf{Z}_j\boldsymbol{\gamma}$ and $\mathbf{V}(\mathbf{Y}_j) = \mathbf{X}_j\mathbf{T}\mathbf{X}_j^t + \sigma^2\mathbf{I}_{(n_j)}$, for all j .

In order to estimate the unknown parameters, namely the fixed effects and the variance components, two main approaches are well-known in literature: the maximum likelihood method (full -ML- and restricted maximum likelihood -REML-; see e.g. Pinheiro and Bates, 2000) and suitable extensions of the generalized least squares (iterative generalized least squares -IGLS- and restricted IGLS -RIGLS-; see Goldstein, 1995). All these methods involve iterative numerical procedures, thus the estimators cannot be expressed in a closed form.

As far as the maximum likelihood approach is concerned, the normality of the error distributions at each level is introduced, so as to obtain the so called Gaussian multilevel model. Specifically, the assumptions are: $\mathbf{U}_j \sim N_q(\mathbf{0}, \mathbf{T})$, $\boldsymbol{\varepsilon}_j \sim N_{n_j}(\mathbf{0}, \sigma^2\mathbf{I}_{(n_j)})$, with $\mathbf{U}_j$ and $\boldsymbol{\varepsilon}_j$ independent ($j = 1, \dots, J$). Goldstein (1995) argued that under normality of all the error terms IGLS/RIGLS estimates coincide with ML/REML ones. Moreover, under the Gaussian multilevel model maximum likelihood estimators for, respectively, fixed effects and variance components are asymptotically uncorrelated and asymptotically normally distributed, with the variance-covariance matrix represented by the inverse of Fisher information matrix (see e.g. Pinheiro and Bates, 2000).

3. Robustness issues in multilevel models

In the current acceptation talking about robustness in multilevel models means studying the performance of the estimators in presence of either outliers or misspecifications of the model error distributions. Broadly speaking, it might also stand for verifying whether the asymptotic properties of the estimators remain valid for finite sample sizes. In this sense, the term "robustness" is used with the meaning of "resistance" to the presence of, respectively, outliers, non normal data or small size samples.

Among the studies given over to these aspects, it is worth recalling the paper by Langford and Lewis (1998), in which the problem of outliers in multilevel data is treated and a variety of procedures for dealing with this kind of units is proposed. As for the influence of misspecifications of the error distributions, the papers by, respectively, Verbeke and Lesaffre (1997) and Maas and Hox (2004) are examples of studies tackling the problem of misspecifications related to the random effects distribution and their impact on the performance of the estimators. What was revealed by these studies is that non normal random effects have little influence on the behaviour of the estimators for fixed effects and level-1 variance. On the contrary, level-2 variance component estimators have

very inaccurate standard errors which should be then corrected by robust procedures. A quite different approach is proposed by Pinheiro, Liu and Wu (1997), who consider robust estimation in linear mixed-effects models by replacing the normal distribution on random effects with the multivariate t distribution. In such a way, a robust version of the model is introduced that is far less sensitive to outlying units. Finally, as regards the framework of finite samples, Kreft's work (1996) represents an interesting review, as Section 1 in Maas and Hox (2004).

As regards the other acceptations of the term robustness, in literature there are not yet systematic theoretic contributions in which robust estimation methods for multilevel models are developed, such as the M-estimator or the Least Median of Squares estimator for the parameters of linear regression models. Among the possible reasons, it might be argued that multilevel models do offer an intrinsic great flexibility by which some kinds of perturbations in the data may be modelled. For instance, as far as the problem of outlying observations is concerned, Langford and Lewis (1998, p. 150) point out that "... it is better to leave data complete, and to treat outliers by fitting separate effects for them rather than omitting units from the model altogether... It is also in part motivated by the complexity of the models, where an outlying unit at one level may cease to be so after treatment of units at other levels in the model".

Of course, how this flexibility could allow us to deal with perturbations in data strictly depends on the very nature of perturbations. In fact, when perturbations may be ascribed to model error terms which are not normally distributed, even roughly, the solution is unlikely to be found in a particular definition of the model, such as the introduction of additional effects, in the sense just indicated, or transforming data. Hence, it seems advisable to study firstly the behaviour of the estimators obtained by applying the standard results based on the normal assumption but in presence of misspecifications of the error distributions. Secondly, where the estimators fail the expected performance, it seems necessary to develop suitable corrections to fix the situation.

