Journal of Modern Applied Statistical Methods

Volume 2 | Issue 2 Article 9

11-1-2003

Random Regression Models Based On The Elliptically Contoured Distribution Assumptions With Applications To Longitudinal Data

Alfred A. Bartolucci
University of Alabama at Birmingham, albartol@uab.edu

Shimin Zheng
University of Alabama at Birmingham

Sejong Bae
University of North Texas

Karan P. Singh
University of North Texas

DOI: 10.22237/jmasm/1067645340

Random Regression Models Based On The Elliptically Contoured Distribution Assumptions With Applications To Longitudinal Data

Cover Page Footnote
The authors acknowledge assistance from Robert H. Lyles with SAS and S-Plus programming; and the HERS Study Group for providing the Human Immunodeficiency Virus (HIV) Epidemiology Research Study data.

Journal of Modern Applied Statistical Methods
November, 2003, Vol. 2, No. 2, 359-370

1538 – 9472/03/$30.00

Random Regression Models Based On The Elliptically Contoured Distribution Assumptions With Applications To Longitudinal Data

Alfred A. Bartolucci, Shimin Zheng
Department of Biostatistics
University of Alabama at Birmingham and
Nanjing Audit University, P. R. China

Sejong Bae, Karan P. Singh
Department of Biostatistics
Health Science Center at Fort Worth
University of North Texas

We generalize Lyles et al.'s (2000) random regression models for longitudinal data, accounting for both undetectable values and informative drop-outs in the distribution assumptions. Our models are constructed on the generalized multivariate theory which is based on the Elliptically Contoured Distribution (ECD). The estimation of the fixed parameters in the random regression models are invariant under the normal or the ECD assumptions. For the Human Immunodeficiency Virus Epidemiology Research Study data, ECD models fit the data better than classical normal models according to the Akaike (1974) Information Criterion. We also note that both univariate distributions of the random intercept and random slope and their joint distribution are non-normal short-tailed ECDs, and that the error term is distributed as a non-normal long-tailed ECD if we don't use the low undetectable limit or half of it to replace the undetectable values. Instead, we use the ECD cumulative distribution function to calculate the contribution to the likelihood due to the undetectable values.

Key words: Generalized multivariate analysis, power exponential distributions, Gamma distributions, maximum likelihood functions, censoring, informative drop-outs, empirical Bayes

Introduction

In clinical studies of human immunodeficiency virus (HIV) infection the number of copies of HIV ribonucleic acid (RNA) per milliliter of plasma is often used to measure the progression of the disease. When the number of copies per milliliter is below or equal to 500, the observation is considered as undetectable, missing, or left-censored, since the copy numbers below 500 are not quantifiable.

Correspondence regarding this article should be emailed to Alfred A. Bartolucci: albartol@uab.edu. The authors acknowledge assistance from Robert H. Lyles with SAS and S-Plus programming; and the HERS Study Group for providing the Human Immunodeficiency Virus (HIV) Epidemiology Research Study data.

On the other hand, illness or death caused by an early drop-out is known as an informative drop-out. If either a left-censored or an informative drop-out is present, as Lyles et al. (2000) pointed out, random effects linear models (Laird & Ware, 1982) and generalized estimating equations (GEE) (Liang & Zeger, 1986) produce biased estimates of key parameters, such as the population average HIV RNA slope and intercept. Louis (1982) used asymptotic approximation methods to deal with the problem of left-censored and informative drop-out data. Both Hughes (1999) and Schluchter (1992) implemented Maximum Likelihood (ML) estimation via Expectation and Maximization (EM) algorithm to handle the problem of left-censored and informative drop-out data. Lyles et al. (2000) combined the approaches of Hughes (1999) and Schluchter (1992) into a single likelihood integrating subject-specific random slopes and intercepts which took both informative drop-out and undetectable data into account. Then, they maximized the likelihood function with respect to fixed effects and other variables. Our

approach follows Lyles et al. (2000) and we extend their normal distribution assumptions to the ECD assumptions since when the number of undetectable observations exceeds a certain number, or when the random intercept and random slope have a bell shaped and long-tailed or short-tailed distribution the ECD distribution improves the fit of the data over the normal distribution.

We used the data from the study of Lyles et al. in this paper. From April 1993 to June 1998 there were 528 HIV-infected women (16-55 years old) in the HIV Epidemiology Research Study (HERS) and 1,864 RNA measurements were collected. Overall, there were 25 (4.7%) drop-out events which resulted in 77 informative drop-out observations, according to Lyles et al.'s (2000) definition.

