MIXED LINEAR MODELS

JAN DE LEEUW

ABSTRACT. This is an entry for The Encyclopedia of Statistics in Behavioral Science, to be published by Wiley in 2005.

A mixed linear model is a regression model of the form

$$\underline{y} = X\beta + Z\underline{\gamma} + \underline{\epsilon}.$$

The convention of underlining random variables shows that the regression coefficients β are fixed constants, while the regression coefficients γ are modeled as random variables. Thus γ is allowed to vary under repeated sampling. If we replicate our experiment, we will expect to see the same fixed regression coefficients, but a different realization of the random regression coefficients. In the social and educational sciences the most common mixed linear models are multilevel models, but random coefficient regression models are important in a much wider context, including biometrics and econometrics.

Date: April 29, 2004.

In particular, mixed models often show up in the seemingly more general form

$$\underline{y}_t = X\beta + Z\underline{\gamma}_t + \underline{\epsilon}_t,$$

where we have different, but related, regression models in different contexts, occasions, time points, or groups. One then supposes the $\underline{\gamma}_t$ are independent realizations of the same random vector, and they differ only because of random variation. This is a useful model, intermediate between assuming that all regression coefficients are equal to the same fixed vector γ or assuming they are different unrelated fixed vectors γ_t for different contexts.

Alternatively, we can understand mixed linear models as heteroscedastic regression models with a specific covariance structure for the disturbances. We typically suppose that γ and ϵ both have expectation zero and are uncorrelated, and have dispersions $\mathbf{V}(\underline{\gamma}) = \Omega$ and $\mathbf{V}(\underline{\epsilon}) = \sigma^2 I$. Then

$$\mathbf{E}(\underline{y}) = X\beta,$$

$$\mathbf{V}(\underline{y}) = Z\Omega Z' + \sigma^2 I.$$

Thus we see that the covariance of the disturbances has a factor analytic structure, with known factor loadings Z and with equal unique variances. As in factor analysis, it is often convenient to require Ω to be diagonal, with diagonal elements ω_s^2. In that case we have only variance components, no covariance components, and we can write

$$\mathbf{V}(\underline{y}) = \omega_1^2 V_1 + \cdots + \omega_p^2 V_p + \sigma^2 I,$$

where $V_s = z_s z_s'$. This shows that linear mixed models have a linear model for the expectations, and also a linear model for the variance components. Expected values of $\underline{y}$ are an unknown linear combination of the known columns of X, and the dispersion matrix of $\underline{y}$ is an unknown non-negative linear combination of the known positive semi-definite matrices V_s.

Linear mixed models are plausible whenever the covariance structure of the observations can plausibly be modeled by one of these factor analysis type specifications. If we compare the mixed model with the fixed model $\mathbf{E}(\underline{y}) = X\beta + Z\gamma$ and $\mathbf{V}(\underline{y}) = \sigma^2 I$ we see that in mixed models more effort goes into modeling the dispersions, while in fixed models more effort goes into modeling the expectations. The fixed model says that if we remove X and Z the errors are homoscedastic, the mixed model says that if we remove X only the errors have factor structure with loadings $Z\Omega Z' + \sigma^2 I$. For diagonal Ω the number of parameters of both models is the same.

There is a voluminous literature on estimating mixed linear models [5]. Most methods are based on the multinormal likelihood. This amounts to minimizing the deviance function

$$\Delta(\beta,\theta) = \log \mathbf{det}\,\Sigma(\theta) + (y - X\beta)'\Sigma^{-1}(\theta)(y - X\beta),$$

where $\Sigma(\theta)$ is a nonnegative linear combination of known positive semi-definite matrices (one of which is the identity). This function is easily minimized, for instance, by alternating minimization over β for fixed θ and minimization of θ for fixed β until convergence. But it is also possible to treat the random coefficients as missing data and apply the EM algorithm, or apply Newton's method to the marginal likleihood obtained by minimizing the full likelihood over β [1]. Models for nonlinear mixed effects models are discussed in [2].

We illustrate the various options with a small dataset taken from the classical paper by Pothoff and Roy [3]. Distances between pituitary gland and pterygomaxillary fissure were measured using x-rays in $n = 27$ children (16 males and 11 females) at $m = 4$ time points, at ages 8, 10, 12, and 14. Data can be collected in a $n \times m$ matrix Y. We also use a $m \times p$ matrix X of the first $p = 2$ orthogonal polynomials on the m time-points.