4. Multilevel models with MEP distributed random effects

Our work is a part of the studies given over to the problem of misspecifications of random effect distributions. In a series of papers (Ferrari and Solaro, 2002; Solaro and Ferrari, 2006a, 2006b) a more general distributional assumption on the random effects than the normal has been assumed. This is represented by the family of the Multivariate Exponential Power (MEP) distributions (Gómez *et al.*, 1998), while for the level-1 errors the normal assumption has been still assumed.

The family of MEPs belongs to the class of symmetric, elliptical multivariate Kotz type distributions (Fang, Kotz and Ng, 1990) and it also includes the multivariate normal as a special case. By specifying the probability density function (p.d.f.) of MEPs in the case of random effects, in model (1) we have for the vectors $\mathbf{U}_j$:

$$f(\mathbf{u}_j; \mathbf{\Phi}, \kappa) = \frac{q\Gamma(q/2) \exp \left\{ -\frac{1}{2} [\mathbf{u}_j^t \mathbf{\Phi}^{-1} \mathbf{u}_j]^{\frac{\kappa}{2}} \right\}}{\pi^{q/2} \Gamma \left(1 + \frac{q}{\kappa} \right) 2^{1+\frac{q}{\kappa}} |\mathbf{\Phi}|^{1/2}}, \quad (2)$$

for all j , which we will shortly indicate with: $\mathbf{U}_j \sim \text{MEP}_q(\mathbf{0}, \mathbf{\Phi}, \kappa)$. Specifically, the parameter κ , always positive, is called the non-normality parameter, since depending on

its value either leptokurtic ($\kappa < 2$) or platikurtic ($\kappa > 2$) distributions of random effects may be obtained, whereas the multivariate normal is attained for $\kappa = 2$. Furthermore, the matrix Φ is positive-definite and it links up with the variance-covariance matrix of $\mathbf{U}_j$ by the relation: $V(\mathbf{U}_j) = c(\kappa, q)\Phi$, being $c(\kappa, q) = 2^{2/\kappa}\Gamma((q+2)/\kappa)/(q\Gamma(q/\kappa))$, with $\Gamma(\cdot)$ denoting the gamma function.

Investigations relating to the influence of this kind of misspecification on the performance of the ML and REML estimators have been undertaken through simulation studies. In fact, it is pointed out that the likelihood function with MEP random effects is defined by such a complex formulation that any analytical treatment in closed form is ruled out, with the exception of $\kappa = 2$ (Ferrari and Solaro, 2002).

In our works the reference model is a random intercept-and-slope two-level model including one explanatory variable at each level and a cross-level interaction term. Furthermore, our simulation settings are defined by combining a range of experimental conditions which concern the group number, the group size and the variance-covariance structure of the random effects. Then, under these experimental conditions pseudo-observations y_j are obtained by random variates generation from MEP random effects and from normal level-1 errors. All estimates are determined by means of the S-Plus library nlme3 by Pinheiro and Bates (2000).

All REML and ML estimates are obtained under leptokurtic as well as platikurtic MEP distributed random effects. The normal case is also considered. The performance of the corresponding estimators is investigated with respect to both parameter typologies (i.e. fixed effects, level-2 and level-1 variance components), and a single parameter at a time. Then their main properties, such as bias, variability and skewness, are focused on. The significance of the effects considered, namely MEP, group number, group size, variance-covariance structure of random effects, is evaluated by means of Friedman's and Page's nonparametric tests. The main results up to now achieved (Solaro and Ferrari, 2006a) as well as some new recent ones are summed up in Table 1.

