

Linear Calibration with Grouped Data and its Implementation in R

Brandon M. Greenwell* and Christine M. Schubert**

Department of Mathematics and Statistics, Air Force Institute of Technology,

Wright-Patterson AFB, OH 45433

**email*: brandon.greenwell@afit.edu

***email*: christine.schubertkabban@afit.edu

SUMMARY: We consider the following extension of the linear calibration problem. In the first stage of the calibration experiment, for each of m subjects, n_i ($i = 1, 2, \dots, m$) independent determinations of a response, \mathcal{Y} , are made giving $N = \sum_{i=1}^m n_i$ observations in all. These are often referred to as *repeated measures data*. In the second stage, a new observation, denoted \mathcal{Y}_0 , is observed for which the corresponding predictor value, denoted x_0 , is unknown. Observations belonging to the same subject cannot be considered as independent, hence, we must account for within-subject correlation when making inference. An efficient way for doing so is to use a mixed-effects model. In particular, we consider the case of calibration with the linear mixed-effects model (LMM). We propose a simple estimator for the unknown x_0 and argue its utility by showing that it coincides with the maximum likelihood (ML) estimator for a simple case. We also extend the usual $1 - \alpha$ Wald and inversion confidence intervals for the unknown x_0 to the LMM case. Finally, we propose a simple parametric bootstrap algorithm that can be used to either obtain confidence intervals for x_0 directly or to adjust the inversion interval by relaxing certain normality assumptions. We illustrate the effectiveness and simplicity of our approach with real assay data using the R software.

KEY WORDS: Calibration; Linear mixed-effects model; Prediction; Repeated measures.

1. Introduction

Calibration for LMMs has already been considered by Oman (1998). Oman proposed an approximate parametric bootstrap based on sampling from the asymptotic distribution of the estimated model parameters. This treats the new observation, denoted \mathcal{Y}_0 , as fixed in the bootstrap simulation. In a true calibration problem, however, \mathcal{Y}_0 is a random variable with variance σ_0^2 (say). Thus, Oman's approach underestimates the true variation which will likely lead to confidence intervals that are too short. The bootstrap algorithm we propose in this paper is not approximate, but rather, fully parametric. In fact, our algorithm can be seen as a parametric version of the bootstrap procedure described in Jones and Rocke (1999) extended to the case of the LMM. More importantly, this algorithm is easy to implement in standard statistical software.

We consider LMMs of the form

$$\mathcal{Y}_{ij} = \underbrace{\mu(x_{ij}; \boldsymbol{\beta})}_{\text{FIXED}} + \underbrace{R(x_{ij}; \mathcal{B}_i)}_{\text{RANDOM}} + \epsilon_{ij}, \quad (1)$$

$$i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n_i,$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, \mathcal{B}_i is a $q \times 1$ vector of random effects, and the functions μ and R are linear in $\boldsymbol{\beta}$ and \mathcal{B}_i , respectively. Furthermore, we assume that $\mathcal{B}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$ and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and that $\{\mathcal{B}_i\}$ and $\{\epsilon_{ij}\}$ are mutually independent. Since $R(x_{ij}; \mathcal{B}_i)$ is random, it follows that $\text{Var}[\mathcal{Y}_{ij}] = \text{Var}[R(x_{ij}; \mathcal{B}_i)] + \sigma_\epsilon^2$, which may or may not depend on x_{ij} . For instance, if $R(x_{ij}; \mathcal{B}) = \mathcal{B} + \epsilon$ (i.e., a random intercept model), then $\sigma_{ij}^2 = \sigma_{\mathcal{B}}^2 + \sigma_\epsilon^2$, whereas if $R(x_{ij}; \mathcal{B}) = \mathcal{B}x + \epsilon$ (i.e., a random slope model), then $\sigma_{ij}^2 = x_{ij}^2 \sigma_{\mathcal{B}}^2 + \sigma_\epsilon^2$ which depends on x_{ij} . In full matrix form, we can write the LMM as

$$\mathcal{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathcal{B} + \boldsymbol{\epsilon}, \quad (2)$$

where \mathbf{X} and \mathbf{Z} are known design matrices,

$\mathcal{B} = (\mathcal{B}'_1, \mathcal{B}'_2, \dots, \mathcal{B}'_m)'$, and

$$\begin{bmatrix} \mathcal{B} \\ \epsilon \end{bmatrix} \sim \mathcal{N} \left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_m \otimes \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma_\epsilon^2 \mathbf{I}_N \end{bmatrix} \right\}.$$

