Distribution Free Methods for Longitudinal Survey Data Models

Leandro Vitral Andraos (leandro.andraos@ibge.gov.br)
Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brazil
and
Marcel de Toledo Vieira* (marcel.vieira@ufjf.edu.br)
Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil

Abstract: This paper investigates distribution free statistical methods with computational support for longitudinal survey data. Inference procedures for mixed effects models parameters are evaluated considering alternative longitudinal covariance structures. Estimation methods that consider the sampling design and those which ignore are compared through a simulation study and the behaviour of estimators that are based on fitting functions is evaluated. Maximum likelihood, pseudo maximum likelihood and distribution free generalized least squares point estimators are considered. The performance of the estimators is also evaluated considering different scenarios where data are generated from various probability distributions, including both symmetrical and asymmetrical ones.

Keywords: mixed effects models; covariance structures; fitting functions; BHPS.

1. Introduction

What makes longitudinal research so indispensable? Longitudinal data are currently the main sources of information related to changing demographic, social and economic conditions in the population. In longitudinal studies variables are measured repeatedly for the same individuals at different times of time.

This paper studies statistical methods with computational support of computational for panel data analysis considering inference procedures for mixed effects models. Comparisons of methods that consider the sampling design and those that do not account for that is performed by means of a simulation study aiming at evaluating the behaviour of the estimators. The following point estimation methods are based on fitting functions and are considered here: maximum likelihood (ML), pseudo maximum likelihood (PML) and distribution free generalized least squares (GLS and GLSC, the latter considering the sampling plan). The performance of the estimators is also evaluated in situations where data are generated from different probability distributions, including both symmetric and asymmetric distributions.

Data used in this paper are from British Househould Panel Survey (BHPS) which is a panel study (see Taylor *et al*., 2001). Data collection began in 1991 and the adopted sampling design is stratified and clustered in multiple stages. In the analysis a subset consisting of men and women who gave complete answers in 1991, 1993, 1995, 1997, and 1999 is considered. Variables include: age at first occasion, a gender role score, gender, education level, economic activity, among others. The gender role score is the response variable of the models, higher values represent more egalitarian attitudes toward gender roles. In addition, the data file includes a longitudinal weight variable that allows for the sample design. More information about the sample design of BHPS and the variables can be found in Vieira (2009).

2. Longitudinal Regression Model

Marginal models and the mixed effects models are among the most discussed longitudinal models in the statistical literature. Mixed effects models parameters are useful in evaluating changes in individual parameters and marginal models are used to evaluate the aggregate changes. In this paper only the mixed effects model is considered. More information about this model are presented in, for example, Diggle *et al*. (2002), Liang and Zeger (1986) and Vieira (2009).

Let the finite population be fixed in occasions 1,...,T and $\underline{Y_i} = (Y_{i1}, \dots, Y_{iT})'$ be a random vector with T repeated observations of the study variable for units i=1,2,..., N. For simplicity the mixed effects model is given here by

$$Y_{it} = \underline{x}_{it}\underline{\beta} + \varepsilon_{it}, \qquad (1)$$

Where ε_{it} is the model random term and $E(\varepsilon_{it}) = 0$. In (1), let i represent individuals, t occasions, x_{it} are $1\times q$ vectors with q covariates, β is a $q\times 1$ vector with unknown coefficients for covariates x and ε_{it} is the residual term which may be decomposed in different random effects.

Random effects imply random coefficients (intercepts and slopes) that take into account the variability between units. This class of models allows the unexplained response variable variability to be decomposed into different components. In this paper only random effects for the intercept are considered as in Vieira (2009).

The model covariance structure depends on how the correlation between the residuals over time is processed. Alternative correlation matrices can be considered to model the residuals, but only exchangeable correlation structure represented by the uniform correlation model (UCM) is adopted here. Let the UCM be given by

$$Y_{it} = \underline{x}_{it}\underline{\beta} + u_i + v_{it}\text{, with } i = 1,..., N \text{ and } t = 1,\ldots,T. \qquad (2)$$

In (2), u_i are permanent random effects (specific unobserved factors) and v_{it} are transitory random effects. Notice that u_i does not have an index for time as it represents and individual permanent characteristic and that v_{it} has an index *t* because it represents and individual characteristic that is variable over time.