Table 1: *Synoptic table reporting the significant effects of our experimental factors (NG: $J = 5, 10, 20, 50$ and GS: $n = 5, 10, 15, 20, 25, 30$) or of MEP-type departures ($\kappa = 1, 1.5, 2, 8, 14$) on REML and ML estimators. The symbol $\downarrow$ intends that values of the statistics tend to reduce as the values of experimental factors or of κ parameter increase.*

	Fixed effects	Level-2 variance components	Level-1 variance
Bias	—	NG $\downarrow$	NG $\downarrow$; GS $\downarrow$
Variability	NG $\downarrow$	NG $\downarrow$; GS $\downarrow$; MEP $\downarrow$	NG $\downarrow$; GS $\downarrow$
Skewness	NG $\downarrow$	NG $\downarrow$; MEP $\downarrow$	NG $\downarrow$; GS $\downarrow$

One of the main features that it is worth pointing out concerns the significant influence that MEP distributed random effects bear on the variability of both REML and ML estimators for level-2 variance components, which is synthetized by the Monte Carlo (MC) standard errors. Specifically, as for both REML and ML estimators, greater MC standard errors are observed under leptokurtic MEPs than those obtained under normal random effects, whereas smaller MC standard errors are achieved under platikurtic MEPs than under normality.

This regular tendency has motivated a more in-depth inspection (Solaro and Ferrari, 2006b), in which the performance of the model-based standard errors, based on the inverse of the empirical Fisher information matrix, and the MC standard errors are compared. Another interesting feature has been then brought to light, namely a systematic underestimation of the *true* variability of these estimators in the presence of leptokurtic MEP random effects and, viceversa, a systematic overestimation of their true variability in the presence of platikurtic MEP random effects.

Under- or over-estimating the standard errors lead obviously to wrong inferences. For instance, nominal coverages of confidence intervals are no longer guaranteed. Hence, conceiving corrections for the model-based standard errors seems absolutely necessary.

5. Correcting the model-based standard errors for level-2 variance components

In the literature a type of correction often recommended for the standard errors is represented by the so-called “sandwich estimator”, in the Huber-White sense. Let $V(\hat{\Theta}) = B$ be the model-based variance-covariance matrix of the estimator $\hat{\Theta}$ for a vector θ of parameters. Then, the sandwich-type correction produces the new variance-covariance matrix: $V^{sw}(\hat{\Theta}) = BCB$, where C is a matrix which operates the correction. There are different ways to define C ; see, e.g., Goldstein’s approach in the framework of IGLS and RIGLS estimation methods (Goldstein, 1995). In Verbeke and Lesaffre (1997) the proposed corrections rely on standard results concerning maximum likelihood estimators that are suitably adapted by the authors to linear mixed-effects models for longitudinal data.

Here we focus on this latest type of correction, since it is based upon the maximum likelihood approach, which is the one just implemented in S-Plus. Let $\psi_j(y_j; \gamma, \sigma^2, T)$ denote the p.d.f. of y_j in the marginal model, for all j . The log-likelihood function, for the model (1) over all the J groups and under normality of both random effects and level-1 errors, is given by: $l(\gamma, \sigma^2, T|y) = \sum_{j=1}^J \ln \psi_j(y_j; \gamma, \sigma^2, T)$, where y is the vector concatenating all the responses y_j . Let $\theta = (\gamma, \sigma^2, \tau)^t$ be the vector encompassing all the parameters, where τ contains the non redundant elements of matrix T . Denoting by $d_l^{(j)}(\theta)$ the first partial derivative of $\ln \psi_j(y_j; \theta)$ with respect to θ_l and by $s_{lm}^{(j)}(\theta)$ its second derivative with respect to θ_l and θ_m , the following matrices may be defined:

$$D(\theta) = \left[\frac{1}{J} \sum_{j=1}^J d_l^{(j)}(\theta) d_m^{(j)}(\theta) \right]_{l,m} \quad \text{and} \quad S(\theta) = \left[-\frac{1}{J} \sum_{j=1}^J s_{lm}^{(j)}(\theta) \right]_{l,m} \quad (3)$$

with $l, m = 1, \dots, P$, being P the total number of non redundant parameters involved in the model. Analogously to standard results on maximum likelihood estimators, the asymptotic variance-covariance matrix of the estimator $\hat{\Theta}$ is given by firstly taking the expectation of matrices D and S in (3), denoted by D_E and S_E , respectively. Then, the asymptotic variance-covariance matrix is given by:

$$V^{as}(\hat{\Theta}) = \frac{1}{J} S_{E,*}^{-1} D_{E,*} S_{E,*}^{-1}, \quad (4)$$

where index “ $*$ ” denotes evaluation at the true, unknown value θ^* . Anyway, when the assumed model is correct: $D_{E,*} = S_{E,*}$, so that simply: $V^{as}(\hat{\Theta}) = \frac{1}{J} S_{E,*}^{-1} = \mathcal{I}^{-1}(\theta^*)$,

being $\mathcal{I}(\theta^*)$ the (expected) Fisher information matrix. What was proved is that the result in (4) remains valid if matrices $\mathbf{S}_{E,*}$ and $\mathbf{D}_{E,*}$ are replaced by the corresponding versions $\hat{\mathbf{S}}$ and $\hat{\mathbf{D}}$ evaluated at the maximum likelihood estimate $\hat{\theta}$. Hence:

$$\hat{\mathbf{V}}^{as}(\hat{\Theta}) = \frac{1}{J} \hat{\mathbf{S}}^{-1} \hat{\mathbf{D}} \hat{\mathbf{S}}^{-1} = \hat{\mathcal{I}}^{-1} J \hat{\mathbf{D}} \hat{\mathcal{I}}^{-1}, \quad (5)$$

where $\hat{\mathcal{I}}$ denotes the empirical Fisher information matrix. Formula (5) represents a sandwich-type correction for the model-based variance-covariance matrix of $\hat{\Theta}$, which was seen empirically to perform better than the uncorrected one in presence of misspecifications in the random effects distribution (Verbeke and Lesaffre, 1997).

Verbeke and Lesaffre (1997) implemented this type of correction in the software SAS. Implementation in S-Plus is unfortunately not so straightforward. In fact, in the library `nlme3` by Pinheiro and Bates (2000) estimation proceeds by starting from the profile log-likelihood that depends exclusively on the level-2 variance components, which are previously transformed according to some convenient unconstrained parameterization (Pinheiro and Bates, 1996). Then estimates for fixed effects γ and level-1 variance σ^2 are obtained consequently as functions of level-2 variance component estimates. Therefore, though the matrix $\hat{\mathcal{I}}$ in (5) is provided, it is the matrix $\hat{\mathbf{D}}$ that cannot be derived directly through this approach and then it will have to be suitably computed.

6. Statistical Calibration

According to previous remarks, linear multilevel models will be here applied in the calibration framework. Statistical calibration deals with the inference on unknown values of a p -dimensional explanatory variable $\mathbf{X}$ given a vector of q -dimensional response variables $\mathbf{Y}$. For example, such a situation may arise when two different instruments for the measurement of the same phenomenon are considered. The first one producing the measure $\mathbf{X}$ is more accurate, but also more difficult and expensive to obtain than the second one producing $\mathbf{Y}$. On the basis of the relationship between $\mathbf{Y}$ and $\mathbf{X}$ it is possible to determine the measure $\mathbf{X}$ when only the measure $\mathbf{Y}$ is observed.

The purpose of predicting unknown values of explanatory variables from further observed responses can be reached by the following two steps (Brown, 1982):

1) *The calibration step.* A first experiment of randomly drawing N observations on q response variables $Y_1, Y_2, \dots, Y_q$ and p explanatory variables $X_1, X_2, \dots, X_p$ is run in order to identify the transfer function that links $\mathbf{Y}$ to $\mathbf{X}$. Suppose that the transfer function is linear, then the calibration model is given by:

$$\mathbf{Y}_1 = \mathbf{1}\alpha^t + \mathbf{X}\mathbf{B} + \mathbf{E}_1 \quad (6)$$

where $\mathbf{Y}_1(N \times q)$ and $\mathbf{E}_1(N \times q)$ are random matrix, $\mathbf{X}(N \times p)$ is a matrix of fixed constants, $\mathbf{1}(N \times 1)$ is the unit vector, $\mathbf{B}(p \times q)$ and $\alpha(p \times 1)$ are, respectively, a matrix and a vector of parameters.

2) *The prediction step.* Analogously to the previous step, suppose that new m observations of $\mathbf{Y}$ are available and that they are related to an unknown vector of values ξ . Then this leads to the following prediction model:

$$\mathbf{Y}_2 = \mathbf{1}\alpha^t + \mathbf{1}\xi^t\mathbf{B} + \mathbf{E}_2 \quad (7)$$

where $\mathbf{Y}_2(m \times q)$ and $\mathbf{E}_2(m \times q)$ are random matrix and $\boldsymbol{\xi}(p \times 1)$ is the vector of unknown values of $\mathbf{X}$ that have to be estimated.

In both $\mathbf{E}_1$ and $\mathbf{E}_2$ the row vectors are $N_q(\mathbf{0}, \mathbf{\Gamma})$ and they are mutually independent.

When $q = p$ the multivariate classical estimator for the unknown values $\boldsymbol{\xi}$ of $\mathbf{X}$:

$$\hat{\boldsymbol{\xi}}_C = \left(\hat{\mathbf{B}}\mathbf{S}^{-1}\hat{\mathbf{B}}^t \right)^{-1} \hat{\mathbf{B}}\mathbf{S}^{-1}(\bar{\mathbf{y}}_2 - \hat{\boldsymbol{\alpha}}) \quad (8)$$

where $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\alpha}}$ are the least-squares estimators of $\mathbf{B}$ and $\boldsymbol{\alpha}$ in the model (6), $\bar{\mathbf{y}}_2$ is the mean vector of $\mathbf{Y}$ in the prediction experiment and $\mathbf{S}$ is the pooled covariance matrix. $\hat{\boldsymbol{\xi}}_C$ is also the maximum likelihood (ML) estimator of $\boldsymbol{\xi}$. For further details and for $q \neq p$, see Brown (1993).

As it is well known standard regression procedures are very sensitive to the presence of misspecifications or perturbations. Since calibration directly stems from regression, it is affected from the same problems.

Very little is known about robust calibration, nevertheless some attempts have been made. A first one consists in extending robust regression estimators in the calibration context. These techniques give robust calibration estimators (Cheng and Van Ness, 1997) but extensions of these methods when $p > 1$ and $q > 1$ lead to difficulties. A second effort, applicable for $q > p$, is based on the so-called inconsistency diagnostic $\mathbf{R}$, which is a central statistic in the multivariate calibration for diagnostic checking (Brown and Sundberg, 1989).

Another approach, the third one, consists in performing a classical multivariate estimator on data after rejecting outliers. Two ways are actually proposed in order to find the bulk of data: genetic algorithms (Walczak, 1995) and the forward search (Riani, Salini 2005). In particular the forward search appears to be effective especially in the presence of multiple outliers or groups, whereas calibration estimators derived by robust regression estimators present the same problems of the classical ones (Salini, 2006).

Undoubtedly, the forward search allows both finding the groups and monitoring the group effects on the estimator performance, but the main guidelines on how a calibration estimator should be suitably constructed have to be found elsewhere. In this regard, we feel that the multilevel model methodology might offer a powerful tool in the evaluation of the relevance of grouping structures and then in the development of calibration estimators that take into account the dependency among the observations due to the presence of clusters. With this sense we will indicate as “robust” the multilevel calibration estimator that we are going to introduce in the next section.

7. Multilevel Calibration Estimator

As above mentioned, we deal with the calibration problem by assuming that data are clustered in just assigned J groups. For simplicity, here we focus on the univariate model (i.e. $q = p = 1$ in formulas (6)–(8)). In other words, we assume that both the variables Y and X are unidimensional and that n observations are randomly drawn from each of the J groups. Moreover, the group effect is assumed to be not negligible.