We used δ as an indicator which was set to 1 if an observation was an informative drop-out and to 0 otherwise. For these 25 individuals the time on study was set as the minimum of the time from the base-line to death or the time from the base-line to 3 months beyond the last visit. For other non-informative drop-out women the censored time was set equal to the time from the base-line to the last visit date. Overall, 745 (40%) out of 1,864 HIV RNA observations were undetectable or left-censored (below 500 copies per milliliter).

Power Exponential Distributions and Models

The power exponential distributions can be used to model both light and heavy tailed, symmetric and unimodal continuous data sets. Gomez et al. (1998) generalized the Univariate Power Exponential (UPE) distribution, which was established by Subbotin (1923), to the Multivariate Power Exponential (MPE) distribution. Both Johnson (1979) and Gomez et al. (1998) discussed the relationship between the UPE distribution and a Gamma distribution. Gomez et al. (1998) studied the properties of MPE intensively, including the stochastic representation, the moments, the characteristic function and the marginal and conditional distributions and asymmetry and kurtosis coefficients. Obviously, the family of MPE distribution is a subset of the class of ECDs. Gomez et al. (1998) defined the MPE distribution as follows:

$$f(y;\mu,\Sigma,\beta)=\frac{n\Gamma(\frac{n}{2})}{\pi^{\frac{n}{2}}\sqrt{|\Sigma|}\Gamma\left(1+\frac{n}{2\beta}\right)2^{\left(1+\frac{n}{2\beta}\right)}}\exp\left(-\frac{1}{2}\left[(y-\mu)'\Sigma^{-1}(y-\mu)\right]^{\beta}\right), \quad (1)$$

where $-\infty<\mu<\infty$, $\Sigma>0$, $0<\beta<\infty$. If y is distributed as an MPE distribution with parameters μ, Σ and β, we write y ~ MPE (μ, Σ, β) and we write y~ UPE (μ, σ, β) if n=1. The parameter β is called the shape parameter.

We use the following linear random-effects regression model (LRRM):

$$y_{ij}=\alpha+a_i+(\beta+b_i)t_{ij}+e_{ij}. \quad (2)$$

We take the response y_{ij} to be the base 10 logarithm of HIV RNA measured at the *jth* time point t_{ij} (j = 1,2,…, n_i) for the *ith* woman (i = 1,…,528, $1 \le n_i \le 5$ for our data set). We assume that the error terms e_{ij} are distributed as UPE (μ, σ^2, ν_1), the random intercept deviations a_i are distributed as UPE (μ, σ_1^2, ν_2) and the random slope deviations b_i are distributed as UPE (μ, σ_2^2, ν_2) with $cov(a_i,b_i)=c\sigma_{12}$ where c is the correction coefficient and ν_2 is a shape parameter. The joint distribution of a_i and b_i is MPE_2 (0, Σ_2, ν_2), where $\Sigma_2=(\sigma_{ij})$. Based on the trivariate normal distribution model (Schluchter, 1992) we assume the 3-dimensional random vector $(a_i,b_i,T_i^0)'$ distributed as trivariate power exponential, i.e.

$$\begin{pmatrix} a_i \\ b_i \\ T_i^0 \end{pmatrix} \sim MPE_3(\mu,\Sigma_3\nu_2),$$

where

$$\mu=\begin{pmatrix} 0 \\ 0 \\ \mu_t \end{pmatrix}, \quad \Sigma_3=\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{at} \\ \sigma_{12} & \sigma_2^2 & \sigma_{bt} \\ \sigma_{at} & \sigma_{bt} & \sigma_t^2 \end{pmatrix}, \quad rk(\Sigma_3)=3.$$

The joint pdf of $(a_i,b_i,T_i^0)'$ is given as

$$f(a_i, b_i, T_i^0) = \frac{3\Gamma(\frac{3}{2})}{\pi^{\frac{3}{2}}\sqrt{|\Sigma_3|}\Gamma\left(1+\frac{3}{2\nu_2}\right)2^{\left(1+\frac{3}{2\nu_2}\right)}}$$

$$\exp\left(-\frac{1}{2}\left[\left(a_i, b_i, T_i^0 - \mu_t\right)\Sigma_3^{-1}\left(a_i, b_i, T_i^0 - \mu_t\right)'\right]^{\nu_2}\right),$$

where T_i^0 is the natural logarithm of the "survival" time for subject *i*.

Maximum Likelihood Functions

In this section we utilize general integrated likelihood expressions given by Lyles et al. (2000), in order to facilitate estimation and inference for the ECD case.