It is further assumed that (i) the observations are equally spaced in time, and (ii) the number of individuals in the population, *N* is relatively large compared to the number of observations per individual (T). It is also assumed that (iii) the sample is selected by probability sampling on the first occasion and then the same sample units visited in each *T*-1 subsequent occasions in the survey.

3. Longitudinal Covariance Structure Models

Pourahmadi (2000), and Pan and Mackenzie (2003), among other authors, discussed the adoption of models for the structure of covariance for estimation of variance components in the longitudinal data context assuming that the observations for different individuals are independent and identically distributed (iid).

The focus of this section is the study of inference procedures for mixed effects models for longitudinal data, initially ignoring the complex sample design and then allowing for it in the estimation of the parameters. Our parameters of interest are

σ_u^2 = between individual variation,
σ_v^2 = within individual variation, and
γ = correlation over time.

Let $\underline{\theta}$ be the b×1 parameter vector of interest. For the UCM presented above, $\underline{\theta} = (\sigma_u^2,\ \sigma_v^2)$. Here, alternative methods for estimating $\underline{\theta}$ are considered. It is the aim of this paper to evaluate the behaviour of methods that allow for the sampling design and to study the impacts of ignoring such feature. Ignoring the sampling design in the analysis and modelling survey data as iid may result in bias in both point and variance estimates and the distribution of test statistics (Vieira, 2009).

Let Σ be a population structure equation model given by the following *T*×*T* matrix,

$$\Sigma = COV(\underline{Y}_i) = E\{[\underline{Y}_i - \underline{\mu}_i][\underline{Y}_i - \mu_i]'\} = \Sigma(\underline{\theta}), \qquad (3)$$

where *COV*(.) denotes the population covariance.

The covariance matrix may be defined in accordance with the parametric structure defined by the model considered for y_{it}, i.e., $\Sigma = \Sigma(\underline{\theta})$. Let $F(S_w, \Sigma)$ and $F(S, \Sigma)$ be fitting (or discrepancy) functions which through $\underline{\hat{\theta}}$ minimise the distance between S_w and Σ, and between S and Σ, respectively. Notice that S_w and S are estimators of Σ, and therefore we aim at finding estimators which minimises such distances. For the case of S_w, for example, the fitting function is given by

$$F\left(S_w, \Sigma\left(\underline{\hat{\theta}}\right)\right) = \min_{\underline{\theta}\in\Theta} F\left(S_w, \Sigma\left(\underline{\theta}\right)\right). \qquad (4)$$

If the model fits the data perfectly, $F(\underline{\theta}) = 0$. Thus, the smaller the value of $F(\underline{\theta})$ is the better will the model fitting. According to Browne (1982, 1984) the fitting function has the following properties: (*i*) $F(\underline{\theta})$ is a scalar; (*ii*) $F(\underline{\theta}) \geq 0$; (*iii*) $F(S_w, \Sigma(\underline{\theta})) = 0$ if and only if $S_w = \Sigma$; and (*iv*) $F(\underline{\theta})$ is a continuously differentiable of S_w and Σ.

4. Estimation Methods

4.1 Maximum likelihood (ML)

In the current context, the maximum likelihood (ML) estimator for $\underline{\theta}$ is obtained by minimising the following fitting function

$$F(\underline{\theta})_{ML} = tr\left[S\Sigma(\underline{\theta})^{-1}\right] - \log\left|S\Sigma(\underline{\theta})^{-1}\right| - T. \qquad (5)$$

The differentiation of $F(\underline{\theta})_{MV}$ with respect to each of the b components of the parameter vector $\underline{\theta}$ (Wiley, Schmidt e Bramble, 1973), results in the solution of the following system of equations

$$\frac{\partial F(\underline{\theta})_{ML}}{\partial \theta_{\mathrm{b}}} = tr\left\{\Sigma(\underline{\theta})^{-1} \cdot [\Sigma(\underline{\theta}) - \mathrm{S}] \cdot \Sigma(\underline{\theta})^{-1} \cdot \frac{\partial \Sigma(\underline{\theta})}{\partial \theta_{\mathrm{b}}}\right\} = 0 .$$

The estimation strategy adopted here is based on initially estimating $\underline{\beta}$ separately and then minimising the fitting function with respect only to $\underline{\theta}$. Considering the model we assume for the dependent variable there is not a analytical solution for the minimisation problem. Therefore the Newton-Raphson numeric minimisation procedure has been adopted. Note that the ML estimator assumes the observations are IID.

4.2 Pseudo maximum likelihood (PML)
The Pseudo maximum likelihood (PML) estimator is described below. The PML estimator allows for the complex sampling design for estimating $\underline{\theta}$, when minimising the following fitting function (Vieira and Skinner, 2008)

$$F(\underline{\theta})_{PML} = tr\left[\mathrm{S}_w \Sigma(\underline{\theta})^{-1}\right] - \log\left|\mathrm{S}_w \Sigma(\underline{\theta})^{-1}\right| - T. \qquad (6)$$

Note that $F(\underline{\theta})_{PML}$ reduces to $F(\underline{\theta})_{ML}$ when the sampling weights w_i are constant. Again, the Newton-Raphson numeric minimisation procedure has been adopted, for the same reason stated above.

4.3 - Distribution free generalized least squares (GLS and GLSC)
The Generalised least squares (GLS) estimation method proposed by Jöreskog and Goldberger (1972) in the context of covariance structure models is also considered. The GLS estimator accounts for the inverse of the matrix with the covariance of the residuals. Elements of Σ with larger sample variance receive smaller weights.

Two versions of the GLS estimator are presented here. The first one, which is based on the fitting function presented below, does not account for the sampling design,

$$F(\underline{\theta})_{GLS} = \{vech[S] - vech[\Sigma(\underline{\theta})]\}' \hat{C}^{-1} \{vech[S] - vech[\Sigma(\underline{\theta})]\}. \qquad (7)$$

Where $vech[S]$ and $vech[\Sigma(\underline{\theta})]$ are k × 1 vectors that include non-duplicated elements from matrices S and $\Sigma(\underline{\theta})$. Moreover, $\hat{C}$ is a k × k positive definite matrix which estimated the covariance matrix of the matrix E, which is the T × T residual matrix, without allowing for the sampling design. For further details see Vieira and Skinner (2008) and Vieira (2009).

A second version of the GLS estimator allows for the complex sampling design and it is denoted here as GLSC, which is weighted by the sampling weights w_i. The GLSC fitting function is given by (Vieira and Skinner, 2008)

$$F(\underline{\theta})_{GLSC} = \{vech[\mathrm{S}_w] - vech[\Sigma(\underline{\theta})]\}' \hat{\mathrm{C}}_c^{-1} \{vech[\mathrm{S}_w] - vech[\Sigma(\underline{\theta})]\}. \qquad (8)$$

Where $\hat{C}_c$ is the estimator of the covariance matrix of E, introduced by Vieira and Skinner (2008) and further described by Vieira (2009). Note that $\hat{C}_c$ allows for the sampling design and is an asymptotically distribution free estimator. Nevertheless, according to Bollen (1989, p. 432) and Satorra (1992), distribution free methods should be used carefully in the context of covariance structure models as these are dependent of large sample sizes in order to be robust to violation of distributional assumptions.

4. Simulation Study
This section presents the characteristics and the results of a simulation study that aims to evaluate the statistical properties of point estimation procedures discussed above. We use the same criteria adopted by Vieira (2009) to evaluate the behaviour of the estimators, i.e., (i) mean square error (MSE), (ii) relative bias, and (iii) the coefficient of variation (CV) .

A number 1000 sample replicates were simulated. The values of x are held fix and values of Y_{it} are simulated from a real BHPS data subset with 1340 individuals considering different models, independently for each replicate. For evaluating the effects of sample size variations alternative sample sizes were considered: $n_{sim}=100$, $n_{sim}=500$, $n_{sim}=10000$, and $n_{sim}=20000$. In this simulation exercise, for the same model presented above residuals were generated from different probability distributions in addition to the Normal: (i) Student t with υ = 4 degrees of freedom; (ii) Asymmetric Normal with asymmetry parameter $\lambda = 3$; and (iii) asymmetric t with υ = 4 and $\lambda = 3$. The value for the asymmetry parameter was chosen such that the residuals are considerably asymmetric (Ferreira, Bolfarine and Lachos, 2013).

For the generation of the values of the variable of interest considering the models and distributions described above , we consider the methodology defined by Vieira (2009). For all scenarios UCM were fitted. For the implementation and achievement of the simulation results the R software (version 2.15.3) were adopted.

For each simulation the four estimators are used studied in this work , of which the first two are unweighted while the last two estimates are weighted by longitudinal sample weights of BHPS. Thus the following estimators are considered,

(i) generalised least squares ($\underline{\hat{\theta}}_{GLS}$),

(ii) maximum likelihood ($\underline{\hat{\theta}}_{ML}$),

(iii) weighted generalised least squares ($\underline{\hat{\theta}}_{GLSC}$, Vieira and Skinner, 2008) and

(iv) pseudo maximum likelihood ($\underline{\hat{\theta}}_{PML}$, Vieira and Skinner, 2008).

Note that (i) and (iii) distribution free methods (Vieira and Skinner, 2008; and Vieira, 2009).

5. Results

5.1 Normally Distributed and Student t Distributed Residuals

For the scenarios that considered normally distributed or Student t distributed residuals, differences were observed in terms of MSE for the distribution free methods when compared to the other methods. According to Satorra (1992) one should be always careful when working with distribution free methods in the current context with sample sizes smaller than 5000. In the current work, the GLSC estimator had the largest MSE among all methods for estimating σ_u^2 and the GLS estimator had the largest MSE for estimating σ_v^2 , even though in terms of c.v. both methods behaved very similarly. For larger sample sizes c.v. and MSE have diminished and the GLSC method became the most efficient method. Tables 1 and 2 include results obtained for the largest sample size considered here.

Parameter		Relative Bias	$cv(\hat{\theta})$	MSE $(\times 1000)$
$\underline{\hat{\theta}}_{GLS}$	$\hat{\sigma}_u^2$	-0.87%	1.28%	16.198
	$\hat{\sigma}_v^2$	-0.23%	0.52%	1.071
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-0.84%	1.18%	16.051
	$\hat{\sigma}_v^2$	-0.15%	0.48%	0.833
$\underline{\hat{\theta}}_{GLSC}$	$\hat{\sigma}_u^2$	-0.93%	1.18%	16.021
	$\hat{\sigma}_v^2$	-0.22%	0.48%	0.902
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-0.82%	1.27%	15.899
	$\hat{\sigma}_v^2$	-0.22%	0.52%	0.944

Table 1 – Results considering residuals with normal distribution and $n_{sim}=20000$.

Parameter		Relative Bias	$cv(\hat{\theta})$	MSE $(\times 1000)$
$\underline{\hat{\theta}}_{GLS}$	$\hat{\sigma}_u^2$	-4.87%	2.07%	701.250
	$\hat{\sigma}_v^2$	-1.39%	1.23%	8.504
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-4.75%	1.97%	687.289
	$\hat{\sigma}_v^2$	-0.32%	1.47%	5.626
$\underline{\hat{\theta}}_{GLSC}$	$\hat{\sigma}_u^2$	-4.75%	1.97%	685.934
	$\hat{\sigma}_v^2$	-0.28%	1.18%	5.454
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-4.76%	2.06%	684.071
	$\hat{\sigma}_v^2$	-0.32%	1.58%	6.443

Table 2 – Results considering residuals with $t_{\upsilon=4}(0,1)$ distribution and $n_{sim}=20000$.

5.2 Residuals with Asymmetric Normal Distribution

Results produced when considering residuals with Asymmetric Normal distribution had bias, variance, and consequently MSE much larger than the scenario which considered symmetric residuals. The GLSC estimator had the largest MSE among all methods for σ_u^2 while the GLS estimator had the largest MSE for σ_v^2. However, in terms of c.v. both distribution free methods behaved similarly to the other methods. The performance difference between the distribution free methods and the other methods are much smaller for larger sample sizes. For the largest sample size considered the GLSC method had the smaller c.v. for estimating σ_v^2 and a MSE similar to the ML method. Table 3 presents results considering the current scenario.

Parameter		Relative Bias	$cv(\hat{\theta})$	MSE $(\times 1000)$
$\underline{\hat{\theta}}_{GLS}$	$\hat{\sigma}_u^2$	-29.81%	1.33%	9067.81
	$\hat{\sigma}_v^2$	-28.74%	0.59%	4100.91
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-29.80%	1.25%	9062.78
	$\hat{\sigma}_v^2$	-28.72%	0.54%	4095.00
$\underline{\hat{\theta}}_{GLSC}$	$\hat{\sigma}_u^2$	-29.81%	1.25%	9064.38
	$\hat{\sigma}_v^2$	-28.73%	0.54%	4095.21
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-29.83%	1.33%	9063.55
	$\hat{\sigma}_v^2$	-28.72%	0.59%	4095.47

Table 3 – Results considering residuals with asymmetric normal distribution and $n_{sim} = 20000$.

5.3 Residuals with Asymmetric Student t Distribution

When comparing the MSE of the estimators for the asymmetric $t_{\upsilon=4}(0,1)$ residuals with the asymmetric normal results several differences in the behaviour of the estimators may be noticed. When comparing results from Table 3 with Table 4, below, there are several differences regarding the behaviour of the estimators with the most noticeable one being a better performance of the estimators for the asymmetric Student t distribution. Moreover, as in other considered scenarios, larger efficiency gains were noticed for distribution free methods in comparison with other methods when the sample size was increased.

Parameter		Relative Bias	$cv(\hat{\theta})$	MSE $(\times 1000)$
$\underline{\hat{\theta}}_{GLS}$	$\hat{\sigma}_u^2$	-4.41%	1.03%	331.805
	$\hat{\sigma}_v^2$	-2.43%	1.34%	2.630
$\underline{\hat{\theta}}_{ML}$	$\hat{\sigma}_u^2$	-5.05%	1.25%	299.117
	$\hat{\sigma}_v^2$	-0.31%	0.21%	1.365
$\underline{\hat{\theta}}_{GLSC}$	$\hat{\sigma}_u^2$	-5.28%	1.12%	303.470
	$\hat{\sigma}_v^2$	-2.24%	0.95%	1.560
$\underline{\hat{\theta}}_{PML}$	$\hat{\sigma}_u^2$	-5.04%	1.45%	302.896
	$\hat{\sigma}_v^2$	-0.38%	0.78%	1.870

Table 4 – Results considering residuals with asymmetric $t_{\upsilon=4}(0,1)$ distribution and $n_{sim} = 20000$.

6. Concluding Remarks

In this paper, the behaviour of methods proposed by Vieira and Skinner (2008) for longitudinal data analysis taking into account the complex sample design was further investigated. Initially, review of estimation procedures for parameters of longitudinal regression models assuming simple random sampling and IID observations, and also the complex survey data case, including maximum likelihood, pseudo maximum likelihood and generalized least squares.