The first class of models we consider is $\underline{Y} = BX' + \underline{E}$ with B a $n \times p$ matrix of regression coefficients, one for each subject, and with $\underline{E}$ the $n \times m$ matrix of disturbances. We suppose the rows of $\underline{E}$ are independent, identically distributed centered normal vectors, with dispersion Σ. Observe that the model here tells us the growth curves are straight lines, not that the deviations from the average growth curves are on a straight line.

The deviance for this class of fixed coefficient regression models is

$$\Delta(B, \Sigma) = n \log \mathbf{det}\, \Sigma + \mathbf{tr}\, (Y - BX')\Sigma^{-1}(Y - BX)'$$

Within this class of models we can specify various submodels. The most common one supposes that $\Sigma = \sigma^2 I$, in which case the regression coefficients are estimated simply by $\hat{B} = YX$. But many other specifications are possible. We can, on the one hand, require Σ to be scalar, diagonal, or free. And we can, on the other hand, require the regression coefficients to be all the same, the same for all boys and the same for all girls, or free (all different). These are all fixed regression models. The minimum deviances are shown in the first three rows of Table 1. In some combinations the deviance is unbounded below and the minimum does not exist.

	equal	gender	free
scalar	307(3)	280(5)	91(55)
diagonal	305(6)	279(8)	$-\infty$(58)
free	233(12)	221(14)	$-\infty$(64)
random	240(6)	229(8)	$-\infty$(58)

TABLE 1. Mixed Model Fit

We show the results for the simplest case, with the regression coefficients "free" and the dispersion matrix "scalar". The estimated growth curves are in Figure 1. Boys are solid lines, girls are dashed. The estimated σ^2 is 0.85.

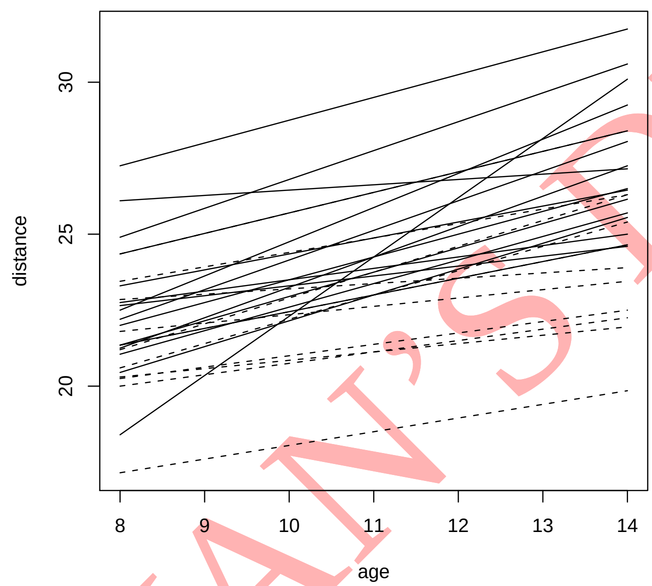

FIGURE 1. Growth Curves for the Free/Scalar Model

We also give the results for the "gender" regression coefficients and the "free" dispersion matrix. The two regression lines are in Figure 2. The regression line for boys is both higher and steeper than the one for girls. There is much less room in this model to incorporate the variation in the data using the regression coefficients, and thus we expect the estimate of the residual variance to be larger. In Table 2 we give the variances and

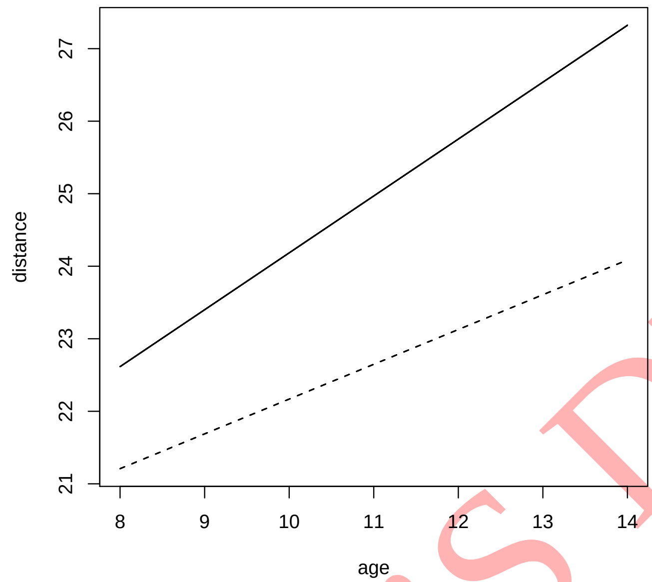

FIGURE 2. Growth Curves for the Gender/Free Model

correlations from the estimated Σ. The estimated correlations between the errors are clearly substantial.

