The Annals of Statistics
1989, Vol. 17, No. 2, 624–642

ASSESSING NORMALITY IN RANDOM EFFECTS MODELS

BY NICHOLAS LANGE[1] AND LOUISE RYAN[2]

Brown University and Harvard University

When one uses the unbalanced, mixed linear model $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$, $i = 1, \dots, n$ to analyze data from longitudinal experiments with continuous outcomes, it is customary to assume $\boldsymbol{\varepsilon}_i \sim_{\text{ind}} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_i)$ independent of $\boldsymbol{\beta}_i \sim_{\text{iid}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Delta})$, where σ^2 and the elements of an arbitrary $\boldsymbol{\Delta}$ are unknown variance and covariance components. In this paper, we describe a method for checking model adequacy and, in particular, the distributional assumption on the random effects $\boldsymbol{\beta}_i$. We generalize the weighted normal plot to accommodate dependent, nonidentically distributed observations subject to multiple random effects for each individual unit under study. One can detect various departures from the normality assumption by comparing the expected and empirical cumulative distribution functions of standardized linear combinations of estimated residuals for each of the individual units. Through application of distributional results for a certain class of estimators to our context, we adjust the estimated covariance of the empirical cumulative distribution function to account for estimation of unknown parameters. Several examples of our method demonstrate its usefulness in the analysis of longitudinal data.

1. Introduction and results. We develop an approach to the assessment of distributional assumptions for a broad class of random effects models. These models assume dependent and nonidentically distributed observations subject to the presence of multiple random effects for each individual under study and are used in a variety of statistical applications areas [see, for example, Harville (1977), Laird and Ware (1982), Laird, Lange and Stram (1987) and Lange and Laird (1989)]. Even though random effects models are widely used, methods for assessing their goodness of fit are relatively undeveloped.

In ordinary linear regression, one can assess model adequacy through applications of classical goodness of fit procedures to estimated residuals obtained from the fit of an assumed model to observed data. For example, when one fits a normal linear regression model, one can examine *q-q* plots of estimated residuals for evidence of model departure or outlying observations. Assessing goodness of fit for random effects models is more complex, due to their error structures which accommodate correlations between repeated measurements taken on the same individual sampling unit over time or some other metameter. Our goodness of fit procedure uses standardized empirical Bayes estimates of individual random effects (possibly vector-valued) that are linear functions of estimated residuals.

Received July 1987; revised July 1988.

[1]Partially supported by Grant GM29745 from the National Institutes of Health. Work done while this author was an instructor in Applied Mathematics at the Massachusets Institute of Technology.

[2]Partially supported by Grants CA33041 and CA48061 from the National Cancer Institute.

AMS 1980 *subject classifications.* Primary 62J05; secondary 62G30, 62F12, 62P10.

Key words and phrases. Weighted normal plots, longitudinal data analysis, variance and covariance components, empirical Bayes estimation, restricted maximum-likelihood estimation, adjustments for estimated parameters.

The assumed linear patterns for the covariance matrices of the observations, induced by the random effects assumption, allow these standardized empirical Bayes estimates to be independent across individuals; hence classical goodness of fit procedures can be modified to apply to random effects model criticism. Our work applies and extends the ideas of Dempster and Ryan (1985), who, for a simpler one-way random effects model, proposed a weighted normal plot that increases the sensitivity of the classical normal plot to detect certain types of model departure by using appropriately weighted empirical cumulative distribution functions (e.c.d.f.'s).

We establish context in Section 1.1 and describe our method in Section 1.2. We motivate the use of standardized empirical Bayes estimators for assessing model adequacy, define unweighted and weighted e.c.d.f.'s of these quantities and show that when the parameters of the random effects model are known the limiting distribution of these e.c.d.f.'s is that of the Brownian bridge process, with a modification in the weighted case. Accounting for the estimation of unknown parameters requires adjustments to the covariance function of the process, as described in other contexts by Pierce and Kopecky (1979), Loynes (1980), Randles (1982, 1984) and Pierce (1982). Required regularity conditions are essentially those that ensure the asymptotic normality of maximum-likelihood (ML) estimates of unknown parameters. In Section 1.3, we use the limiting distributions of the e.c.d.f.'s to assess the expected variability of the plots, with covariance functions adjusted point-by-point for estimation of parameters. Section 2 gives examples of applications of our method and Section 3 contains derivations of results given in the foregoing sections.

1.1. *Models and notation.* There are several distinct approaches to modeling the dependent error structures present in longitudinal data. One approach assumes that the observations arise by sampling from a continuous-time stochastic process [for example, see Singer and Cohen (1980)]. Alternatively, one may use linear models with error structures that also accommodate dependencies among repeated observations on the same sampling unit, yet in differing ways. Such linear models can be classified according to the assumed pattern, or lack of pattern, of the covariance matrices for the unit observations and fall into the following general classes: (i) those that employ arbitrary covariance matrices possessing no assumed pattern (the general linear model framework), (ii) those that assume an autoregressive error structure (for example, the nonlinear patterned covariance matrices of time-series models) and (iii) those that assume linear patterned covariance matrices. This last category contains the random effects models we discuss. Interested readers may refer to Singer (1985) for an overview of research questions and analysis strategies for longitudinal data and also to Ware (1985) for an overview of the linear model viewpoint.

The general random effects model we study is

$$\underset{t_i\times 1}{\mathbf{y}_i} = \underset{t_i\times p}{\mathbf{X}_i}\ \underset{p\times 1}{\boldsymbol{\alpha}} + \underset{t_i\times r}{\mathbf{Z}_i}\ \underset{r\times 1}{\boldsymbol{\beta}_i} + \underset{t_i\times 1}{\boldsymbol{\varepsilon}_i}, \tag{1}$$

for $i = 1, \dots, n$, where t_i is the number of occasions on which individual i is observed. The matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ are known, nonrandom between-individuals and within-individual design matrices for the fixed effects α and random effects β_i. When the outcomes $\mathbf{y}_i$ are measurements of a continuous random variable, the unbalanced mixed linear model (1) usually assumes $\varepsilon_i \sim_{\text{ind}} \mathscr{N}(\mathbf{0}, \sigma^2\mathbf{I}_i)$ and independent of $\beta_i \sim_{\text{iid}} \mathscr{N}(\mathbf{0}, \Delta)$, where σ^2 and the elements of Δ are unknown variance and covariance components, with Δ arbitrary and $\mathbf{I}_i$ the $t_i \times t_i$ identity matrix. Note that the assumptions of model (1) imply the familiar linear patterned covariance matrices associated with such models,

$$\text{(2)} \qquad \operatorname{cov}(\mathbf{y}_i) = \mathbf{V}_i \equiv \sigma^2\mathbf{I}_i + \mathbf{Z}_i\Delta\mathbf{Z}_i^T, \quad \text{for } i = 1, \dots, n.$$

The popularity of model (1) stems in part from the attractive and often appropriate linear pattern (2) assumed for $\operatorname{cov}(\mathbf{y}_i)$. Also, the assumed random effects β_i have simple interpretations. For example, when the $\mathbf{Z}_i$ consist of a constant column and a column of the times of observation, the random effects correspond to the random intercepts and random slopes specific to each individual.

1.1.1. *Growth curve models.* The general model (1) is a growth curve model with random effects covariance structure when

$$\text{(3)} \qquad \mathbf{X}_i \equiv \mathbf{Z}_i \otimes \mathbf{a}_i^T, \quad \text{for } i = 1, \dots, n,$$

where $\mathbf{a}_i$ is a vector of covariate values for each individual that does not change from occasion to occasion and $\otimes$ is the direct (Kronecker) product, defined such that $\mathbf{A} \otimes \mathbf{B}$ has (g, h)th block $a_{gh}\mathbf{B}$. [Laird, Lange and Stram (1987), pages 98–99, mention that the growth curve model can be written in a slightly more general form.] Lange and Laird (1989) analyzed the balanced and complete case, for which in addition to (3), $\mathbf{Z}_i \equiv \mathbf{Z}$ for all i, and showed that growth curve model specification and parameter estimation involve some relatively simple extensions of basic ideas from ordinary least-squares (OLS) regression and analysis of variance. In Section 2, we give examples of the use of our method with the general model (1) and of its growth curve form in analyses of three data sets.

1.2. *Assessing goodness of fit.* The underdevelopment of goodness of fit theory for random effects models is due in part to the richness of the class of models to which they belong, in contrast to models that assume independent errors of observation both within and between individuals. Assessing goodness of fit for random effects models involves checking the adequacy of assumptions concerning (i) the deterministic component $\mathbf{X}_i\alpha$ and (ii) the random component $\mathbf{Z}_i\beta_i + \varepsilon_i$ for each sampling unit. In this paper we focus on aspect (ii) by developing theory and methods for checking the assumed error structure. In particular, we develop methods that are sensitive to the distributional assumption on the random effects β_i.

1.2.1. *A simplified case.* Consider first the following simplified version of model (1) that assumes a common mean and a single random effect for each individual:

$$y_{ij} = \mu + \beta_i + e_{ij}, \qquad j = 1, \dots, t_i,\ i = 1, \dots, n, \tag{4}$$

where $e_{ij} \sim_{\text{iid}} \mathscr{N}(0, \sigma^2)$ and $\beta_i \sim_{\text{iid}} \mathscr{N}(0, \delta)$. If the random effects β_i were observable, then classical approaches could be used to check their normality. Even though the β_i are not observable, one can "estimate" these unobservable random variables through the use of empirical Bayes techniques. We propose to examine estimates of the β_i suitably standardized to have the standard normal distribution under the assumed model (4). Specifically, we consider empirical Bayes estimators $\hat{\beta}_i$, i.e., the means of the posterior distributions of the β_i given the $\mathbf{y}_i$. We standardize each $\hat{\beta}_i$ by its marginal (or sampling) standard deviation. This standard deviation differs from the posterior standard deviation of $\hat{\beta}_i$, as discussed for example by Harville (1977) and by Laird and Ware (1982) in the context of making inferences on β_i. In our context, the approach used here is consistent with that advocated by Box [(1980), page 384] who argued that "... sampling theory is needed for exploration and ultimate *criticism* of an entertained model in the light of current data, while Bayes' theory is needed for *estimation* of parameters conditional on the adequacy of the entertained model."

For the simple model (4), we show in Section 3 that standardized estimators of the β_i can be written as

$$z_i = \frac{\hat{\beta}_i}{\mathrm{SD}(\hat{\beta}_i)} = \frac{\bar{y}_i - \mu}{\sqrt{\sigma^2/t_i + \delta}}, \tag{5}$$

where $\bar{y}_i = (1/t_i)\sum_{j=1}^{t_i} y_{ij}$ and assuming for the present that μ, σ^2 and δ are known. Thus, for the simple model, our approach reduces to calculating a standardized average residual for each individual. When the model holds, the z_i are a random sample from a standard normal distribution. Let $F_n(x)$ denote the e.c.d.f. of the z_i, that is, let

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x - z_i),$$

where $I(x - z_i) = 1$ if $x \geq z_i$ and 0 otherwise. Also, let $\Phi(x)$ denote the cumulative standard normal distribution. To check the normality assumption on the β_i, one could compare $F_n(x)$ with $\Phi(x)$, its expected value under the assumed model (4), using a *q-q* plot, a *p-p* plot or other classical goodness of fit techniques.

If one is interested specifically in checking the normality assumption on the random effects, a comparison of $F_n(x)$ and $\Phi(x)$ would not be as sensitive as it could be in detecting departures from the assumed model. The estimated random effects $\hat{\beta}_i$ each possess differing variances consisting of two components σ^2/t_i and δ. A goodness of fit procedure for assessing the normality assumption should highlight the contributions of individuals whose component δ is large relative to σ^2/t_i. To achieve this, Dempster and Ryan (1985) proposed a weighted normal plot that compares a weighted e.c.d.f. $F_n^*(x)$ to $\Phi(x)$ using weights that are functions of the variance components σ^2 and δ [Dempster and Ryan (1985), page

845, (4) and (5)]. The importance of accounting for the differing variances of the estimated random effects was perhaps first pointed out by J. W. Tukey, although he noted that the general problem has a long history. Tukey (1974) explored a procedure he called *faceless reflation*, and stated [Tukey (1974), page 125] that his proposed method "... comes at a stage that is still too early for very formal probability models... at a stage where trying out approaches can be of considerable help in guiding later model building."

1.2.2. *An extension to more complex cases.* Extension of the results of Dempster and Ryan (1985) to apply to models for longitudinal data requires accounting for fixed-effects covariates and multiple random effects for each individual. One simple and useful generalization of (5) would be to obtain a "centered" $\bar{y}_i$ for each individual, standardized under the assumptions of model (1), that is, one could use

$$z_i = \frac{\mathbf{1}^T(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha})}{(\mathbf{1}^T\mathbf{V}_i\mathbf{1})^{1/2}}, \quad \text{for } i = 1, \dots, n. \tag{6}$$

A normal plot of the standardized $\bar{y}_i$ given at (6) could provide a broad assessment of the adequacy of model (1). In fact, as a general check, one could examine *any* suitably standardized linear combination of each individual's estimated residuals. However, if one is interested in the normality assumption for a particular random effect, a more natural generalization of (5) is to use standardized empirical Bayes estimators of the $\boldsymbol{\beta}_i$.

For the general model (1), with vectors of covariates and vectors of random effects for each individual, Laird and Ware (1982) showed that an empirical Bayes estimator of $\boldsymbol{\beta}_i$ is

$$\begin{aligned}\hat{\boldsymbol{\beta}}_i &= E[\boldsymbol{\beta}_i|\mathbf{y}_i] \\ &= \boldsymbol{\Delta}\mathbf{Z}_i^T\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha}),\end{aligned}$$

assuming $\boldsymbol{\alpha}$ known, and the marginal (or sampling) covariance of $\hat{\boldsymbol{\beta}}_i$ is given by

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}_i) = \boldsymbol{\Delta}\mathbf{Z}_i^T\mathbf{V}_i^{-1}\mathbf{Z}_i\boldsymbol{\Delta}. \tag{7}$$

We continue to assume, until Section 1.3.1, that all parameters are known.

Toward producing normal plots, there are several approaches to obtaining standardized versions of the $\hat{\boldsymbol{\beta}}_i$. The approach taken in this paper is to examine a *q-q* plot of some linear combination

$$z_i \equiv \frac{\mathbf{c}^T\hat{\boldsymbol{\beta}}_i}{[\mathbf{c}^T\operatorname{cov}(\hat{\boldsymbol{\beta}}_i)\mathbf{c}]^{1/2}}, \quad \text{for } i = 1, \dots, n. \tag{8}$$

Goodness of fit is then assessed by comparing the e.c.d.f. $F_n(x)$ of the z_i defined in (8) with its expected value under model (1). Through appropriate choices of the vector $\mathbf{c}$, the plot can be made sensitive to different types of model departures. In practice, the random effects often have direct interpretations and there is an advantage to treating components of the $\boldsymbol{\beta}_i$ vectors separately. For example, if the model assumes two random effects for each individual, for random intercept and for random slope, one could produce two marginal q-q plots by setting $\mathbf{c}_1 = (1,0)^T$ and $\mathbf{c}_2 = (0,1)^T$. However, the use of such linear combinations, which simply select components of the $\boldsymbol{\beta}_i$, does not account for possible nonzero correlations between these components. In practice, we have found it useful in such cases to produce a set of plots ranging from one marginal to the other by letting $\mathbf{c}_u = (1-u, u)^T$ for some moderate number of values $0 \le u \le 1$. Extensions of such an approach to $r \ge 3$ are obvious, yet may yield too many plots to examine. Use of a method such as projection pursuit [Friedman and Stuetzle (1981) and Huber (1985)] may be helpful in such cases. In Section 2.3, we give an example for $r = 2$ of the use of marginal projections, projections in-between the two marginals and also of projection pursuit.

1.2.3. *Generalized weighted normal plots.* The argument of Dempster and Ryan (1985) applies when the aim is to assess the normality of the random effects in model (1): The normal plot should be weighted to reflect the differing sampling variances of the estimated random effects. Laird and Ware [(1982), (4.1)–(4.6)] gave expressions for obtaining estimates of the variance and covariance components of model (1) using an expectation-maximization (EM) algorithm [Dempster, Laird and Rubin (1977)]. These equations provide a heuristic justification for the choice of weights which we now describe. For estimating Δ, one equates the following unconditional and conditional expectations at each E step,

$$E\left[\sum_{i=1}^{n} \boldsymbol{\beta}_i\boldsymbol{\beta}_i^T|\Delta\right] = E\left[\sum_{i=1}^{n} \boldsymbol{\beta}_i\boldsymbol{\beta}_i^T|\Delta, \mathbf{y}_i\right]. \tag{9}$$

Reexpressing (9), we have

$$\begin{aligned}\sum_{i=1}^{n} \hat{\boldsymbol{\beta}}_i\hat{\boldsymbol{\beta}}_i^T &= \sum_{i=1}^{n} [\Delta - \operatorname{cov}(\boldsymbol{\beta}_i|\mathbf{y}_i)]\\ &= \sum_{i=1}^{n} \operatorname{cov}(\hat{\boldsymbol{\beta}}_i),\end{aligned} \tag{10}$$

where $\operatorname{cov}(\hat{\boldsymbol{\beta}}_i)$ is the marginal covariance given at (7). The latter equality in (10) is obtained by using a conditional variance argument and by noting that $\operatorname{cov}(\hat{\boldsymbol{\beta}}_i|\mathbf{y}_i)$ is independent of $\mathbf{y}_i$ and hence equals $E[\operatorname{cov}(\hat{\boldsymbol{\beta}}_i|\mathbf{y}_i)]$ [Dempster, Rubin and Tsutakawa (1981), page 342].

Consider now the linear combination z_i of the estimated random effects defined at (8). Pre- and postmultiplying (10) by $\mathbf{c}^T$ and $\mathbf{c}$, respectively, implies setting

$$\frac{\sum_{i=1}^{n} w_i z_i^2}{\sum_{i=1}^{n} w_i} = 1,$$

where

$$\begin{aligned}w_i &\equiv \mathbf{c}^T \operatorname{cov}(\hat{\boldsymbol{\beta}}_i)\mathbf{c}\\ &= \mathbf{c}^T \Delta \mathbf{Z}_i^T \mathbf{V}_i^{-1} \mathbf{Z}_i \Delta \mathbf{c},\end{aligned} \tag{11}$$

the marginal variance of $\mathbf{c}^T\hat{\boldsymbol{\beta}}_i$. Thus, the MLE of $\boldsymbol{\Delta}$ can be thought of as that value $\hat{\boldsymbol{\Delta}}$ which sets the weighted variance of the z_i equal to unity for some fixed $\mathbf{c}$. This relation suggests that to assess the normality of a particular set of $\boldsymbol{\beta}_i$, a weighted plot of z_i versus $\Phi^{-1}[F_n^*(z_i)]$ should be used, with

$$(12) \qquad F_n^*(x) \equiv \frac{\sum_{i=1}^n w_i I(x - z_i)}{\sum_{i=1}^n w_i}.$$

Note that when the longitudinal data are balanced and complete, $\mathbf{Z}_i = \mathbf{Z}$ and $\mathbf{V}_i = \mathbf{V}$ for all i. The variances and covariances of the random effects, and therefore the weights, are identical for all individuals in such cases, and thus the unweighted and weighted plots are also identical.

A reviewer has suggested an alternative to the empirical Bayes approach taken in this paper. Consider the following "fixed effects" estimates of the $\boldsymbol{\beta}_i$ given by

$$\tilde{\boldsymbol{\beta}}_i = \left(\mathbf{Z}_i^T\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i^T(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha}), \quad \text{for } i = 1, \dots, n,$$

whose sampling variation under the assumed model (1) is

$$(13) \qquad \operatorname{cov}\left(\tilde{\boldsymbol{\beta}}_i\right) = \sigma^2\left(\mathbf{Z}_i^T\mathbf{Z}_i\right)^{-1} + \boldsymbol{\Delta}.$$

One could use (13) to obtain standardized quantities, as in (8), to construct a normal plot. In such a context, it would be natural to use the inverse variances of the $\tilde{\boldsymbol{\beta}}_i$, with some fixed $\mathbf{c}$, as the w_i in a weighted normal plot, in contrast to our use of direct variances as weights. It is of interest to note that despite their different motivations, these two approaches yield identical results for the simple case described in Section 1.2.1.

1.3. *Distributional properties.* We now derive distributional properties of the unweighted and weighted e.c.d.f.'s defined in the preceding section. Assuming that all parameters are known, it has been shown [see Ross (1983), pages 187–189 and Dempster and Ryan (1985)] that under the assumed model the mean and covariance functions of the unweighted and weighted e.c.d.f.'s $F_n(x)$ and $F_n^*(x)$ are identical to those of the Brownian bridge process, with an adjustment for the covariance function in the weighted case. That is,

$$(14) \qquad E[F_n(x)] = E[F_n^*(x)] = \Phi(x)$$

and, for $x \leq y$,

$$(15) \qquad \operatorname{cov}[F_n(x), F_n(y)] = \Phi(x)[1 - \Phi(y)]/n.$$

For the weighted e.c.d.f., Dempster and Ryan (1985) showed that

$$(16) \qquad \operatorname{cov}[F_n^*(x), F_n^*(y)] = \left(1 + \frac{v}{m^2}\right)\operatorname{cov}[F_n(x), F_n(y)],$$

where m and v are the mean and variance of the weights w_i. The factor v/m^2 is the square of their sample coefficient of variation. Although the weighted plot is more variable than its unweighted counterpart under the assumed model, the weighted plot is more sensitive to certain departures from the modeling assumptions.

1.3.1. *Adjustments for estimated parameters.* Model and data criticism requires acknowledging that in practice $F_n(x)$ or $F_n^*(x)$ are estimated by some $\hat{F}_n(x)$ or $\hat{F}_n^*(x)$, with the unknown parameters replaced by estimates. Under such replacements, the formulas for the mean and covariance functions given at (14)–(16) are incorrect. Dempster and Ryan [(1985), page 849 bottom] adjusted for the estimation of δ assuming $\mu = 0$ and σ^2 known in the context of a simple, one-way comparisons model (4). It was found in such cases that adjustments for parameter estimation reduce the unadjusted naive variance by about 30% for $|x| \approx 1.5$, and that adjusted and unadjusted variances are equal when $x = 0$ and are also equal when $x = \pm\infty$. In Section 2.2 we report analogous behavior of more general adjustments for model (1).

Adjustments for estimated parameters in other contexts have been considered previously by Pierce and Kopecky (1979) for the simple linear regression problem, and more generally by Loynes (1980), Randles (1982, 1984) and Pierce (1982). We demonstrate that similar results apply here. We do not, however, establish weak convergence of $\hat{F}_n(x)$ or of $\hat{F}_n^*(x)$ uniformly in x, as was done for example by Loynes (1980) in a different context, but instead derive limiting distributions at fixed values of x.

We first consider the distribution of the unweighted e.c.d.f. $\hat{F}_n(x)$. Let the row vector

$$\hat{\boldsymbol{\theta}}_n \equiv \left[\hat{\boldsymbol{\alpha}}_n^T, \hat{\sigma}_n^2, \mathrm{uvec}^T(\hat{\Delta}_n)\right]$$

denote an efficient estimate of unknown parameters in model (1). (The "upper vector" operator uvec strings out the upper-triangular part of Δ into a column vector.) In the following, we assume that $\hat{\boldsymbol{\theta}}_n$ is the MLE of $\boldsymbol{\theta}$; we discuss the use of restricted maximum-likelihood (REML) estimates [Patterson and Thompson (1971)] of σ^2 and Δ in Section 1.3.2. Pierce (1982) showed how to derive the limiting distribution of certain types of statistics when efficient estimates have been substituted for unknown parameters. To apply Pierce's result here, we require two conditions.

CONDITION 1. The e.c.d.f. and the parameter estimator are jointly asymptotically normal, with

$$n^{1/2}\begin{bmatrix} [F_n(x) - \Phi(x)] \\ (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \end{bmatrix} \to_{\mathscr{L}} \begin{bmatrix} T \\ \boldsymbol{\Gamma} \end{bmatrix} \sim \mathscr{N}\left(\mathbf{0}, \begin{bmatrix} W_{00} & \mathbf{W}_{01} \\ \mathbf{W}_{01}^T & \mathbf{W}_{11} \end{bmatrix}\right). \tag{17}$$

CONDITION 2. One can find some column vector $\mathbf{a}(x)$ such that

$$\begin{aligned} \hat{T}_n &= n^{1/2}\left[\hat{F}_n(x) - \Phi(x)\right] \\ &= n^{1/2}\left[F_n(x) - \Phi(x)\right] + n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})\mathbf{a}(x) + o_p(1). \end{aligned} \tag{18}$$

Under conditions (17) and (18), one can use Pierce's result to show

$$\frac{n^{1/2}[\hat{F}_n(x) - \Phi(x)]}{[W_{00} - \mathbf{a}^T(x)\mathbf{W}_{11}\mathbf{a}(x)]^{1/2}} \to_{\mathscr{L}} Z \sim \mathscr{N}(0,1). \tag{19}$$

Essentially, (19) implies that to find the limiting distribution of $\hat{F}_n(x)$, one simply substitutes estimated parameters, treating them as known, and then makes an appropriate adjustment to the variance of the process.

From (15) we see that for a single point x, $W_{00} = \Phi(x)[1 - \Phi(x)]$, the unadjusted naive asymptotic variance at x. The matrix $\mathbf{W}_{11}$ in (19) can be obtained from expressions given by Searle [(1956), page 740] and by Harville [(1977), page 326 middle]. Under ML estimation, a consistent estimate of $\mathbf{W}_{11}$ is

$$n \cdot \operatorname{diag}\left([\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}]^{-1}, \left[\frac{1}{2}\operatorname{trace}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_g}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_h}\right)\right]_{g,h}^{-1}\right), \tag{20}$$

for $g, h = 1, \dots, r^*$, where $r^* = 1 + r(r+1)/2$ is the number of variance and covariance components in the model. The matrix $\mathbf{V}$ is the block-diagonal matrix of the $\mathbf{V}_i$ defined previously at (2), i.e., $\mathbf{V} \equiv \operatorname{diag}(\mathbf{V}_1, \dots, \mathbf{V}_n)$. The matrix $\mathbf{X}$ is the matrix of the $\mathbf{X}_i$ stacked below one another, i.e., $\mathbf{X} \equiv (\mathbf{X}_1^T : \cdots : \mathbf{X}_n^T)^T$. The large square brackets in (20) denote an $r^* \times r^*$ matrix with $[\cdot]_{g,h}$ as its g, h element.

Miller (1977) gave results which establish condition (17) in our case; see our Section 3.2. Randles (1982, 1984) provided a method to demonstrate that condition (18) holds, and also to derive the following expression for $\mathbf{a}^T(x)$; see also our Section 3.2. Letting $\gamma = [\gamma_1, \gamma_2, \gamma_3]$ denote a mathematical variable serving in the role of the unknown parameter $\theta = [\alpha^T, \sigma^2, \operatorname{uvec}^T(\Delta)]$, we find

$$\mathbf{a}^T(x) = \frac{1}{n}\sum_{i=1}^{n} \nabla\mu_i(x,\gamma)\Big|_{\gamma=\theta}, \tag{21}$$

where ∇ is the gradient operator, and

$$\mu_i(x,\gamma) \equiv \Phi\left[(xs_i(\gamma) - m_i(\gamma))(\mathbf{p}_i^T(\gamma)\mathbf{V}_i(\theta)\mathbf{p}_i(\gamma))^{-1/2}\right], \tag{22}$$

$$\mathbf{p}_i^T(\gamma) \equiv \mathbf{c}^T\Delta(\gamma_3)\mathbf{Z}_i^T\mathbf{V}_i^{-1}(\gamma), \tag{23}$$

$$m_i(\gamma) \equiv \mathbf{p}_i^T(\gamma)\mathbf{X}_i(\alpha - \gamma_1^T) \tag{24}$$

and

$$s_i(\gamma) \equiv [\mathbf{p}_i^T(\gamma)\mathbf{V}_i(\gamma)\mathbf{p}_i(\gamma)]^{1/2}, \tag{25}$$

where $\Delta(\gamma_3)$ is a matrix such that $\operatorname{uvec}^T(\Delta(\gamma_3)) = \gamma_3$ and $\mathbf{V}_i(\gamma) = \gamma_2\mathbf{I}_i + \mathbf{Z}_i\Delta(\gamma_3)\mathbf{Z}_i^T$. Note that the covariance matrices $\mathbf{V}_i$ in (22) are evaluated at θ, as in (2), and not at γ as in (25). By use of the rules for matrix differentiation given in Section 3.3, we determine that

$$\nabla\mu_i(x,\gamma)|_{\gamma=\theta} = \phi(x)\left[\frac{1}{\sqrt{w_i}}\mathbf{X}_i^T\mathbf{p}_i, \frac{x}{2w_i}\mathbf{p}_i^T\mathbf{p}_i, \frac{x}{2w_i}\mathbf{v}_i\right], \tag{26}$$

where $\phi(x) = d\Phi(x)/dx$, the w_i are as defined in (11), $\mathbf{p}_i = \mathbf{p}_i(\theta)$, $\mathbf{v}_i$ is a $1 \times (r^* - 1)$ row vector whose kth component is

$$v_{ik} \equiv \mathbf{p}_i^T \mathbf{Z}_i \mathbf{D}_k \mathbf{Z}_i^T \mathbf{p}_i$$

and $\mathbf{D}_k$ is the $r \times r$ matrix of 0's and 1's defined by

$$\mathbf{D}_k \equiv \frac{\partial \Delta(\gamma_3)}{\partial \gamma_{3k}}, \quad \text{for } k = 1, \dots, r^* - 1.$$

Pierce [(1982), page 476 middle] provided a natural relation between $\mathbf{a}(x)$ and the asymptotic covariance matrix in (17), namely $\mathbf{a}(x) = -\mathbf{W}_{01}\mathbf{W}_{11}^{-1}$, noting also that $\mathbf{W}_{01}$ is often difficult to compute directly; hence our Section 3.3.

Consider now the limiting distribution of the weighted e.c.d.f. defined at (12). By an approach similar to the preceding argument for the unweighted case, one can demonstrate that

$$\text{(27)} \qquad \frac{n^{1/2}\left[\hat{F}_n^*(x) - \Phi(x)\right]}{\left[W_{00}^* - \mathbf{a}^{*T}(x)\mathbf{W}_{11}\mathbf{a}^*(x)\right]^{1/2}} \to_{\mathscr{L}} Z^* \sim \mathscr{N}(0,1),$$

where

$$W_{00}^* = \left(1 + \frac{v}{m^2}\right)\Phi(x)[1 - \Phi(x)]$$

and

$$\mathbf{a}^{*T}(x) = \left[\sum_{i=1}^{n} w_i\right]^{-1} \sum_{i=1}^{n} w_i \nabla \mu_i(x, \gamma)\bigg|_{\gamma = \theta}.$$

Thus, when efficient estimates have replaced unknown parameters, $n^{1/2}[\hat{F}_n^*(x) - \Phi(x)]$ has also a limiting normal distribution with mean zero and variance equal to the variance of $n^{1/2}[F_n^*(x) - \Phi(x)]$, minus an adjustment which involves a weighted analog to $\mathbf{a}^T(x)$ defined at (21).

1.3.2. *Restricted maximum-likelihood estimation.* Our approach has thus far dealt with the derivation of the limiting distributions of unweighted and weighted e.c.d.f.'s, accounting for the substitution of MLE's for the unknown parameters of the model. In practice, however, REML estimators for the variance and covariance components are often preferred to MLE's, as the latter ignore degrees of freedom lost through estimation of fixed effects and are thus biased downward. REML estimates are obtained through maximization of a likelihood based upon "error contrasts" rather than the full observed data likelihood. Harville [(1974); (1977), pages 324–325] described justifications for the REML approach from both Bayesian and sampling theory viewpoints. Furthermore, Harville (1977) claimed through a sufficiency argument that this approach loses no information and thus REML estimates are efficient in the same sense as are ML estimators. Hence, the results of the previous section also apply when REML estimates of the variance and covariance components have been substituted for unknown parameters.

2. Applications of results. In this section we apply our method to three data sets. There is no natural ordering of the repeated measures taken on the sampling units in the first two examples, whereas the third example contains a time metameter and is longitudinal in nature. The $\hat{\sigma}^2$ and $\hat{\Delta}$ used in all of our examples to obtain in the $\hat{\beta}_i$ are the REML estimates, computed by the methods and software described by Laird, Lange and Stram (1987). Computations and graphics in our examples were obtained through use of the new release of the S system running on a Sun 3/160 workstation. The projection pursuit example was obtained from a version of ISP running on an IBM/AT. Figures 1–3 display generalized weighted normal plots of the z_i defined at (8) and include bands that are ± 1 SD, adjusted point-by-point for estimation of parameters. The bands on the plots are not to be interpreted as "confidence bands"; they have been included only as aid for assessing the expected variability of the plots. Bickel and Doksum [(1977), pages 381–383] have provided general guidelines for the interpretation of q-q plots.

2.1. *Example*: *Boston housing data.* Harrison and Rubinfeld (1978) reported on a study of housing prices in the Boston Standard Metropolitan Statistical Area and their study provides our first example of a generalized weighted normal plot. For a description of these data, see also Belsley, Kuh and Welsch [(1980), pages 229–261], who used an OLS regression model ($r = 0$, no random effects) and also a robust regression model in their analysis. We fit an unbalanced random effects model to the Boston data in order to investigate how our plot performs in this case. Following Belsley, Kuh and Welsch (1980), we fit $p = 14$ fixed effects. It seems natural to assume that housing prices would be clustered according to the towns represented in the Boston sample. Thus, we treat *town* as the sampling unit, *census tracts* within towns as the repeated measures and assume a single random intercept ($r = 1$) for each of the $n = 92$ towns. Figure 1 shows a generalized weighted normal plot of the z_i defined at (8) for the Boston data, with adjusted ± 1 SD bands. The weights w_i vary according to the differing numbers of census tracts per town (from 1 for Cohasset to 30 for Cambridge), as well as according to the two variance components in the model.

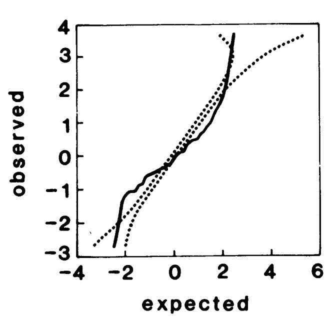

FIG. 1. *Generalized weighted normal plot with adjusted* ± 1 *SD bands for the Boston housing data.*

The ratio of $\max\{w_i\}$ to $\min\{w_i\}$ is only about 5/3. The sample estimate of the coefficient of variation of the weights is $(v/m^2)^{1/2} \approx 7/50$, and it is this low value that accounts for the similarity in shape of the weighted plot to its unweighted counterpart (not shown). We notice familiar indications of heavy tails, as the plot is concave for $z_i < 0$ and convex for $z_i > 0$. It is not surprising that Figure 1 exhibits a violation of the normality assumption for the β_i. The z_i are linear combinations of Studentized residuals from an OLS fit and a classical normal probability plot of such residuals also exhibits "very substantial departures from normality" [Belsley, Kuh and Welsch (1980), Exhibit 4.24, page 233]. We include this example to show the results of fitting a random effects model to these data and also to show how the departure from the normality assumption can be assessed by our method.

2.2. *Example*: *A teratology experiment.* Hartsfield (1986) examined the effects of a common treatment for epilepsy (phenytoin) on the weights of laboratory mice at birth. Fifty-four female mice were divided into three groups, one group receiving the drug, and weights were recorded for 401 of their offspring, in litter sizes varying between 1 and 10. A random effects model was fit to these

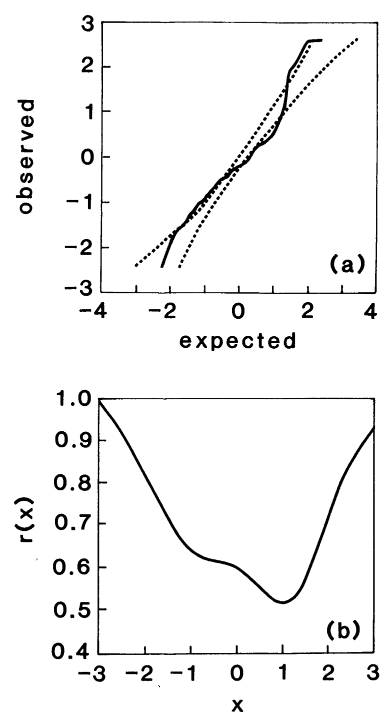

FIG. 2. (a) *Generalized weighted normal plot with adjusted* ± 1 *SD bands for the teratology data.* (b) *The effect of adjustment for estimated parameters in the teratology data: A plot of* $r(x) = [(W_{00}^* - \mathbf{a}^{*T}(x)\mathbf{W}_{11}\mathbf{a}^*(x))/W_{00}^*]^{1/2}$.

data with $p = 4$ fixed covariates (group indicators for the mothers and a gender indicator for the offspring). To allow for correlation within the litters (so-called "litter effects"), the model included a single random intercept term ($r = 1$). Figure 2(a) is a generalized weighted normal plot for these data, with adjusted bands. The normality assumption for the distribution of random effects seems a reasonable one. As in the Boston housing data example, the sample estimate of the coefficient of variation of the weights is small, with $(v/m^2)^{1/2} \approx 1/10$.

We use this example to demonstrate the importance of adjusting the variance of $\hat{F}_n^*(x)$ for estimation of parameters. Figure 2(b) shows a plot of x versus

$$r(x) = \left[\frac{W_{00}^* - \mathbf{a}^{*T}(x)\mathbf{W}_{11}\mathbf{a}^*(x)}{W_{00}^*} \right]^{1/2}$$

for the teratology data and shows that the reduction in variance can be quite substantial, reducing the width of the naive ± 1 SD bands by up to about 50% near the center of the plot. The asymmetry of Figure 2(b) about $x = 0$ is due to the influence of fixed-effects covariates.

2.3. *Example*: *AIDS data.* Incomplete serial measurements of immune function (T-helper cell counts) were obtained by the San Francisco Men's Health Study over a period of about three years for $n = 425$ patients infected with the AIDS virus (HIV). In this preliminary analysis, we fit a growth curve model with no between-individuals covariates ($\mathbf{X}_i \equiv \mathbf{Z}_i$ for all i) and with random intercept and random slope effects ($r = 2$). Figure 3(a) shows a bivariate scatterplot of estimated random intercept effects versus estimated random slope effects. The line at about $-5°$ consists of points for the 59 individuals who were measured on only one occasion and is thus somewhat of an artifact of the incomplete sample. Figures 3(b) and (c) display generalized weighted normal plots of the estimated random intercept and slope effects, respectively, standardized by the estimated variances of these marginal projections. Figure 3(b) exhibits an overall convex shape indicative of a distribution for the random intercept effects with a right tail heavier than the normal, whereas Figure 3(c) indicates that the normality assumption for the random slopes is adequate. In addition, a set of nine plots (not shown), ranging between the two marginals by letting $\mathbf{c}_u = (1 - u, u)^T$ for $u = 0:1/8:1$, indicated that the overall convexity of the weighted normal plot for intercepts is persistent, up until the final, marginal plot of the slope effects.

To examine further the apparent nonnormality of the random intercept effects, we used projection pursuit to reveal the most interesting, i.e., nonnormal, projections of the 425×2 matrix of estimated random effects. We chose a projection index which is an asymptotic equivalent of a χ^2 test for normality [cf. Huber (1985), pages 445–446], using 20 bins of equal expected occupancy. We selected the projection $\hat{\mathbf{c}} = (0.985, 0.167)^T$ that corresponded to a large value of this projection index, which is a rotation of about 10°, and display the resulting generalized weighted normal plot in Figure 3(d). (We ignored a projection nearly orthogonal to the line in the scatterplot with a slightly larger index value which revealed an expectedly peaked distribution, but not nonnormality.) Note that

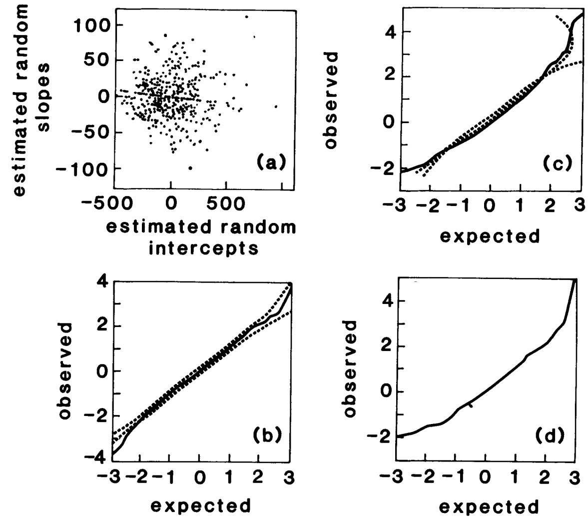

FIG. 3. (a) *A bivariate scatterplot of estimated random intercept and slope effects for the AIDS data. Generalized weighted normal plot for the AIDS data*: (b) *intercept projection* $\mathbf{c} = (1,0)^T$; (c) *slope projection* $\mathbf{c} = (0,1)^T$; (d) *projection pursuit* $\hat{\mathbf{c}} = (0.99, 0.17)^T$.

the use of projection pursuit results in a data-dependent choice of $\mathbf{c}$ and thus invalidates our distributional results; hence the $\hat{\mathbf{c}}$ and the absence of adjusted bands in Figure 3(d). It is also of interest to note that the rotation chosen by projection pursuit is very close to the $\mathbf{c} = (1,0)^T$ marginal intercept projection, as could perhaps be expected since r is so small.

3. Derivations.

3.1. *The standardized empirical Bayes estimator in the simple case.* For a simple, one-way random effects model (4), the covariance matrices $\mathbf{V}_i$ of the observations each possess an assumed compound symmetric pattern, that is,

$$\mathbf{V}_i = \sigma^2\mathbf{I}_i + \delta\mathbf{J}_i,$$

where $\mathbf{J}_i$ is a $t_i \times t_i$ matrix of 1's. By Schur's binomial inverse theorem [cf. Rao (1973), page 33] we rewrite $\mathbf{V}_i^{-1}$ as

$$\mathbf{V}_i^{-1} = \sigma^{-2}\left[\mathbf{I}_i - \frac{\delta}{t_i\delta + \sigma^2}\mathbf{J}_i\right].$$

Assuming all parameters known, we have

$$\hat{\beta}_i = \left[\frac{\delta}{\sigma^2/t_i + \delta}\right]\frac{1}{t_i}\sum_{j=1}^{t_i}(y_{ij} - \mu)$$

and

$$\mathrm{var}(\hat{\beta}_i) = \frac{\delta^2}{\sigma^2/t_i + \delta},$$

which together yield (5).

3.2. *Adjustments for estimated parameters.* Condition (17) requires the joint asymptotic normality of $n^{1/2}[F_n(x) - \Phi(x)]$ and $n^{1/2}(\hat{\theta}_n - \theta)$. The asymptotic normality of $n^{1/2}[F_n(x) - \Phi(x)]$ has been well established [for example, see Ross (1983), page 188]. For our class of mixed linear models, the consistency and asymptotic normality of $n^{1/2}(\hat{\theta}_n - \theta)$ has been established by Miller (1977). When we rewrite (1) in Miller's form, as follows, it is a straightforward procedure to check that the conditions required to apply his main result [Miller (1977), Assumptions 2.1–2.6, 3.1–3.5 and Theorem 3.1, pages 748–752] are satisfied by the sequence of models defined by (1) as the number of individual units $n \to \infty$, yet as $t_i < \infty$ for all i. We also assume that the number of fixed effects p remains constant, so that rank($\mathbf{X}$) is clearly bounded as $n \to \infty$.

Stack the $\mathbf{y}_i$, $\mathbf{X}_i$ and the $\boldsymbol{\varepsilon}_i$ below one another, and the $\mathbf{Z}_i$ in a block-diagonal matrix, to rewrite (1) as

$$\underset{N\times 1}{\mathbf{y}} = \underset{N\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\alpha}} + \underset{N\times nr}{\mathbf{Z}}\ \underset{nr\times 1}{\boldsymbol{\beta}} + \underset{N\times 1}{\boldsymbol{\varepsilon}},$$

where $N = \sum_{i=1}^{n} t_i$, $\mathbf{Z} = \mathrm{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_n^T)^T$, and rewrite $\mathbf{Z}\boldsymbol{\beta}$ as

$$\mathbf{Z}\boldsymbol{\beta} = \sum_{k=1}^{r} \mathbf{U}_k \mathbf{b}_k,$$

where

$$\underset{N\times n}{\mathbf{U}_k} \equiv \mathbf{Z}\mathbf{J}_k, \qquad \underset{n\times 1}{\mathbf{b}_k} \equiv \mathbf{J}_k^T\boldsymbol{\beta} \quad \text{and} \quad \underset{nr\times n}{\mathbf{J}_k} \equiv \mathbf{I}_n \otimes \underset{r\times 1}{\mathbf{j}_k},$$

with $\mathbf{j}_k$ a column of 0's, except for a 1 in its kth slot, $k = 1, \dots, r$. Making these replacements, we may reexpress our model (1) in Miller's form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \sum_{k=1}^{r} \mathbf{U}_k \mathbf{b}_k + \boldsymbol{\varepsilon}. \tag{28}$$

Suitable regularity conditions on the metameter measurements may be applied to write the $\mathbf{U}_k$ of (28) as sums of matrices whose elements are either 0 or 1, with exactly one 1 in each row and at least one 1 in each column, to comply with Assumption 2.6 of Miller (1977). The joint asymptotic normality of $n^{1/2}[F_n(x) - \Phi(x)]$ and $n^{1/2}(\hat{\theta}_n - \theta)$ may then be established by expressing $(\hat{\theta}_n - \theta)$ in terms of efficient scores and applying a multivariate central limit theorem. Miller [(1977), page 757] has given conditions under which $\hat{\theta}_n$ is an asymptotically efficient estimator for θ.

We use Theorem A.9 of Randles (1984) in order to verify that condition (18) holds in our context and to derive expression (21) for $\mathbf{a}^T(x)$. Randles' proof exploits the fact that the e.c.d.f. is a U-statistic, with the indicator $I(x - z_i)$

serving as its kernel function. His general result concerns the limiting distribution of a U-statistic when estimated parameters replace unknown quantities. The conditions necessary for Randles' theorem to apply are satisfied in our case so long as the $\mathbf{X}_i$ and the $\mathbf{Z}_i$ are bounded, with the $\mathbf{Z}_i$ also bounded away from 0. In our context, Randles' theorem implies that

$$n^{1/2}[\hat{F}_n(x) - \Phi(x)] - n^{1/2}[F_n(x) - \Phi(x)] - n^{1/2} \sum_{i=1}^{n} \nabla\mu_i(x, \gamma)\Big|_{\gamma=\theta} (\hat{\theta}_n - \theta)^T \to_p 0,$$

where $\nabla\mu_i(x, \gamma)|_{\gamma=\theta}$ is the gradient vector of

$$\mu_i(x, \gamma) = E_\theta[I(x - z_i(\gamma))]$$

evaluated at θ. The scalar $z_i(\gamma)$ represents a standardized estimator of the unobserved random effect β_i viewed as a function of a mathematical variable γ, that is,

$$z_i(\gamma) = \frac{\mathbf{p}_i^T(\gamma)(\mathbf{y}_i - \mathbf{X}_i\gamma_1^T)}{s_i(\gamma)},$$

with $\mathbf{p}_i^T(\gamma)$ and $s_i(\gamma)$ as defined at (23) and (25). When model (1) holds, $z_i(\gamma)$ has a normal distribution with mean

$$E_\theta[z_i(\gamma)] = \frac{m_i(\gamma)}{s_i(\gamma)},$$

where $m_i(\gamma)$ is as defined at (24), and with variance

$$\operatorname{var}_\theta[z_i(\gamma)] = \frac{\mathbf{p}_i^T(\gamma)\mathbf{V}_i(\theta)\mathbf{p}_i(\gamma)}{s_i^2(\gamma)}.$$

Note that when $\gamma = \theta$, $z_i \sim \mathcal{N}(0,1)$. It follows that

$$\mu_i(x, \gamma) = \Pr\{z_i(\gamma) \le x\} = \Phi\left[\frac{x s_i(\gamma) - m_i(\gamma)}{(\mathbf{p}_i^T(\gamma)\mathbf{V}_i(\theta)\mathbf{p}_i(\gamma))^{1/2}}\right],$$

which is expression (22).

For the weighted case, one can apply the approach given by Randles (1982) to show that

$$n^{1/2}[\hat{F}_n^*(x) - \mu_n^*(x, \hat{\theta}) - F_n^*(x) + \Phi(x)] \to_p 0,$$

where

$$\mu_n^*(x, \gamma) = \frac{\sum_{i=1}^n w_i \mu_i(x, \gamma)}{\sum_{i=1}^n w_i}$$

and $\mu_i(x, \gamma)$ is as defined previously. Given this result, it is straightforward to establish that

$$n^{1/2}[\hat{F}_n^*(x) - \Phi(x)] - n^{1/2}[F_n^*(x) - \Phi(x)] - n^{1/2}(\hat{\theta}_n - \theta)\mathbf{a}^*(x) \to_p 0.$$

3.3. *Vector and matrix derivatives.* For convenience, we include the following list of algebraic rules for matrix differentiation [all of which can be found, for example, in Rogers (1980), Chapters 6 and 7]. Application of these rules and evaluation of

$$\nabla\mu_i(x,\gamma) = [\partial\mu_i/\partial\gamma_1,\ \partial\mu_i/\partial\gamma_2,\ \partial\mu_i/\partial\gamma_3]$$

at $\gamma = \theta$ yields expression (26). We list general versions of such rules, which in essence retain the array structure of element-by-element derivatives and are required when analyzing more complex versions of random effects models than those used in the preceding examples.

RULE 1. For constant matrices $\mathbf{A}$ and $\mathbf{B}$ and any conformable matrix $\mathbf{X}_{m\times n}$

$$\partial(\mathbf{AX})/\partial\mathbf{X} = \mathrm{vec}(\mathbf{A}^T)\mathrm{vec}^T(\mathbf{I}_n),$$

$$\partial(\mathbf{XB})/\partial\mathbf{X} = \mathrm{vec}(\mathbf{I}_m)\mathrm{vec}^T(\mathbf{B})$$

and

$$\partial(\mathbf{AXB})/\partial\mathbf{X} = \mathrm{vec}(\mathbf{A}^T)\mathrm{vec}^T(\mathbf{B}).$$

RULE 2. For a column vector $\mathbf{x}$ and a symmetric, constant matrix $\mathbf{A}$,

$$\partial(\mathbf{x}^T\mathbf{A}\mathbf{x})/\partial\mathbf{x}^T = 2\mathbf{x}^T\mathbf{A}.$$

We have the following "chain rule" for matrices.

RULE 3. For w a scalar function of a matrix $\mathbf{Y}$ which is a function of a matrix $\mathbf{X}$,

$$\partial w/\partial\mathbf{X} = (\partial w/\partial\mathbf{Y}) \star \partial\mathbf{Y}/\partial\mathbf{X}.$$

The operator $\star$ is the "star product" [cf. Rogers (1980), page 26] defined for $\mathbf{X}_{m\times n}$ and the partitioned matrix $\mathbf{Y}_{mp\times nq}$ as

$$\mathbf{X}\star\mathbf{Y} \equiv \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}\mathbf{Y}_{ij},$$

each $\mathbf{Y}_{ij}$ being of dimension $p \times q$. We have the following special cases of the "star product":

RULE 4. When $\mathbf{X}$ and $\mathbf{Y}$ are of the same dimension ($p = q = 1$),

$$\mathbf{X}\star\mathbf{Y} = \mathrm{trace}(\mathbf{X}^T\mathbf{Y}).$$

RULE 5.

$$\mathbf{Y}\star\mathrm{vec}(\mathbf{X})\mathrm{vec}^T(\mathbf{Z}^T) = \mathbf{XYZ}.$$

4. Conclusions. We have motivated the use of generalized weighted normal plots as a simple graphical analysis method for assessing goodness of fit for a class of models for longitudinal data. Our approach involves plotting standardized empirical Bayes estimators of unobservable random effects. We have not proposed any formal goodness of fit test, but instead have proposed pointwise bands of ± 1 SD, adjusted for estimation of parameters, as a guide to the expected variability of the plots given the model and sample size.

Our results show the importance of adjustments for estimation of unknown parameters when assessing the expected variability of the plots. In the examples considered, we have found the reduction in the width of naive bands to be as large as 50% near the center of the plot. We have also stressed the theoretical importance of using a generalized weighted normal plot that accommodates the differing variances of estimated random effects. Although the empirical difference between weighted and unweighted plots can often be small, we advise the use of a weighted plot whenever possible. We also recommend the use of our approach to justify nonnormal or nonparametric models, the need for transformations and/or robust estimation techniques for longitudinal data problems.

Acknowledgments. We thank several anonymous reviewers and an Associate Editor for their very helpful suggestions. We also thank Murray Aitkin for his suggestion to use the Boston data example, Peter Huber for his help with the projection pursuit example, Robert Gray and Nan Laird for their advice and comments on the paper, Burton Singer for bringing the work of Tukey (1974) to our attention, the San Francisco Men's Health Study (contract NO1-AI-32519) for use of their data and the Statistics Center, Massachusetts Institute of Technology, for use of their computing resources.

REFERENCES

BELSLEY, D. A., KUH, E. and WELSCH, R. E. (1980). *Regression Diagnostics: Identifying Influential Data.* Wiley, New York.

BICKEL, P. J. and DOKSUM, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, San Francisco.

BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modeling and robustness (with discussion). *J. Roy. Statist. Soc. Ser. A* **143** 383–430.

DEMPSTER, A. P. and RYAN, L. M. (1985). Weighted normal plots. *J. Amer. Statist. Assoc.* **80** 845–850.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood with incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

DEMPSTER, A. P., RUBIN, D. B. and TSUTAKAWA, R. K. (1981). Estimation in covariance components models. *J. Amer. Statist. Assoc.* **76** 341–353.

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

HARRISON, D. and RUBINFELD, D. L. (1978). Hedonic prices and the demand for clean air. *J. Environ. Econom. Manage.* **5** 81–102.

HARTSFIELD, J. K., JR. (1986). Phenytoin embryopathy: The effect of epoxide hydrolase inhibitor on chronic phenytoin exposure *in utero* in C57BL-6J mice. M.M.Sc. thesis, Harvard Univ.

HARVILLE, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61** 383–385.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–340.

HUBER, P. J. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13** 435–525.

LAIRD, N. M. and WARE, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38** 963–974.

LAIRD, N., LANGE, N. and STRAM, D. (1987). Maximum likelihood computations with repeated measures: Application of the EM algorithm. *J. Amer. Statist. Assoc.* **82** 97–105.

LANGE, N. and LAIRD, N. M. (1989). The effect of covariance structure on variance estimation in balanced growth curve models with random parameters. *J. Amer. Statist. Assoc.* **84** 241–247.

LOYNES, R. M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Statist.* **8** 285–298.

MILLER, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746–762.

PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58** 545–554.

PIERCE, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10** 475–478.

PIERCE, D. A. and KOPECKY, K. J. (1979). Testing goodness of fit for the distribution of errors in regression models. *Biometrika* **66** 1–5.

RANDLES, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10** 462–475.

RANDLES, R. H. (1984). On tests applied to residuals. *J. Amer. Statist. Assoc.* **79** 349–354.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

ROGERS, G. S. (1980). *Matrix Derivatives.* Dekker, New York.

ROSS, S. M. (1983). *Stochastic Processes.* Wiley, New York.

SEARLE, S. R. (1956). Matrix methods in variance and covariance components analysis. *Ann. Math. Statist.* **27** 737–748.

SINGER, B. (1985). Longitudinal data analysis. *Encyclopedia of Statistical Sciences* **5** 142–155. Wiley, New York.

SINGER, B. and COHEN, J. E. (1980). Estimating malaria incidence and recovery rates from panel surveys. *Math. Biosci.* **49** 273–305.

TUKEY, J. W. (1974). Named and faceless values: An initial exploration in memory of Prasanta C. Mahalanobis. *Sankhyā Ser. A* **36** 125–176.

WARE, J. H. (1985). Linear models for the analysis of longitudinal studies. *Amer. Statist.* **39** 95–101.

DEPARTMENT OF COMMUNITY HEALTH
DIVISION OF BIOLOGY AND MEDICINE
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02912

DEPARTMENT OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
HARVARD UNIVERSITY
BOSTON, MASSACHUSETTS 02115