Simulation results suggest that the evaluated estimation methods behave very similarly when the residuals follow the normal distribution, confirming results of Vieira and Skinner 2008. Similar conclusions can be found in scenarios in which the residuals followed the $t_{\upsilon=4}(0,1)$ distribution, but with results showing higher bias and MSE. For scenarios that considered residuals generated from asymmetric distributions, we noticed large differences in the bias values, c.v. and MSE for the different estimation methods. For the asymmetric normal, the results showed generally large bias and low efficiency. With the $t_{\upsilon=4}(0,1)$ distribution, the results had lower bias and MSE. The fact of the t distribution has heavier tails (thereby putting a greater probability to events that occur in their tails than the shorter tail distributions) seems to make the asymmetry to cause less problems for estimation methods than the normal distribution case.

The methods of ML (and PML) present estimators with generally good performance in terms of bias and variance independently of the sample size. As for the GLS and GLSC methods, these offer good performance in terms of bias and variance, in general, for sample sizes greater than 10,000. Therefore, we recommend that distribution free methods should be used with caution in situations where only small samples are available

Another relevant observed result is that in most cases the MSE of the estimators of σ_u^2 are much higher than for estimators of σ_v^2. The parameter σ_u^2 measures the variability of permanent random effects and σ_v^2 measures the variability of transitory random effects. Thus, because σ_v^2 varies in time, it considers *NT* observations while σ_u^2 considers only *N* observations.

Furthermore, several different paths could be taken in future works. For example, changing the asymmetry parameter λ = 3 it would be possible to study the impacts of different degrees of asymmetry in the estimation procedures. Additionally, other asymmetric distributions with heavier tails could be used as, for example, slash, contaminated normal, among others. Other sample sizes could be used as well as other estimation methods and the model itself, could also have expanded their complexity allowing, in addition to random intercepts, random slopes could be considered.

References

Bollen, K. A. (1989) *Structural Equations with Latent Variables*. New York, John Wiley & Sons.

Browne, M. W. (1982) Covariance Structures. In Hawkins, D. M. eds. *Topics in Applied Multivariate Analysis*. Cambridge, Cambridge University Press.

Browne, M. W. (1984) Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.

Diggle, P. J., Heagerty, P., Liang, K. & Zeger, S. L. (2002) *Analysis of Longitudinal Data*. 2nd ed. Oxford, Oxford University Press.

Ferreira, C. S, Lachos, V. H., and Bolfarine, H. (2014) *Inference and Diagnostics in Skew Scale Mixtures of Normal Regression Models*. Accepted for publication in Journal of Statistical Computation and Simulation, 85, 517-537.

Jöreskog, K. G. and Goldberger, A. S. (1972) Factor Analysis by Generalized Least Squares. 37, 243-260.

Liang, and Zeger, S.L. (1986). Longitudinal Data Analysis Using Generalised Linear Models. *Biometrika*, 73, 13–22.

Pan, J. and Mackenzie, G. (2003) On Modelling Mean-covariance Structures in Longitudinal Studies. *Biometrika*, Vol. 90, n. 1, 239-2449

Pourahmadi, M. (2000) Maximum Likelihood Estimation of Generalised Linear Models for Multivariate Normal Covariance Matrix. *Biometrika*, Vol. 87, n. 2, 425-435.

R Development Team (2015) *The R Environment for Statistical Computing and Graphics – Reference Index*, Version 2.15.3, The R Foundation for Statistical Computing.

Taylor, M. F. ed, with Brice J., Buck, N. and Prentice-Lane E. (2001) *British Household Panel Survey - User Manual - Volume A: Introduction, Technical Report and Appendices*.

Satorra, A. (1992) Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures. *Sociological Methodology*, Vol. 22, p. 249-278.

Vieira, M. D. T. (2009) *Analysis of Longitudinal Survey Data*. 1. ed. Saarbrücken: VDM Verlag.

Vieira, M. D. T. and Skinner, C. J. (2008) Estimating Models for Panel Survey Data under Complex Sampling. *Journal of Official Statistics*, Vol. 24, No. 3, pp. 343-364.

Wiley, D. E., Schmidt, W. H. and Bramble, W. J. (1973) Studies of a Class of Covariance Structure Models. *Journal of the American Statistical Association*, Vol. 68, N. 342, 317-323.