(a) The Maximum Likelihood (ML) function without accounting for undetects and informative drop-outs: By the conditional probability formulae the ML function without accounting for undetects and informative drop-outs is given by

$$L(\theta, Y, T) = \prod_{i=1}^{k}\left[\int_{-\infty}^{\infty}\prod_{j=1}^{n_i} f\left(Y_{ij} \mid a_i, b_i\right) f\left(a_i \mid b_i\right) f\left(b_i\right) da_i db_i\right], \quad (3)$$

where $\theta = (\alpha, \beta, \sigma_1^2, \sigma_2^2, \sigma_{12}, \sigma^2)'$, Y is a vector consisting of Y_{ij}, T is a vector consisting of t_{ij} and

$$f(Y_{ij} \mid a_i, b_i) = \frac{1}{\sigma\Gamma\left(1+\frac{1}{2}\nu_1\right)2^{\left(1+\frac{1}{2}\nu_1\right)}}$$

$$\exp\left(-\frac{1}{2}\Sigma\left[\frac{Y_{ij} - \left[\alpha + a_i + (\beta + b_i)t_{ij}\right]}{\sigma}\right]^{\nu_1}\right),$$

$$f(a_i \mid b_i) f(b_i) = f(a_i, b_i) = \frac{1}{\pi\sqrt{|\Sigma_2|}\Gamma\left(1+\frac{1}{\nu_2}\right)2^{\left(1+\frac{1}{\nu_2}\right)}}$$

$$\exp\left(-\frac{1}{2}\left[\left(a_i, b_i\right)\Sigma_2^{-1}\left(a_i, b_i\right)'\right]^{\nu_2}\right).$$

(b) The ML function accounting for undetectable values only:

We use d to denote the operable limit of detection. We assume that the first n_{i1} measurements are detectable values and there are $n_i - n_{i1}$ undetectable values for subject *i*. We use the probability distribution function (pdf) to calculate the contribution to the likelihood due to the observed values for subject *i*. On the other hand we use the cumulative distribution function (cdf) to calculate the contribution to the likelihood due to the undetectable values. Therefore, the complete-data likelihood function is given by

$$L(\theta, Y) = \prod_{i=1}^{k}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{j=1}^{n_{i1}} f\left(Y_{ij} \mid a_i, b_i\right)\prod_{j=n_{j1}+1}^{n_i} F_Y\left(d \mid a_i, b_i\right) f\left(a_i \mid b_i\right) f\left(b_i\right) da_i db_i\right],$$

where $f(Y_{ij}|a_i,b_i)$ and $f(a_i|b_i)f(b_i)$ are given in (3) and

$$F_Y(d \mid a_i, b_i) = \int_{-\infty}^{b} \frac{1}{\sigma\Gamma\left(1+\frac{1}{2}\nu_1\right)2^{\left(1+\frac{1}{2}\nu_1\right)}}\exp\left(-\frac{1}{2}\left(\frac{y}{\sigma}\right)^{\nu_1}\right)dy$$

$$= \begin{cases} \frac{1}{2}\left[P\left\{u \le \frac{1}{2}\left(\frac{b}{\sigma}\right)^{2\nu_1}\right\} + 1\right], & \text{if} \quad b \ge 0 \\ \frac{1}{2} - \frac{1}{2}P\left\{u \le \frac{1}{2}\left(-\frac{b}{\sigma}\right)^{2\nu_1}\right\}, & \text{if} \quad b < 0 \end{cases} \quad (4)$$

where $b = y_i - [\alpha + a_i + (\beta + b_i)t_i]$, y_i is the censored value for subject *i* and $u \sim \Gamma\left(1, \frac{1}{2\nu_1}\right)$.

(c) The ML function accounting for informative drop-outs only:

We use T_i^0 to denote the natural logarithm of the "survival" time for subject *i* and c_i to denote the natural logarithm of the time from the base-line to the study end. Let $T_i = \min(T_i^0, c_i)$.

i) If subject *i* did not drop out early we have $\delta_i = 0$ and use $1 - F_T(c_i|a_i,b_i)$ to compute the contribution to the likelihood due to the right censored values, where F is the cdf of T given a_i and b_i. That is

$$F_T(c_i \mid a_i, b_i) = \frac{\frac{3}{2}\Gamma\left(\frac{3}{2}\right)\sqrt{|\Sigma_2|}\Gamma\left(1+\frac{1}{\nu_2}\right)}{\sqrt{\pi}\sqrt{|\Sigma_3|}\Gamma\left(1+\frac{3}{2\nu_2}\right)2^{\left(\frac{1}{2\nu_2}\right)}} \qquad (5)$$

$$\int_{-\infty}^{c_i} \exp\left\{ \begin{array}{l} -\frac{1}{2}\left[(a_i, b_i, z-\mu_t)\sum_3^{-1}(a_i, b_i, z-\mu_t)'\right]^{\nu_2} \\ +\frac{1}{2}\left[(a_i, b_i)\begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}^{-1}\begin{pmatrix}a_i\\ b_i\end{pmatrix}\right]^{\nu_2} \end{array} \right\} dz,$$

where Σ_2 , Σ_3 and ν_2 were defined in LRRM.

(ii) If subject i dropped out early we have $\delta_i = 1$ and $T_i = T_i^0$ and use the pdf $f(T_i^0 \mid a_i, b_i)$ to compute the contribution to the likelihood due to the informative drop-out values. Therefore, the likelihood function accounting for informative drop-outs and the right censored data is given by

$$L(\theta, Y, T) = \prod_{i=1}^{n}\left[\begin{array}{l}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(Y_i \mid a_i, b_i) f(T_i^0 \mid a_i, b_i)^{\delta_i} \\ \left[1 - F_T(c_i \mid a_i, b_i)\right]^{1-\delta_i} f(a_i \mid b_i) f(b_i) da_i db_i\end{array}\right], \qquad (6)$$

where $\theta = (\alpha, \beta, \sigma_1^2, \sigma_2^2, \sigma_{12}, \sigma^2, \mu_t, \sigma_{at}, \sigma_{bt}, \sigma_t^2)'$. Thus, the complete ML function is given by

$$L(\theta, Y, T) = \prod_{i=1}^{528}\left[\begin{array}{l}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{j=1}^{n_{i1}} f(Y_{ij} \mid a_i, b_i)\prod_{j=n_{i1}+1}^{n_i} F_Y(d \mid a_i, b_i) \\ f(T_i^0 \mid a_i, b_i)^{\delta_i}\left[1 - F_T(c_i \mid a_i, b_i)\right]^{1-\delta_i} \\ f(a_i \mid b_i) f(b_i) da_i db_i\end{array}\right]. \qquad (7)$$

Computing Empirical Bayes Estimates of Random Intercepts & Random Slopes

In this section we discuss the calculation of empirical Bayes estimates of random intercepts and random slopes in the presence of drop-outs and undetectable values based on the ECD assumptions. Specifically, we calculate the estimate of the random intercept a_i and random slope b_i by substituting the ML estimators of θ based on the ML function (7) developed in the last section into the analytic expressions for the posterior means given the observed data (Y_i, T_i). Specifically, the empirical Bayes estimates of the random intercept a_i and slope b_i for subject i are given, respectively, by

$$\hat{a}_i = E(a_i \mid Y_i, T_i) = f^*(Y_i, T_i, \theta)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_a(a_i, b_i) da_i db_i,$$

where

$$g_a(a_i, b_i) = a_i f^*(Y_i \mid a_i, b_i) f(T_i \mid a_i, b_i) f(a_i \mid b_i) f(b_i),$$

$$\hat{b}_i = E(b_i \mid Y_i, T_i) = f^*(Y_i, T_i, \theta)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_b(a_i, b_i) da_i db_i,$$

where

$$g_b(a_i, b_i) = b_i f^*(Y_i \mid a_i, b_i) f(T_i \mid a_i, b_i) f(a_i \mid b_i) f(b_i).$$

The above empirical Bayes estimates were given by Lyles et al., (2000). Note that $f^*(Y_i|a_i,b_i)$ is different from $f(Y_i|a_i,b_i)$, the one with asterisk indicates that the data vector Y_i may include one or more undetectable values.

Computation

The software package we have used to obtain the ML estimates of variance components and fixed effects corresponding to models discussed in this chapter is SAS PROC IML. The ML function is constructed within PROC IML first. The initial parameter estimates are obtained from Lyles et al. (2000). The ML function is maximized through the NLPQN routine in IML with respect to the parameters stated in this paper. The double integration was computed by quadrature for each subject. The Hessian matrix (the dispersion matrix of the estimated parameters) was found through the NLPFDD routine in IML. There are no built-in generic non-normal ECD functions in SAS. We used the theorems of relationship between a UPE and a Gamma distribution developed in another paper to compute the probability of UPE distribution below or above a certain point. However, this method can not be used to deal with MPE distribution or the conditional and marginal MPE distributions since there is no

existing useful relationship between an MPE and a Gamma distribution and the conditional or the marginal distributions of an MPE are not necessarily MPEs, which can be much more complicated ECD distributions. We used approximation methods to integrate such integrands. The Simpson's rule has been adopted which requires much less computing time and can reach highly accurate results. S-Plus and SAS PROC IML were used to obtain the empirical Bayes estimates of the random intercept and random slope for each subject and the critical values of UPE distribution and the Simpson's rule has also been used for non-normal situations.

Results

We used the Akaike (1974) Information Criterion (AIC) which was used by Lindsey (1999) and among others to compare the classical multivariate normal model and the multivariate power exponential model. In version 8 of SAS/STAT software AIC is defined as 'smaller-is-better'. Specifically, AIC=2l + 2d, where l denotes the maximum value of the log likelihood, d denotes the dimension of the model, i.e., the number of parameters estimated in the ML function. Six models were considered:

Model 1 (M1): In this model we assumed the normal distributions. There were six parameters (α, β, σ_1^2, σ_2^2, σ_{12}, σ^2) estimated in the ML function accounting for undetectable values which were constructed as in equation (2) of Lyles et al. (2000, p.488).

Model 2 (M2): As in model M1, the normal distributions were assumed. There were ten parameters (α, β, σ_1^2, σ_2^2, σ_{12}, σ^2, μ_t, σ_{at}, σ_{bt}, σ_t^2) estimated in the ML function accounting for both undetects and informative drop-outs which were constructed as in equation (5) of Lyles et al. (2000, p.489).

Model 3 (M3): ECDs were assumed in this model. This model accounts for undetectable values only. Furthermore we assumed that two shape parameters were equal, i.e., $v_1 = v_2$. There were seven parameters (α, β, σ_1^2, σ_2^2, σ_{12}, σ^2, v_1) estimated using the ML function.

Model 4 (M4): This model is the same as M3 except that we don't assume $v_1 = v_2$.

Model 5 (M5): ECDs were assumed in this model. Undetectable, informative drop-out and right censored values were considered at the same time in this model. Also, we assume $v_1 = v_2$.

Model 6 (M6): This model is the same as M5 except that we don't assume $v_1 = v_2$.

Next, we summarize what we have found from the HERS data analysis.

(1). *ECDs* fit the data much better than the classical normal distributions.

Among models M1, M3 and M4 we account for undetectable values only. Model M1 is based on the normal distribution assumptions while model M3 and M4 are based on ECD assumptions. The value of AIC changes from 3932.216 to 3928.101 when the model, M3, is used whereas the value reduces to 3908.833 using the model, M4. Overall, model M4 is the best according to the AIC standard if we consider undetectable values only in our analysis.

Among models M2, M5 and M6 we treat undetects, informative drop-outs and right censored observations simultaneously. Model M2 is based on the normal distribution assumptions, but model M5 and M6 are based on the ECD assumptions. Model M5 reduces AIC from 4083.556 of M2 to 4079.746 (see Table 2). Overall, model M6 (4064.791) is the best by AIC standard if we consider all possible situations.

(2). The dispersion matrix of an MPE random vector is proportional to $\sum$ as defined in section 2. Hence, multiplying the ML estimate $\hat{\Sigma}$ by a coefficient we transformed $\hat{\Sigma}$ to the estimated dispersion matrix whose elements are listed in Table 1. As expected, variance and covariance estimates are very close under the six different models. This proportional relationship provides us a short cut to gain the ML estimates. That is, we can get the ML estimate of the dispersion matrix under the normal distribution assumption first and then utilize this estimated dispersion matrix to estimate the shape parameters. This method is very useful and effective, especially when we have a large number of parameters to estimate or when we deal with a very large data set where computing CPU time and memory space are prohibiting. The estimates of the fixed intercept and the fixed slope for all subjects are almost exactly the same

under the six different models. This is because that $\hat{\alpha}$ and $\hat{\beta}$ only involve the data set which is given and the dispersion matrices of random effects and error terms which are invariant under the normal distribution assumptions and the ECD assumptions as we discussed.

(3). The estimates of the shape parameters in Table 2 strongly suggest that we should make the power exponential distribution assumptions instead of classical normal distribution assumptions since our simulations revealed that less than 0.94 or greater than 1.15 shape parameters indicate the distribution departs significantly from the normal distribution at α = 0.05 level. The shape parameter estimates $\hat{\nu}$=0.6574 (S.E.=0.116) under model M3 and $\hat{\nu}$=0.6997 (S.E.=0.099) under model M5 indicate that 40% undetects contribute to a long tailed non-normal distribution. In model M4 and M6 we don't assume v_1= v_2. The estimate of the second shape parameter is $\hat{\nu}_2$=1.8089 (S.E.= 0.490) in model M4 and $\hat{\nu}_2$=1.3706 (S.E.=0.215) in model M6. The shape parameter estimate $\hat{\nu}_2$ in both models M4 and model M6 are much larger than 1 which shows that both univariate distributions of the random intercept and the random slope and their joint distribution are non-normal. They are thin-tailed *ECDs*, concentrated around 0 means.

Possible Extensions

First, power exponential distributions are just a member of larger ECD family. To extend the power exponential distribution assumptions for the models we have discussed is a challenging task and of great interest in both theory and practice. Second, we used approximation methods to compute probability distribution function values at a certain given point and the probability on some interval or within a certain given high dimension rectangle for the non-normal power exponential distributions. The CPU time and memory space required for this kind of task are prohibitive. This highly intensive computing problem will be eased if we could find an exact or asymptotic relationship between distributions (such as non-normal MPEs and marginal or conditional distributions of a non-normal MPE). Third, we have used simulation methods to assess different distributions, like normal or non-normal characteristics as per the shape parameter. If we could construct a statistic related to the shape parameter and get an explicit, exact or asymptotic distribution of the statistic we could do a formal accurate hypothesis testing about the shape parameter of the distribution. This is another challenging task for future research. All source code provided in this paper is in SAS (Appendix).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Inference Theory*.

Fang, K. T., & Zhang, Y. T. (1993). Generalized multivariate analysis. *Science Press*.

Gomez, E., Gomez-Villegas, M. A., & Marin, J. M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics*, A*27*, 589-600.

Harville, D. A. (1977). Maximum likelihood approached to variance component estimation and to related problems. *Journal of the American Statistical Association*, *722* (358), 320 - 338.

Harville, D. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *The Annals of Statistics*, *4*(2), 384-395.

Hughes, J. P. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*, *55*, 625-629.

Laird, N. and Ware, J. (1982). Random-Effects models for longitudinal data. *Biometrics 38*, 963-974.

Lindsey, J. K. (1999). Multivariate elliptically contoured distributions for repeated measurements. *Biometrics*, *55*, 1277 - 1280.

Lyles, R. H., Lyles, C. M., & Taylor, D. J. (2000). Random regression models for human immunodeficiency virus ribonucleic acid data subject to left censoring and informative drop-outs. *Applied Statistics*, *49* (4), 485 - 497.

Lyles, R. H., Lyles, C. M. & Taylor, D. J. (2000). SAS programs simulation data sets. http://www.blackwellpublishing.com/rss/Volumes/Cv49p4.htm.

Schluchter, M. D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine*, Vol. 11, 1861-1870.

Smith, D. K., Warren, D. L., Vlahov, D., Schuman, P, Stein, M. D., Greenberg, B.L., & Holmberg, S. D. (1997). Design and baseline participant characteristics of the human immunodeficiency virus epidemiology research (HER) study: A prospective cohort study of human immunodeficiency virus infection in US women. *American Journal of Epidemiol*ogy, *146*, 459-469.

Table 1. Results from HERS data: ML Estimates.

	α	β	σ_1^2	σ_2^2	σ_{12}	σ^2	μ_t	σ_{at}	σ_{bt}	σ_t^2
M1	2.89	0.058	0.721	0.037	0.061	0.383	-	-	-	-
	(0.033)	(0.016)	(0.088)	(0.008)	(0.022)	(0.023)				
M2	2.88	0.062	0.718	0.039	0.060	0.382	2.32	0.165	0.035	0.298
	(0.050)	(0.016)	(0.088)	(0.008)	(0.022)	(0.023)	(0.158)	(0.062)	(0.022)	(0.096)
M3	2.91	0.058	0.747	0.040	0.050	0.387	-	-	-	-
	(0.057)	(0.017)	(0.149)	(0.009)	(0.015)	(0.076)				
M4	2.89	0.050	0.695	0.044	0.054	0.410	-	-	-	-
	(0.002)	(0.001)	(0.417)	(0.028)	(0.052)	(0.023)				
M5	2.90	0.062	0.833	0.047	0.059	0.383	2.258	0.173	0.042	0.269
	(0.053)	(0.017)	(0.137)	(0.008)	(0.015)	(0.068)	(0.144)	(0.036)	(0.010)	(0.060)
M6	2.90	0.062	0.833	0.047	0.059	0.383	2.258	0.173	0.042	0.269
	(0.053)	(0.017)	(0.137)	(0.008)	(0.015)	(0.068)	(0.144)	(0.036)	(0.010)	(0.060)

Note. Numbers in parentheses are Standard Errors of the corresponding estimates.

Table 2. Results from HERS data: Shape parameter estimates and AIC.

	ν_1	ν_2	d	*-2 log-likelihood*	AIC
M1	-	-	6	3920.216	3932.216
M2	-	-	10	4063.556	4083.556
M3	0.6574 (0.116)	-	7	3914.101	3928.101
M4	0.4694 (0.060)	1.8090 (0.490)	8	3892.833	3908.833
M5	0.6997 (0.099)	-	11	4057.746	4079.746
M6	0.5173 (0.055)	1.3706 (0.215)	12	4040.791	4064.791

Note. Numbers in parentheses are Standard Errors of the corresponding estimates.

Appendix

SAS Program for Taking Left-censored into Account

```
*****************************************************************
Acknowledgments:  The following program was created originally by Dr.
Robert H. Lyles. We have changed his distribution assumptions normal
to ECD and added five nonlinear constraints. We really appreciate Dr.
Lyles's providing this program.
Description: Calculation of the ML estimates of the fixed effects and
the variance matrix.
We assume the underlying distributions are ECDs. Also, we take the
undetectable observations into account under the model described by
the likelihood equation (4) in this paper.
*****************************************************************
data test;
infile '/herscens1.dat';
input obsn  id time nondet response fail survtyrs logsurvt;
*Compute ML estimates via PROC MIXED on complete data (which would not
  be available in practice).  That is, using the actual values for the
  response and all 1864 measurements;

proc mixed data=test method=ml;
     class id;
     model response=time / s ddfm=bw;
     random intercept time /type=un subject=id;
     title2 "ml estimates for full data set (unavailable in
practice)"; run;

data test2;
  set test;
if nondet=1 then do;
  observed=0;
end;
else if nondet=0 then do;
   observed=1;
end;
   label response="Base 10 log HIVRNA value"
         time    ="time of measurement"
         id      ="subject id"
         observed="indicator for whether value was observed"
         fail="indicator for whether subject dropped out"
         survtyrs="Years to dropout"
         logsurvt="Natural log of dropout time";

* Compute ML estimates ignoring left censoring and drop-outs
  using PROC MIXED with random intercept and slope.  These naive
  estimates will be used as starting values for the six parameters
  of the mixed effects model;

proc mixed data=test2 method=ml;
     class id;
```

```
     model response=time / s ddfm=bw;
     random intercept time /type=un subject=id;
     title2 "ml estimates ignoring left censoring and dropouts";
run;

***Create dataset to be read into IML for maximizing likelihood in
Eqn. 2, accounting for left censoring: *;

data test; set test;
  if nondet=1 then do;
   observed=0;
 end;

 else if nondet=0 then do;
   observed=1;
 end;  run;

proc iml worksize=999216000 symsize=999999900;

*************************************************************
* define IML function which will be used to maximize the likelihood
*************************************************************;

start likeli1(parms);

* lower and upper boundaries and stepsize for numerical integration;
   nsteps=31;
a_l   =-5;
a_u   = 5;
step_a=(a_u-a_l)/(nsteps-1);
b_l   = -1.5;
b_u   =  1.5;
step_b=(b_u-b_l)/(nsteps-1);
pi=2*arsin(1);

* variables corresponding to input parameters from vector 'parms';
   sigsq1  =parms[1];  * random intercept effect variance;
sig12   =parms[2];  * covariance between random intercept and slope;
sigsq2  =parms[3];  * random slope effect variance;
sigsq   =parms[4];  * within subject variance;
alpha   =parms[5];  * fixed effect intercept;
beta    =parms[6];  * fixed effect slope;
v       =parms[7];

* determine number of subjects in dataset;
use test;
   read all var {id} into subjects;
   close test;

* compute number of subjects and create vector for each subjects
  contribution to the likelihood;
subjects=ncol(unique(subjects));
```

```
terms=j(subjects,1,.);

* get vector of indicators for observed vs. censored responses for
subject i;
do i=1 to subjects;
   use test;
   read all var {observed} into d_i where (id=i);
   close test;

* number of observations, number of observed values, and number of
censored
  values, respectively, for subject i;
   n_i=nrow(d_i);
   o_i=sum(d_i);
   c_i=n_i-o_i;

* create vectors of censored values and the associated time of
measurement;
   if c_i>0 then do;
      use test;
      read all var {response} into cens_i where (id=i & observed=0);
      read all var {time} into c_time_i where (id=i & observed=0);
      close test;
      end;

* create vectors of observed values and the associated time of
measurement;
   if o_i>0 then do;
      use test;
      read all var {response} into y_i where (id=i & observed=1);
      read all var {time} into time_i where (id=i & observed=1);
      close test;
      end;

 * set initial value for likelihood contribution by subject i to zero;
   func_i=0;

* define quadrature points for numerical integration;
   do a_i=a_l to a_u by step_a;
   do b_i=b_l to b_u by step_b;

      * contribution to likelihood due to observed values for subject
i;
      if o_i=0 then func_i1=1;
         else do;
         t_i1=(y_i-alpha-beta*time_i);
         t_i2=(a_i+b_i*time_i);
         func_i1=(1/(sqrt(sigsq)*gamma(1+0.5/v)*(2##(1+0.5/v)))**o_i)*
                 exp(-0.5*sum(((t_i1-t_i2)##2/sigsq)##v));
         end;
```

```
      * contribution to likelihood due to censored values for subject
i;
      func_i2=1;
      if c_i>0 then
       do j=1 to c_i;
            b=cens_i[j,1]-alpha-a_i-beta*c_time_i[j,1]-
b_i*c_time_i[j,1];
             if b >= 0 then
temp_i2=0.5*(1+probgam(0.5*(b/sqrt(sigsq))**(2*v),(1/(2*v))));
                  else temp_i2=0.5*(1-probgam(0.5*(-
b/sqrt(sigsq))**(2*v),(1/(2*v))));
         func_i2=func_i2*temp_i2;
       end;

      * compute correlation coefficient between intercept and  slope;
      r=sig12/sqrt(sigsq1*sigsq2);

      * compute joint distribution of intercept and slope;
      w=(sigsq1||sig12)//(sig12||sigsq2);
       u=det(w);
      y=inv(w);
       x_i=(a_i||b_i);

      func_i3=(2/(pi*sqrt(u)*gamma(1+1/v)*(2##(1+1/v))))*
                exp(-0.5*(x_i*y*x_i`)##v);

      * compute contribution of subject 'i' to objective function;

      func_i=func_i+(func_i1*func_i2*func_i3*step_a*step_b);
      end;
      end;

* add subject i's contribution to vector of likelihood terms;

   terms[i,1]=func_i;
   end;

* compute -2 log likelihood;

   loglik2=-2*sum(log(terms));
   return(loglik2);

finish likeli1;


*****************************************************************
The following is the main body of the program (which calls the
minimization function, computes the Hessian, etc.)
*****************************************************************
;
* initial estimates from preliminary analysis;
parms={.24 -.012 .015 .201 3.21 .039 1.0};
```

```
* options vector for minimization function;
* matrix of lower (row 1) and upper (row 2) bound contraints on
parameters
   (sigsq1 > 0, sig12 <> 0, sigsq2 > 0, sigsq > 0, alpha <> 0, beta <>
0);
/*  con={1E-5 . 1E-5 1E-5 . .,
     . . . . . .}; */

* The following are five non-linear restrictions;

start c_h(parms);
  c=j(5,1,0.);
  c[1]=parms[1];
  c[2]=parms[3];
  c[3]=parms[4];
  c[4]=parms[1]-(parms[2]##2/parms[3]);
  c[5]=parms[7];
    return(c);
finish c_h;

* call function minimizer in IML;
optn=j(1,11,.); optn[1]=0; optn[2]=3; optn[10]=5; optn[11]=0;
call nlpqn(rc, xres, "likeli1", parms, optn) nlc="c_h";

* create vector of mle's computed using function minimizer;
parms=xres`;

* compute numerical value of Hessian (and covariance matrix) using
  mle's calculated above;
call NLPFDD(crit, grad, hess, "likeli1", parms);
cov_mat=2*inv(hess);
se_vec =sqrt(vecdiag(cov_mat));
print cov_mat se_vec;

*****************************************************;
The following program is used to transform MLE of ECD Sigma matrix
int the variance matrix;

proc iml;
sig1={ 0.221828 -0.014873 0.011839};
sig = 0.141499; a = 2.906699; b = 0.057614;
beta=0.657391;
c1=2**(1/beta)*gamma(2/beta)/(2*gamma(1/beta));
c2=2**(1/beta)*gamma(1.5/beta)/(gamma(0.5/beta));
sig11=c1*sig1;  sig0=c2*sig;
sigma=sig11||sig0||a||b;
print sigma;    /* with ECD    */
sigmaOld={0.720710 -0.060955 0.037333 0.382976 2.886360 0.058335};
print sigmaOld; /* without ECD */
```