Three different approaches may be referred to in setting up the calibration estimator. The first one, which is the most common practice, is to ignore the grouping structure at all and to construct an unique calibration model over the whole dataset. Then, by specifying

formula (8) to this case the classical calibration estimator $\hat{\xi}_C$ for ξ is given by:

$$\hat{\xi}_C = \frac{y_2 - \hat{\alpha}}{\hat{B}}. \quad (9)$$

Anyway, as just mentioned in Section 6, the classical calibration model does not produce reliable estimators when a grouping structure is present and the relationship between the X measure and the Y measure is thereby influenced. Moreover, what was seen is that calibration estimators derived by robust regression estimators may be a good solution when outliers are included in data, but unfortunately they may produce unreliable results whenever data are organized according to a grouping structure.

A second approach, which might be seen somewhat fairly natural, consists in defining separate models for each group and then estimating the parameters by applying the least squares estimation method. Therefore, in our case a total number of $2J$ estimates $\hat{\alpha}_j$ and $\hat{\beta}_j$ have to be computed so as to obtain the “within-group” calibration estimator:

$$\hat{\xi}_{C_j}^w = \frac{y_2 - \hat{\alpha}_j}{\hat{B}_j}, \quad j = 1, \dots, J. \quad (10)$$

In such a way more efficient estimators are yielded, but several parameters are involved. This approach represents actually an overfitting of the data, which might be unadvisable if the main objective consists of making prediction on further data rather than inferring from the present ones. Moreover, an unified formulation of the model is often preferable since it is much more manageable for practical purposes, as for instance it may occur when the experiment is carried out in order to obtain a calibration model for constructing an engineering measuring instrument.

A third approach relies on the multilevel model methodology. In fact, multilevel models may be a powerful tool also in the calibration context, settling by a compromise the opposite requirements of efficiency of estimators, on one hand, and simplicity and flexibility in the formulation of the model, on the other hand. In the literature an effort in this direction is the proposal by Oman (1998), who introduces a calibration model with random slopes for repeated measures.

Here we consider a specific multilevel model definition, namely a random intercept model. Furthermore, for convenience we assume that $m = 1$ in the prediction step. By specifying the model (1) with respect to the calibration context and in a more compact way over the J groups all together, the random intercept calibration model may be expressed by:

1) at the calibration step:

$$\mathbf{Y}_1 = [\mathbf{1}, \mathbf{x}] \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \mathbf{U}_{0,1} + \boldsymbol{\varepsilon}_1; \quad (11)$$

2) at the prediction step:

$$Y_{2j^*} = \gamma_{00} + \gamma_{10}\xi_{j^*} + U_{0,2j^*} + \epsilon_{2j^*}, \quad (12)$$

where $\mathbf{Y}_1$ is the response variable vector of the calibration step concatenating all the $\mathbf{Y}_j$ by column; Y_{2j^*} is the response variable of the prediction step referred to the group j^* , which is one among the J groups to whom Y_2 variable may be referred; $\mathbf{x}$ is a column

vector of values of variable X ; ξ_{j^*} is the unknown value of X that corresponds to Y_{2j^*} ; γ_{00} and γ_{10} are, respectively, the general intercept and the general regression coefficient; $\mathbf{U}_{0,1}$ is the vector concatenating by column all the random effects $\mathbf{U}_j$ related to the intercept; $U_{0,2j^*}$ is the random effect related to γ_{00} and to the group j^* ; $\boldsymbol{\varepsilon}_1$ is the vector concatenating the level-1 errors ε_j ; finally, ε_{2j^*} is the level-1 error in the prediction step within the group j^* . On the random vectors $\boldsymbol{\varepsilon}_1$ and $\mathbf{U}_{0,1}$ the usual assumptions, including normality, are adopted. Furthermore, it is worth pointing out that this approach assumes that the belonging to the group j^* is known.

On this basis, the multilevel calibration estimator is given by:

$$\hat{\xi}_{j^*}^m = \frac{y_{2j^*} - \hat{\gamma}_{00}^m - \hat{U}_{0,2j^*}}{\hat{\gamma}_{10}^m}, \quad (13)$$

where $\hat{\gamma}_{00}^m$ and $\hat{\gamma}_{10}^m$ denote REML or ML estimators for fixed effects γ_{00} and γ_{10} , whereas $\hat{U}_{0,2j^*}$ is the empirical Bayes estimator for the value u_{0j^*} of $U_{0,2j^*}$ (see e.g. Pinheiro and Bates, 2000).

To gain preliminary indications on the performance of the calibration estimators (9), (10) and (13), a simulation study has been carried out by considering $J = 5, 10, 20, 50$ groups and $n = 10, 50, 100$ units within group. Experimental conditions are then given by combinations of J and n . For each of these, $K = 1,000$ datasets have been artificially constructed: pseudo-observations $\mathbf{y}_1$ have been computed by applying model (11) with independent $U_{0j} \sim N(0, 10)$ and $\varepsilon_{ij} \sim N(0, 20)$. Similarly, values of vector $\mathbf{x}$ have been randomly drawn from $N(30, 40)$, but they have been kept fixed over the K replications. Subsequently, all the estimates derived from the classical calibration estimator (9), the within-group calibration estimator (10) and the REML and ML multilevel calibration estimator (13) have been computed. All simulations have been carried out in R software; whenever the multilevel model is involved the library `nlme` by Pinheiro and Bates (2000) has been used.

Table 2 displays the main results achieved for the Monte Carlo Mean Squares Error, which is given by:

$$MSE(\hat{\xi}) = \frac{1}{nJ} \frac{1}{K} \sum_{i=1}^{nJ} \sum_{k=1}^K (\xi_i - \hat{\xi}_{ik})^2 \quad (14)$$

where ξ_i is the true value adopted in the simulation design and $\hat{\xi}_{ik}$ generically indicates its estimate given by (9), (10) or (13) according to the model considered.

This simulation study confirms our expectations. As it may be readily seen from Table 2, under each experimental conditions the classical estimator is characterized by the greatest MSE . On the contrary, the within-group estimator has the smallest MSE , while the REML and ML multilevel estimators have a sort of “intermediate” MSE . Nevertheless, this gap tends to become negligible as the sample size increases.

To wrap up, we believe that the multilevel calibration estimator represents a valid solution to the problem of calibration when data are clustered and when the number of groups is large. In particular, this latest point represents the main motivation towards the need of an unified formulation of the model.

Table 2: Simulation Study, $J = 5, 10, 20, 50$ groups and $n = 10, 50, 100$ units within group, Monte Carlo Mean Square Error for REML estimator ($\hat{\xi}_{j^*}^{\text{REML}}$), ML estimator ($\hat{\xi}_{j^*}^{\text{ML}}$), classical calibration estimator ($\hat{\xi}_C$) and within-group calibration estimator ($\hat{\xi}_{C_j}^w$).

n	MSE	Number of groups (J)			
		5	10	20	50
10	multilevel REML	1.411	1.424	1.430	1.421
	multilevel ML	1.419	1.427	1.431	1.421
	classical	1.730	1.780	1.811	1.817
	within-group	1.350	1.351	1.352	1.350
50	multilevel REML	1.473	1.476	1.476	1.459
	multilevel ML	1.474	1.476	1.476	1.459
	classical	1.757	1.798	1.800	1.814
	within-group	1.462	1.464	1.463	1.424
100	multilevel REML	1.483	1.482	1.481	1.483
	multilevel ML	1.483	1.482	1.480	1.483
	classical	1.760	1.790	1.809	1.821
	within-group	1.476	1.475	1.475	1.477

8. Conclusion

In this paper we have considered the connection “robustness” and “multilevel models” according to a dual direction of analysis: robustness in multilevel models and multilevel models for robustness. As regards the former, we have briefly discussed the current acceptations of the term robustness in multilevel model framework. Then, we have summed up the main results achieved in our studies, which focus on the performance of maximum likelihood estimators in multilevel models with MEP distributed random effects. As for the second point, we have suggested using multilevel models to evaluate the group effect and then to set up a calibration estimator that takes into account the grouping structure. A first study on the performance of this approach has encouraged us to proceed in this direction, by extending the approach to more complex multilevel models and to multivariate calibration.