	8	10	12	14
Correlations	1.00			
	0.54	1.00		
	0.65	0.56	1.00	
	0.52	0.72	0.73	1.00
Variances	5.12	3.93	5.98	4.62

TABLE 2. Σ from Gender/Free Model

The general problem with fixed effects models in this context is clear from both the figures and the tables. To make models realistic we need a lot of parameters, and if there are many parameters we cannot expect the estimates to be very good. In fact in some cases we have unbounded likelihoods and the estimates we look for do not even exist. Also, it is difficult to make sense of so many parameters at the same time, as Figure 1 shows.

Next consider random coefficient models of the form $\underline{Y} = \underline{B}X' + \underline{E}$, where the rows of $\underline{B}$ are uncorrelated with each other and with all of $\underline{E}$. By write $\underline{B} = B + \underline{\Delta}$ with $B = \mathbf{E}(\underline{B})$ we see that we have a mixed linear model of the form $\underline{Y} = BX' + \underline{\Delta}X' + \underline{E}$. The deviance now is

$$\Delta(B, \Sigma, \Omega) = n \log \mathbf{det}\,(X\Omega X' + \Sigma) + \\ + \mathbf{tr}\,(Y - BX')(X\Omega X' + \Sigma)^{-1}(Y - BX')',$$

where Ω is the dispersion of the rows of $\underline{\Delta}$. It seems that we have made our problems actually worse by introducing more parameters. But allowing random variation in the regression coefficients makes the restrictive models for the fixed part more sensible. We fit the "equal" and "gender" versions for the regression coefficients B, together with the "scalar" version of Σ, leaving Ω "free".

Deviances for the random coefficient model are show in the last row of Table 1. We see a good fit, with a relatively small number of parameters. To get growth curves for the individuals we compute $\mathbf{E}(\underline{B}|Y)$, which turns out to be

$$\mathbf{E}(\underline{B}|Y) = \tilde{B}[I - \Omega(\Omega + \sigma^2 I)^{-1}] + \hat{B}\Omega(\Omega + \sigma^2 I)^{-1},$$

where $\tilde{B}$ is the mixed model estimate and $\hat{B} = YX$ is the least squares estimate portrayed in Figure 1. Using the "gender" restriction on the regression coefficients the conditional expectations, also known as the *best linear unbiased predictors* or BLUP's [4], are plotted in Figure 3.

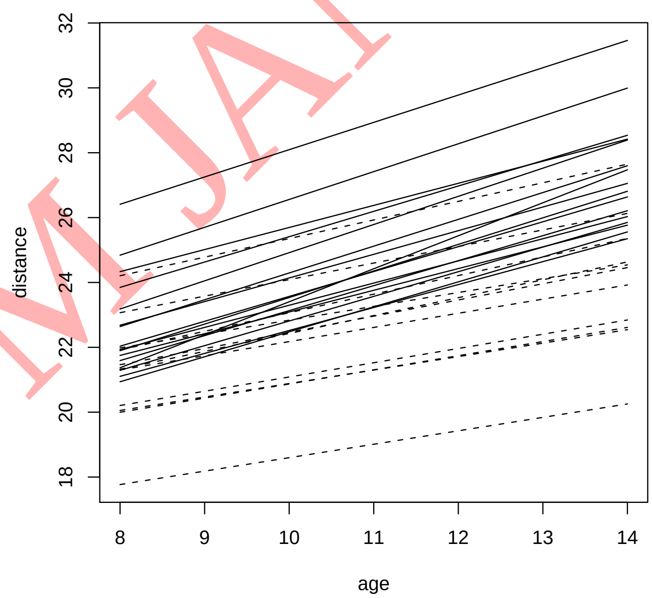

FIGURE 3. Growth Curves for the Mixed Gender Model

We see they provide a compromise solution, that shrinks the ordinary least squares estimates in the direction of the "gender" mixed model estimates. We more clearly see the variation of the growth curves for the two genders around the mean gender curve. The estimated σ^2 for this model is 1.72.

References

[1] M. Lindstrom and D. Bates. Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Replicated Measures Data. *Journal of the American Statistical Association*, 83:1014–1022, 1988.

[2] J.C. Pinheiro and D.M. Bates. *Mixed Effects Models in S and S-Plus*. Springer Verlag, New York, 2000.

[3] R.F. Pothoff and S.N. Roy. A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems. *Biometrika*, 51:313–326, 1964.

[4] G.K. Robinson. That BLUP is a Good Thing: the Estimation of Random Effects (with Discussion). *Statistical Science*, 6:15–51, 1991.

[5] S.R. Searle, G. Casella, and C.E. McCulloch. *Variance Components*. Wiley, New York, NY, 1992.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu