

Random Location and Scale Effects: Model Building Methods for a General Class of Models

William S. Cleveland Lorraine Denby Chuanhai Liu
Bell Labs Avaya Labs Bell Labs
Murray Hill, NJ Basking Ridge, NJ Murray Hill, NJ

Abstract

We describe an approach to building models for data with random effects. The modeling includes both random location effects and random scale effects. The approach consists of a sequence of steps. In each step, model building tools based on certain statistics are used to study the structure of the data and identify or check an aspect of the model; the results of most of the steps are used explicitly in subsequent steps. This paper presents the properties of the statistics for a general class of location and scale models, and briefly describes the sequence and the tools. The paper is a supporting document for (Cleveland, Denby, and Liu 2000), which introduces the approach and describes by example how it is used in practice for model building.

1 Introduction

Models used in practice for data with random effects — both Bayesian and frequentist models — are often complex. Added to the structure of the fixed effects, which by itself can be complex, are the assumptions about the random-effects distributions. The model complexity has impeded progress in the development of model building tools. For example, it is commonly the case in practice that random-effects distributions are simply assumed and not checked.

In applications, random-effects components of models describe the variation in *units*. For example, in an experiment in rat growth, the units are the rats. In almost all applications, unit variability is modeled by random location effects. However, scales can vary across units as well, and often do. Still, random scale variation is seldom modeled. A small number of examples may be found in (Lindley 1971; Leonard 1975; Brady 1985; Cox and Solomon 1986; James, Venables, Dry, and Wiskich 1994; Chinchilli, Esinhart, and Miller 1995; Nusser, Carriquiry, Dodd, and Fuller 1996; Johnson 1997; Lin, Raz, and Harlow 1997; Clark, Cleveland, Denby, and Liu 1999; Cleveland, Denby, and Liu 2000). Responses treated in these references are subjective-scale rating data, university course grades, food nutrient intakes, lengths of menstrual cy-

cles, fabric breaking strengths, heart pulse rates, respiration rates in soybean mitochondria, and serum cholesterol concentrations. But it seems likely that there are random scale effects in many other types of data.

Cleveland, Denby, and Liu (2000) present an approach to model building that consists of a sequence of steps in which the results for each step are used in subsequent steps. They treat models with random scale effects as well as location effects. In the context of examples, they present motivation, background, the sequence of steps, and the tools used in each step. The tools depend on statistics derived from the data, and in the paper, the statistics and their properties are presented for the specific models in the examples.

This paper is a supporting document for (Cleveland, Denby, and Liu 2000). We present a general class of models and give the statistics and their properties for the class. We also briefly describe the sequence and the tools, but there are no examples. Section 2 presents the general model. Section 3 presents formulas on which the model building tools are based. Section 4 presents the model building sequence; it contains references to Section 3 but we suggest a cursory reading before reading Section 3, and a more detailed reading after. Overall, this paper gives an efficient presentation of the methods, and sets out needed formulas, but is quite dry.

2 A General Model for Data with Random Location and Scale Effects

The discussion here is carried out for a general class of random-effects models with normal errors. The class provides adequate fits to a wide class of data sets with responses measured on a continuous scale. The normality of the errors could be replaced by another distribution but would require considerations beyond those of this paper.

Suppose that r_{uj} for $u = 1$ to m and $j = 1$ to n_u are measurements of a response. Each unit u has n_u observations of the response. Suppose there are k_α fixed-effects explanatory variables and k_β random-effects explanatory variables. Let $v_{(k)uj}$ be the measurements of the k -th fixed-effects vari-

able and let $w^{(k)}_{uj}$ be the measurements of the k -th random-effects variable. Let $N = \sum_{u=1}^m n_u$ be the number of observations of the response and of each explanatory variable.

The model is

$$r_{uj} = \sum_{k=1}^{k_\alpha} \alpha^{(k)} v^{(k)}_{uj} + \sum_{k=1}^{k_\beta} \beta^{(k)}_u w^{(k)}_{uj} + \gamma_u \epsilon_{uj}. \quad (1)$$

The $\alpha = (\alpha_{(1)}, \dots, \alpha_{(k_\alpha)})'$ are fixed-effect parameters. The $\beta_u = (\beta_{(1)u}, \dots, \beta_{(k_\beta)u})'$, the parameters of the random location effects, are i.i.d. multivariate random variables. We will suppose that the random-effects explanatory variables are contained in the space spanned by the fixed-effects explanatory variables, and take $E(\beta^{(k)}_u) = 0$. The ϵ_{uj} , error terms, are i.i.d. normal with $E(\epsilon_{uj}) = 0$. If the standard deviation (scale) of the ϵ_{uj} is constant, it will be denoted by $\sigma(\epsilon)$. We will allow for one form of fixed scale effects — the scale depends on the value of a categorical explanatory variable, z_{uj} , normally one of the fixed-effects variables $v^{(k)}_{uj}$. In such a case we denote the error scale by $\sigma_z(\epsilon_{uj})$. The γ_u , the random scale effects, are i.i.d. scalar random variables, and $E(\gamma_u^2) = 1$ because we have parameterized the variance of ϵ_{uj} . Finally, we suppose that the three collections of random variables — β_u , γ_u , and ϵ_{uj} — are mutually independent.

Throughout, we will use $\sigma^2(\delta)$ to denote the variance of the scalar random variable δ . We will take some liberties with subscripts. If we have a collection of variables δ_k all of which have the same variance, we will simply use $\sigma^2(\delta)$ instead of $\sigma^2(\delta_k)$; we have already begun this convention with $\sigma(\epsilon)$. If the variance of δ_k depends on a categorical variable, z_k , we write $\sigma_z(\delta_k)$. We have already begun this convention with $\sigma_z(\epsilon_{uj})$. A similar convention holds for the covariance matrix $\Sigma(\delta)$ of a multivariate random variable δ .

3 Unit Regressions: The Statistics and Their Properties

Most of our model building methods are based on *unit regressions*: r_{uj} is regressed on $w_{(1)uj}, \dots, w_{(k_\beta)uj}$ for each $u = 1$ to m . When there are fixed location or scale effects, we correct r_{uj} for these effects before carrying out the regressions.

3.1 Adjustment of r_{uj} for Fixed Location and Scale Effects

Suppose for the moment that $\alpha^{(k)}$ and $\sigma_z(\epsilon_{uj})$ are known. We will correct the response for the fixed location and scale effects using these known values and then carry out the unit regressions. Of course, in practice, we must correct with estimates, and in Section 3.8 we discuss estimation.

Let t_{uj} be r_{uj} adjusted for the fixed location effects:

$$t_{uj} = r_{uj} - \sum_{k=1}^{k_\alpha} \alpha^{(k)} v^{(k)}_{uj}. \quad (2)$$

Let y_{uj} be r_{uj} adjusted for both the fixed location and scale effects:

$$y_{uj} = \frac{t_{uj}}{\sigma_z(\epsilon_{uj})}. \quad (3)$$

Even when the error variance is constant, we adjust for the scale; as we shall see, in this constant case the adjustment serves the useful purpose of scaling the error term in a convenient way.

3.2 The Unit Regression Equation

Let

$$\begin{aligned} x^{(k)}_{uj} &= \frac{w^{(k)}_{uj}}{\sigma_z(\epsilon_{uj})} \\ \theta_u &= \frac{\beta_u}{\sigma_z(\epsilon_{uj})} \\ \zeta_{uj} &= \frac{\epsilon_{uj}}{\sigma_z(\epsilon_{uj})} \\ \tau_{uj} &= \gamma_u \zeta_{uj}. \end{aligned}$$

Then Equation 3 becomes

$$y_{uj} = \sum_{k=1}^{k_\beta} \beta^{(k)}_u x^{(k)}_{uj} + \tau_{uj}, \quad (4)$$

or

$$y_{uj} = \sum_{k=1}^{k_\beta} \theta^{(k)}_u w^{(k)}_{uj} + \tau_{uj}. \quad (5)$$

Equation 4 maintains the random-location parameters in the original units and rescales the $w^{(k)}_{uj}$, and Equation 5 maintains the $w^{(k)}_{uj}$ in the original units and rescales the β_u . In some applications the first is convenient, and in others the second is convenient. In the remainder of this paper we will use the scaling of Equation 4.

For fixed u , conditional on β_u and γ_u , Equation 4 for $j = 1$ to n_u is the unit regression equation for unit u . The unknown parameters are β_u and γ_u , and the errors ζ_{uj} are normal with mean 0 and variance 1. The unit regression equation in matrix notation is

$$y_u = X_u \beta_u + \gamma_u \zeta_u = X_u \beta_u + \tau_u,$$

where

$$\begin{aligned} y_u &= (y_{u1}, \dots, y_{un_u})' \\ \tau_u &= (\tau_{u1}, \dots, \tau_{un_u})' \\ \zeta_u &= (\zeta_{u1}, \dots, \zeta_{un_u})' \end{aligned}$$

and X_u is an $n_u \times k_\beta$ matrix whose (j, k) -th element is $x_{(k)uj}$. Let

$$P_u = X_u(X_u'X_u)^{-1}X_u' \quad (6)$$

be the projection matrix onto the space spanned by the columns of X_u , and let $\bar{P}_u = I_{n_u} - P_u$ where I_{n_u} is the $n_u \times n_u$ identity matrix. Let p_{uij} be the (i, j) -th element of P_u , and let \bar{p}_{uij} be the (i, j) -th element of \bar{P}_u .

3.3 Estimates of Random Location Parameters

The least squares estimate of β_u from the u -th unit regression is

$$\begin{aligned} \hat{\beta}_u &= (X_u'X_u)^{-1}X_u'y_u \\ &= \beta_u + \gamma_u(X_u'X_u)^{-1}X_u'\zeta_u \\ &= \beta_u + \gamma_u\xi_u \end{aligned} \quad (7)$$

where

$$\xi_u = (X_u'X_u)^{-1}X_u'\zeta_u \sim N(0, (X_u'X_u)^{-1}).$$

The distribution of $\hat{\beta}_u$ is an additive-multiplicative convolution of the distributions of β_u , γ_u , and ξ_u . The distribution depends on both β_u and γ_u , and it changes with X_u . Let $\Sigma(\beta)$ be the variance-covariance matrix of β_u , and let $\Sigma(\hat{\beta}_u)$ be the variance-covariance matrix of $\hat{\beta}_u$. Then

$$\Sigma(\hat{\beta}_u) = \Sigma(\beta) + (X_u'X_u)^{-1}. \quad (8)$$

3.4 Residuals

The residuals from the unit regression are

$$\begin{aligned} \hat{\tau}_u &= (\hat{\tau}_{u1}, \dots, \hat{\tau}_{un_u})' \\ &= y_u - X_u\hat{\beta}_u \\ &= \gamma_u\bar{P}_u\zeta_u \\ &= \gamma_u\hat{\zeta}_u. \end{aligned}$$

where

$$\hat{\zeta}_u = (\hat{\zeta}_{u1}, \dots, \hat{\zeta}_{un_u})' = \bar{P}_u\zeta_u \sim N(0, \bar{P}_u).$$

Thus

$$\sigma^2(\hat{\zeta}_{uj}) = \bar{p}_{ujj}. \quad (9)$$

3.5 Standardized Residuals

The *standardized residuals* are

$$\hat{\psi}_{uj} = \frac{\hat{\tau}_{uj}}{\sqrt{\bar{p}_{ujj}}} = \gamma_u \frac{\hat{\zeta}_{uj}}{\sqrt{\bar{p}_{ujj}}} \sim \gamma_u N(0, 1). \quad (10)$$

Let

$$\hat{\psi}_u = (\hat{\psi}_{u1}, \dots, \hat{\psi}_{un_u})'.$$

The distribution of $\hat{\psi}_{uj}$ is independent of β_u , depends on γ_u , and does not vary with n_u . The distribution is a multiplicative convolution of the γ_u and the standard normal. For example, suppose

$$\gamma_u^2 \sim IG(h, h-1) \quad (11)$$

where $h > 2$. $IG(h, \lambda)$ is an inverse gamma distribution with shape h and scale λ , which means that λ divided by the random variable is $G(h, 1)$; the mean is $\lambda(h-1)^{-1}$ and the variance is $\lambda^2(h-1)^{-2}(h-2)^{-1}$. Suppose furthermore in our example that $d = 2h$ is an integer (greater than 4). Then

$$\hat{\psi}_{uj} \sim T(d, 0, 1-2/d)$$

where $T(d, 0, \lambda^2)$ is a t distribution with d degrees of freedom, location 0, and scale λ .

The effect of $\sigma^2(\gamma^2) > 0$ is to make the tails of the $\hat{\psi}_{uj}$ heavier than those of a normal, at least as measured by the standard coefficient of kurtosis, which is zero for the normal. The variance of $\hat{\psi}_{uj}$ is

$$E(\hat{\psi}_{uj}^2) = 1,$$

the variance of γ_u^2 is

$$\sigma^2(\gamma^2) = E\gamma_u^4 - 1,$$

and

$$E(\hat{\psi}_{uj}^4) = 3(\sigma^2(\gamma^2) + 1)$$

because the fourth moment of a $N(0, 1)$ variable is 3. The coefficient of kurtosis of the $\hat{\psi}_{uj}$ is

$$\frac{E(\hat{\psi}_{uj}^4)}{E^2(\hat{\psi}_{uj}^2)} - 3 = 3\sigma^2(\gamma^2). \quad (12)$$

As $\sigma^2(\gamma^2)$ increases, the coefficient increases.

3.6 Residual Variance

Let s_u^2 be the residual variance, the residual sum of squares divided by the degrees of freedom:

$$s_u^2 = \frac{\hat{\tau}_u'\hat{\tau}_u}{n_u - k_\beta} = \gamma_u^2 \frac{\hat{\zeta}_u'\hat{\zeta}_u}{n_u - k_\beta}. \quad (13)$$

Then

$$s_u^2 \sim \gamma_u^2 MSQ(n_u - k_\beta) \quad (14)$$

where $MSQ(d)$ is a *mean-square distribution* with d degrees of freedom, the distribution of a chi-square random variable divided by its degrees of freedom. The distribution of s_u^2 is independent of β_u , depends on γ_u , and varies with n_u . For

example, if $\gamma_u^2 \sim IG(h, h-1)$, and $d = 2h$ is an integer greater than 4, then

$$s_u^2 \sim (1 - 2/d)F(n_u - k_\beta, d)$$

where $F(f_1, f_2)$ is an F -distribution with f_1 and f_2 degrees of freedom.

If $\sigma^2(\gamma^2) = 0$, then

$$\sigma^2(s_u^2) = \frac{2}{n_u - k_\beta}. \quad (15)$$

The effect of $\sigma^2(\gamma^2) > 0$ is to inflate $\sigma^2(s_u^2)$ because

$$\sigma^2(s_u^2) = \frac{2}{n_u - k_\beta} + \sigma^2(\gamma^2) \left(1 + \frac{2}{n_u - k_\beta}\right). \quad (16)$$

3.7 Studentized Residuals

The *studentized residuals* are

$$\hat{\phi}_u = (\hat{\phi}_{u1}, \dots, \hat{\phi}_{un_u})' = \frac{\hat{\psi}_u}{s_u},$$

and we have

$$\hat{\phi}_{uj} = \frac{\sqrt{n_u - k_\beta} \hat{\zeta}_{uj}}{\sqrt{\hat{p}_{ujj}} \sqrt{\hat{\zeta}_u' \hat{\zeta}_u}}. \quad (17)$$

A random variable on $[-1, 1]$ has a *double squared beta distribution*, $DSB(r, s)$, if its distribution is symmetric and its square is distributed $BETA(r, s)$. We can think of this distribution as follows. Start with a $BETA(r, s)$ variable on $[0, 1]$, and take the square root, which has a ‘‘squared beta’’ distribution because the square is a $BETA(r, s)$ (just like a ‘‘log normal’’ is a variable whose log is normal). Now symmetrize the squared beta by reflecting the density about zero; the result is a ‘‘double squared beta’’ (just like a ‘‘double exponential’’, which is a symmetrized exponential).

A simple derivation shows that

$$\hat{\phi}_{uj} \sim \sqrt{n_u - k_\beta} DSB(1/2, (n_u - k_\beta - 1)/2). \quad (18)$$

The distribution is independent of β_u and γ_u , but it changes with n_u .

3.8 Adjustment of r_{uj} for Fixed Effects Using Estimates of the Fixed- Effects Parameters

Our adjustment method is to follow Section 3.1; we estimate the fixed location-effects parameters, both location and scale, assume the estimates are the true parameters and use the adjustment equations in Section 3.1.

We estimate the $\alpha_{(k)}$ by $\hat{\alpha}_{(k)}$, the least squares estimates from regressing r_{uj} on the k_α variables $v_{(k)uj}$ for $k =$

1 to k_α . The least squares estimates are reasonable because, taking expectations across all random variables, $E(\hat{\alpha}_{(k)}) = \alpha_{(k)}$, although in the end we will have more efficient estimates when the whole model structure is taken into account; however, for model building, the least squares estimates are likely to be adequate. Before moving to scale estimation and adjustment, we correct for the fixed location effects assuming $\alpha_{(k)} = \hat{\alpha}_{(k)}$, forming t_{uj} .

Next we estimate the error scale. First, suppose it is constant, equal to $\sigma(\epsilon)$. We carry out the unit regressions without scale adjustment, forming $\hat{\beta}_u$, and then estimate $\sigma(\epsilon)$ by

$$\hat{\sigma}^2(\epsilon) = \frac{\sum_{u=1}^m \sum_{j=1}^{n_u} (t_{uj} - \sum_{k=1}^{k_\beta} \hat{\beta}_{(k)u} w_{(k)uj})^2}{N - mk_\beta}.$$

As stated in Section 3.1, we carry out scale adjustment even in this constant case because it provides a convenient scaling for the model building. Now suppose the scale is not constant, equal to $\sigma_z(\epsilon_{uj})$. Let $Q(z)$ be the set of indices (u, j) for which $z_{uj} = z$, and let n_z be the number of unique values of z_{uj} . From Equations 3 and 4 and Section 3.4 we have the following n_z equations:

$$E \sum_{Q(z)} \left(\frac{t_{uj} - \sum_{k=1}^{k_\beta} \hat{\beta}_{(k)u} w_{(k)uj}}{\sigma_z(\epsilon_{uj})} \right)^2 = \sum_{Q(z)} \bar{p}_{ujj}. \quad (19)$$

Note that each \bar{p}_{ujj} depends on the n_z values of $\sigma_z(\epsilon_{uj})$. We remove the expectations from Equation 19 and form the following n_z equations:

$$\sum_{Q(z)} \left(\frac{t_{uj} - \sum_{k=1}^{k_\beta} \hat{\beta}_{(k)u} w_{(k)uj}}{\sigma_z(\epsilon_{uj})} \right)^2 = \sum_{Q(z)} \bar{p}_{ujj}. \quad (20)$$

Repeated solution of these equations can be used to estimate the $\sigma_z(\epsilon_{uj})$. We begin with estimates of $\hat{\beta}_{(k)u}$ without scale adjustment, that is, assuming a constant error scale. Then we solve the n_z equations in Equation 20 getting estimates $\hat{\sigma}_z(\epsilon_{uj})$. Then we adjust for the error scale using these estimates, recompute the $\hat{\beta}_{(k)u}$, and then recompute $\hat{\sigma}_z(\epsilon_{uj})$. The procedure is iterated until it converges.

3.9 Estimating Variances

In the course of our model building, we will use estimates of the variances $\sigma^2(\gamma^2)$ and $\sigma^2(s_u^2)$, and estimates of the covariance matrices $\Sigma(\beta)$ and $\Sigma(\hat{\beta}_u)$. Equation 12 can be used to form an estimate of $\sigma^2(\gamma^2)$:

$$\hat{\sigma}^2(\gamma^2) = \frac{\sum_{u=1}^m \sum_{j=1}^{n_u} \hat{\psi}_{uj}^4}{3N} - 1. \quad (21)$$

Equations 16 and 21 can be used to form an estimate of $\sigma^2(s_u^2)$:

$$\hat{\sigma}^2(s_u^2) = \frac{2}{n_u - k_\beta} + \hat{\sigma}^2(\gamma^2) \left(1 + \frac{2}{n_u - k_\beta}\right). \quad (22)$$

Equation 8 can be used to form an estimate of $\Sigma(\beta)$:

$$\hat{\Sigma}(\beta) = m^{-1} \sum_{u=1}^m (\hat{\beta}_u \hat{\beta}_u' - (X_u' X_u)^{-1}). \quad (23)$$

Equations 8 and 23 can be used to form an estimate of $\Sigma(\hat{\beta}_u)$:

$$\hat{\Sigma}(\hat{\beta}_u) = \hat{\Sigma}(\beta) + (X_u' X_u)^{-1}. \quad (24)$$

4 A Sequence of Model Building Steps

The model building process consists of a sequence of steps. In each step, model building tools based on certain statistics are used to study the structure of the data and identify or check an aspect of the model; the results of most of the steps are used explicitly in subsequent steps.

4.1 Step 1: An Initial Specification of the Regression Components and Fixed Scale Effects

We need an initial partial model for the data from the model class in Equation 1 that specifies the structure of the fixed-effects regression component

$$\sum_{k=1}^{k_\alpha} \alpha^{(k)} v^{(k)}_{uj},$$

the random-effects regression component

$$\sum_{k=1}^{k_\beta} \beta^{(k)} u w^{(k)}_{uj},$$

and any fixed scale effect $\sigma_z(\epsilon_{uj})$. At this stage we leave open the form of the distributions of β_u and γ_u ; these specifications come in later steps. Also, specifications that already exist in the model class, such as the normality of the ϵ_{uj} and the independence of γ_u and β_u , will be checked in later steps.

Specification of the regression components and fixed scale effects will depend on our knowledge external to the data, but we can explore the data as well using visualization. It is not possible at this level of generality to lay out specific exploration tools. The nature of the explanatory variables determines what is likely to be useful. It is sometimes helpful to visualize the data for each unit separately, for example, if the explanatory variables take on the same values for each unit. It

is sometimes helpful to study the dependence on fixed-effects explanatory variables by visualizations that pool across units; in so doing the variability in the visual displays becomes the error variability together with the random-effects variability.

4.2 Step 2: Initial Adjustment of the Fixed Location and Scale Effects

As described in Section 3.8, we estimate the fixed location-effects parameters, $\alpha^{(k)}$, and the fixed scale-effects parameters $\sigma_z(\epsilon_{uj})$, and then adjust for the fixed effects. We consider the estimates to be the true values in the subsequent steps.

4.3 Step 3: Checking the Assumption of Normality of the Errors

The assumption of normality of the errors, ϵ_{uj} , or equivalently, the standardized errors, ζ_{uj} , can be checked using the studentized residuals, $\hat{\phi}_{uj}$. From Equation 18, under normality,

$$\hat{\phi}_{uj} \sim \sqrt{n_u - k_\beta} DSB(1/2, (n_u - k_\beta - 1)/2).$$

If n_u is constant, the $\hat{\phi}_{uj}$ are identically distributed (i.d.), and we check the assumption by an i.d. quantile plot (Cleveland 1993). If the n_u vary, then the $\hat{\phi}_{uj}$ are not i.d., and we use a mixture quantile plot (Cleveland 2000).

The normal distribution is symmetric, and the standardized residuals, $\hat{\psi}_{uj}$, can be used to check the symmetry of the ζ_{uj} , which provides a partial check of normality. From Equation 10,

$$\hat{\psi}_{uj} = \gamma_u \frac{\hat{\zeta}_{uj}}{\sqrt{p_{ujj}}}.$$

If ζ_{uj} is symmetric, $\hat{\psi}_{uj}$ is symmetric because $\hat{\zeta}_{uj}$ is symmetric and γ_u is a nonnegative random variable. A normal quantile plot of the $\hat{\psi}_{uj}$ is one way to check their symmetry.

Suppose the normal distribution does not appear to be a good approximation. Sometimes a transformation of the response can be the remedy. If not, we must find a non-normal approximation. Often, the diagnostic plots for normality help identify the new distribution. But the above distributional results for the statistics $\hat{\phi}_{uj}$ and $\hat{\psi}_{uj}$ do not hold in general under non-normality; we must find their distributions under other assumptions either by derivation or simulation. In some cases we might want to use other statistics in their place. The subsequent steps in our model building process need alteration in a similar way; the non-normal case proceeds using the same framework as the normal, but the details must change. Here, we will suppose that normality does provide a good approximation.

4.4 Step 4: Checking for the Presence of Random Scale Effects

We now have in place a specification of normality for the ζ_{uj} . Based on this specification, we explore the data using the standardized residuals, $\hat{\psi}_{uj}$, and the residual variances, s_u^2 , to determine if random scale effects appear to be present.

If random scale effects are not present, then, from Equations 10 and 14,

$$\hat{\psi}_{uj} \sim N(0, 1),$$

and

$$s_u^2 \sim MSQ(n_u - k_\beta).$$

We can search for evidence of random scale effects by studying the empirical distributions of $\hat{\psi}_{uj}$ and s_u^2 to see if they follow these no-scale-effect reference distributions. For $\hat{\psi}_{uj}$ we simply use an i.d. normal quantile plot. For s_u^2 , if n_u is constant, we use an i.d. MSQ quantile plot, and if n_u varies, we use a mixture MSQ quantile plot. In the latter case we need the variance of s_u^2 , which, from Equation 15, is

$$\sigma^2(s_u^2) = \frac{2}{n_u - k_\beta}.$$

Quantile plots are typically more effective if the displayed variables have distributions that are symmetric or nearly so. The empirical distribution of s_u^2 is typically strongly skewed toward large values, but the fourth roots are often closer to symmetric, so our practice is to plot fourth root s_u^2 against the fourth roots of the quantiles of the reference distribution.

4.5 Step 5: Identifying the Distribution of γ_u

We now have two specifications in place — normality for the ζ_{uj} and an assumption that random scale effects are present. Based on these specifications, we explore the data using $\hat{\psi}_{uj}$ and s_u^2 to specify the distribution of γ_u .

From Equations 10 and 14,

$$\hat{\psi}_{uj} \sim \gamma_u N(0, 1),$$

and

$$s_u^2 \sim \gamma_u^2 MSQ(n_u - k_\beta).$$

Graphical deconvolution procedures, one based on s_u^2 and another on $\hat{\psi}_{uj}$, are used to identify the unknown distribution of γ_u .

For $\hat{\psi}_{uj}$, we posit a distribution for γ_u^2 , generate (derive or simulate) quantiles of the reference distribution, $\gamma_u N(0, 1)$, and make an i.d. quantile plot of the $\hat{\psi}_{uj}$ using the generated quantiles.

For s_u^2 , we posit a distribution for γ_u^2 , generate quantiles of the reference distribution $\gamma_u^2 MSQ(n_u - k_\beta)$, and make an

i.d. quantile plot if the n_u are equal and a mixture quantile plot if not. To carry out the mixture quantile method with s_u^2 we use the estimate $\hat{\sigma}^2(s_u^2)$ of $\sigma^2(s_u^2)$ from Equation 22.

For the distribution of γ_u^2 we can consider standard families for positive random variables. Four are the following:

1. Gamma: $G(h, \lambda = h^{-1})$, where h is the shape parameter and λ the scale; the mean is λh and the variance is $\lambda^2 h$.
2. Weibull: $W(h, \lambda = \Gamma^{-1}(h^{-1} + 1))$, where h is the transformation parameter and λ the scale; the mean is $\lambda \Gamma(h^{-1} + 1)$ and the variance is $\lambda^2 (\Gamma(2h^{-1} + 1) - \Gamma^2(h^{-1} + 1))$.
3. Inverse Gamma: $IG(h, \lambda = h - 1)$, where h is the shape and λ the scale, and λ divided by the random variable is $G(h, 1)$; the mean is $\lambda(h - 1)^{-1}$ and the variance is $\lambda^2(h - 1)^{-2}(h - 2)^{-1}$.
4. Log Normal: $LN(h, \lambda^2 = -2h)$, where h and λ are the mean and scale (standard deviation) of the natural log of the random variable; the mean is $e^{h+\lambda^2/2}$ and the variance is $e^{\lambda^2} - 1$.

The scale parameter of each of these distributional families has been set to a value that makes $E(\gamma_u^2) = 1$. One parameter, h , remains to specify a member of the family. One approach is to use trial values and make a quantile plot for each trial value. A second is to choose an h so that the variance of the distribution is equal to the estimated variance $\hat{\sigma}^2(\gamma^2)$ in Equation 21. This leads to the following values for h :

1. Gamma: $h = \hat{\sigma}^{-2}(\gamma^2)$.
2. Weibull: h is the solution to $\hat{\sigma}^2(\gamma^2) = 2h\Gamma^{-2}(h^{-1})\Gamma(2h^{-1}) - 1$.
3. Inverse Gamma: $h = \hat{\sigma}^{-2}(\gamma^2) + 2$.
4. Log Normal: $h = -0.5 \log(1 + \hat{\sigma}^2(\gamma^2))$.

4.6 Step 6: Identifying the Distribution of β_u

We now have three specifications in place — normality for the ζ_{uj} , an assumption that random scale effects are present, and a specification for the distribution of γ_u^2 . Based on these specifications, we explore the data using the random location estimates $\hat{\beta}_u$ to specify the distribution of β_u .

From Equation 7,

$$\hat{\beta}_u = \beta_u + \gamma_u \xi_u.$$

The distribution of γ_u^2 has been specified, and $\xi_u \sim N(0, (X_u' X_u)^{-1})$. We posit a distribution for β_u , generate the distribution of $\beta_u + \gamma_u \xi_u$, and compare this reference distribution with that of the empirical distribution of $\hat{\beta}_u$. Visualization tools to carry out the comparison will depend on the posited distribution and the structure of the random-effects regression component. Thus, as in the specification for the γ_u^2 , we use a graphical deconvolution procedure.

To illustrate the considerations in carrying out the specification let us look at a simple special case. Suppose that the random-effects regression component is a single location parameter; that is, $k_\beta = 1$, $\beta_u = \beta_{(1)u}$, and $w_{(1)uj} = 1$. Suppose $n_u = n$ is constant. The simple model then is

$$r_{uj} = \beta_u + \gamma_u \epsilon_{uj} \quad (25)$$

for $u = 1$ to m and $j = 1$ to n . In this case

$$\xi_u = n^{-1} \sum_{j=1}^n \zeta_{uj} = \bar{\zeta}_u,$$

and $\sigma^2(\xi) = n^{-1}$. Because β_u is a scalar, we replace Σ by σ^2 in Equations 8 and 23. Equation 8 becomes

$$\sigma^2(\hat{\beta}_u) = \sigma^2(\beta) + n^{-1},$$

and Equation 23 becomes

$$\hat{\sigma}^2(\beta) = m^{-1} \left(\sum_{u=1}^m \hat{\beta}_u^2 - n^{-1} \right).$$

In positing a distribution for β_u we specify its variance to be $\hat{\sigma}^2(\beta)$. The resulting $\beta_u + \gamma_u \xi_u$ are identically distributed with a reference distribution whose quantiles we compute and compare with the quantiles of $\hat{\beta}_u$ by an i.d. quantile plot. Note that if n is large enough that n^{-1} is small compared with $\hat{\sigma}^2(\beta)$, then the empirical distribution of $\hat{\beta}_u$ provides a direct look at the distribution of β_u . It is tempting, of course, to begin by positing normality for β_u . If this fails and the quantile plot suggests that the distribution is symmetric, but has longer tails than the normal, then we can posit $T(d, 0, \lambda^2)$, a t -distribution with d degrees of freedom, location 0, and scale λ . The variance of this T is $\lambda^2 d / (d - 2)$, so if d is specified we have

$$\lambda^2 = \hat{\sigma}^2(\beta) \frac{d - 2}{d}.$$

We can try different values of d and make a quantile plot for each.

4.7 Step 7: Checking the Dependence of γ_u on β_u

One important assumption of the general model of Equation 1 is that γ_u and β_u are independent. There is always a danger that γ_u depends on β_u because variability frequently increases with an increasing mean level. One way to check for dependence is to plot fourth root s_u^2 against $\hat{\beta}_u$.

4.8 Step 8: Checking for Remaining Correlation

Let us return to the simple model of Equation 25. The correlation between r_{uj} and r_{uk} is $\sigma^2(\beta) / (\sigma^2(\beta) + \sigma^2(\epsilon))$. In other

words, the random location-effects induce within-unit correlation. The correlation structure is, of course, complex for the general model of Equation 1. We can study the standardized residuals and the studentized residuals to search for higher than expected correlation as a way of detecting poor specification of the random-effects regression component. The details of how to search will depend on the structure of the random-effects regression component and the nature of the explanatory variables.

References

- Brady, H. (1985). The Perils of Survey Research: Interpersonally Incomparable Responses. *Political Methodology* 11, 269–292.
- Chinchilli, V. M., J. D. Esinhart, and W. G. Miller (1995). Partial Likelihood Analysis of Within-Unit Variances in Repeated Measurement Experiments. *Biometrics* 51, 205–216.
- Clark, L., W. S. Cleveland, L. Denby, and C. Liu (1999). Modeling Customer Survey Data. In C. Gatsonis, R. Kass, A. Carriquiry, A. Gelman, I. Verdinelli, and M. West (Eds.), *Case Studies in Bayesian Statistics IV*, pp. 3–57. New York: Springer.
- Cleveland, W. S. (1993). *Visualizing Data*. Summit, New Jersey, U.S.A.: Hobart Press.
- Cleveland, W. S. (2000). Mixture Quantile Plots for Studying the Distributions of Non-Identically But Similarly Distributed Random Variables. Technical report, Statistics Research, Bell Labs, Murray Hill, NJ. <http://cm.bell-labs.com/doc/mixturequantile.ps>.
- Cleveland, W. S., L. Denby, and C. Liu (2000). Random Scale Effects. Technical report, Bell Labs. <http://cm.bell-labs.com/doc/randomscale.ps>.
- Cox, D. R. and P. J. Solomon (1986). Analysis of Variability of Large Numbers of Small Samples. *Biometrika* 73, 543–554.
- James, A. T., W. N. Venables, I. B. Dry, and J. T. Wiskich (1994). Random Effects and Variances as a Synthesis of Nonlinear Regression Analysis of Mitochondrial Electron Transport. *Biometrika* 81, 219–235.
- Johnson, V. E. (1997). An Alternative to Traditional GPA for Evaluating Student Performance (with discussion). *Statistical Science* 12, 251–278.
- Leonard, T. (1975). A Bayesian Approach to the Linear Model with Unequal Variances. *Technometrics* 17, 95–102.

Lin, X., J. Raz, and S. D. Harlow (1997). Linear Mixed Models with Heterogeneous Within-Cluster Variances. *Biometrics* 53, 910–923.

Lindley, D. V. (1971). The Estimation of Many Parameters. In V. P. Godambe and D. A. Sprott (Eds.), *Foundations of Statistical Inference*, pp. 435–455. Toronto: Holt, Rinehart, and Winston.

Nusser, S. M., A. L. Carriquiry, K. W. Dodd, and W. A. Fuller (1996). A Semiparametric Transform Approach to Estimating Usual Daily Intake Distributions. *Jour. of the Amer. Statistical Assoc.* 91, 1440–1449.