References

- Atkinson, A.C. and Riani, M. (2000). *Robust Diagnostic Regression Analysis*, Springer-Verlag, New York.
- Atkinson, A.C., Riani, M. and Cerioli, A. (2004). *Exploring Multivariate Data With the Forward Search*, Springer-Verlag, New York.
- Brown, P. J. and Sundberg, R. (1989), Prediction Diagnostic and Updating in Multivariate Calibration, *Biometrika*, 72, 349-361.
- Brown, P. J. (1982), Multivariate Calibration, *JRSS, Series B*, 44, 287-321.
- Brown, P. J. (1993), *Measurement, Regression and Calibration*, Clarendon Press, Oxford.
- Cheng, C. L. and Van Ness, J. W. (1997), Robust Calibration, *Technometrics*, 39, 401-411.

- Fang, K.-T., Kotz, S. and Ng, K.W. (1990), *Symmetric Multivariate and Related Distributions*, Monograph on Statistics and Applied Probability, 36, Chapman and Hall, New York.
- Ferrari, P.A. and Solaro, N. (2002), Una proposta per le componenti erratiche del modello multilivello. In: B.V. Frosini, U. Magagnoli, G. Boari: *Studi in onore di Angelo Zanella*. V&P Università, Milano, 273–291.
- Goldstein, H. (1995), *Multilevel Statistical Models*, 2nd edition, Kendall's Library of Statistics, 3, Arnold Ed., London.
- Gómez, E., Gómez-Villegas, M.A. and Marín, J.M. (1998), A multivariate generalization of the power exponential family of distributions, *Communications in Statistics – Theory and Methods*, 27, 3, 589–600.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986), *Robust Statistics: The Approach based on Influence Functions*. Wiley, New York.
- Huber, P. J. (1981), *Robust Statistics*. Wiley, New York.
- Kreft, I.G.G. (1996), Are multilevel techniques necessary? An overview, including simulation studies, *Technical report*, <http://www.calstatela.edu/faculty/ikreft>
- Laird, N.M. and Ware, J.H. (1982), Random-effects models for longitudinal data, *Biometrics*, 38, 963–974.
- Langford, I.H. and Lewis, T. (1998), Outliers in multilevel data, *JRSS, Series A*, 161, 2, 121–160.
- Maas, C.J.M. and Hox, J. (2004), Robustness issues in multilevel regression analysis, *Statistica Neerlandica*, 58, 2, 127–137.
- Oman S. D. (1998), Calibration with Random Slopes, *Biometrika*, 85, 439–449.
- Pinheiro, J., Liu, C., and Wu, Y. (1997), Robust Estimation in Linear Mixed-Effects Models Using the Multivariate t-distribution, *Technical Report*, Bell Labs.
- Pinheiro, C.J. and Bates, D.M. (1996), Unconstrained parametrizations for variance-covariance matrix, *Statistics and Computing*, 6, 289–296.
- Pinheiro, C.J. and Bates, D.M. (2000), *Mixed-Effects Models in S and S-Plus*, Statistics and Computing, Springer-Verlag, New York.
- Riani M., Salini, S. (2005), Robust and Efficient Multivariate Calibration, *Proceeding of ICORS Conference*. Jyväskylä, Finland.
- Salini, S. (2006), Robust Multivariate Calibration, in: Zani, S., Cerioli, A., Riani, M., Vichi, M. (eds.), *Data Analysis, Classification and Forward Search*. Springer, Berlin (in press).
- Snijders, T. and Bosker R. (1999), *Multilevel Analysis – An Introduction to Basic and Advanced Multilevel Modelling*, Sage Publications, London.
- Solaro, N. and Ferrari, P.A. (2006a), Robustness of parameter estimation procedures in multilevel models when random effects are MEP distributed, to appear in: *Statistical Methods and Applications*, Springer-Verlag.
- Solaro, N. and Ferrari, P.A. (2006b), The effects of MEP distributed random effects on variance component estimation in multilevel models, in: Zani, S., Cerioli, A., Riani, M., Vichi, M. (eds.), *Data Analysis, Classification and Forward Search*. Springer, Berlin (in press).
- Verbeke, G. and Lesaffre, E. (1997), The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data, *Computational Statistics and Data Analysis*, 23, 541–556.
- Walczak, B. (1995), Outlier Detection in Multivariate Calibration, *Chemometrics and Intelligent Laboratory Systems*, 28, 259–272.