Thus, we have $\text{Var}[\mathcal{Y}] = \mathbf{I}_m \otimes (\mathbf{Z}\mathbf{G}\mathbf{Z}' + \sigma_\epsilon^2 \mathbf{I}_N) = \mathbf{V}$. For computational purposes, we often work with the scaled variance-covariance matrix, $\mathbf{G}^\dagger = \mathbf{G}/\sigma_\epsilon^2$, so that we may write $\mathbf{V} = \sigma_\epsilon^2 [\mathbf{I}_m \otimes (\mathbf{Z}\mathbf{G}^\dagger \mathbf{Z}' + \mathbf{I}_N)]$.

In LMMs, there are two standard predictions we can make regarding a future observation: (1) predicting a *new* observation *within* an existing group, and (2) predicting a *new* observation in a *new* group. Since longitudinal studies often aim to make inference for the whole population under study (Jiang and Zhang, 2002), we will restrict our attention to case (2). In particular, for calibration, we will assume that a (single) new observation, denoted \mathcal{Y}_0 , is independent of the current observations and does not belong to an existing group. We wish to make inference regarding the unknown predictor value, x_0 .

2. A Simple Estimator for the Unknown x_0

By assumption, a new observation, \mathcal{Y}_0 (independent of current observations), is distributed as a $\mathcal{N}\{\mu(x_0; \boldsymbol{\beta}), \sigma_0^2\}$ random variable where σ_0^2 may or may not depend on the unknown x_0 . A natural estimator for x_0 is then

$$\hat{x}_0 = \mu^{-1}(\mathcal{Y}_0; \hat{\boldsymbol{\beta}}), \quad (3)$$

where $\hat{\boldsymbol{\beta}}$ is the estimated best linear unbiased estimator (EBLUE) of $\boldsymbol{\beta}$. We shall refer to Equation (3) as the inverse estimator. As it turns out, for the random intercept model

$$\mathcal{Y}_{ij} = \mathcal{B}_i + \beta_0 + \beta_1 x_{ij} + \epsilon_{ij},$$

the inverse estimator $\hat{x}_0 = (\mathcal{Y}_0 - \hat{\beta}_0)/\hat{\beta}_1$ is the ML estimator of x_0 . Furthermore, if the data are balanced (i.e., $n_i = n$, $i = 1, 2, \dots, m$), the EBLUE, $\hat{\boldsymbol{\beta}}$, coincides with the ordinary least

squares estimator of β and \hat{x}_0 is the same as the ML estimator of x_0 (Graybill, 1976) for the ordinary linear calibration problem (see Web Appendix A for proof).

In general, it is difficult to find the ML estimator of x_0 . However, the inverse estimator just discussed is certainly reasonable. Next, we discuss various ways in which to form a $1 - \alpha$ confidence interval for x_0 , called a calibration interval.

For an example of calibration in nonlinear mixed-effects models, see Concordet and Nunez (2000) and the references therein.

3. Extending the Classic Calibration Intervals

In the calibration problem with independent observations, there are two well-known confidence intervals for the unknown x_0 : the *Wald interval* and the *inversion interval*. It is not too difficult to extend these methods to more complex models (see, for example, Davidian and Giltinan (1995, chap. 10)). In this section, we extend the Wald and inversion intervals to linear models with random coefficients (1). Then, we propose a parametric bootstrap algorithm that presumably is more accurate as it takes into account the variability of the estimated variance components. It also relies on fewer assumptions, making the parametric bootstrap slightly more robust. We also discuss how this bootstrap algorithm can be modified in order to compute an *adjusted* inversion interval that does not rely on asymptotic normality of the predictive pivot.

3.1 The Wald Interval

The Wald interval for x_0 , while not very accurate, is simple to compute. It has the form: $\hat{x}_0 \pm \gamma_{1-\alpha/2} \widehat{\text{se}}[\hat{x}_0]$, where $\gamma_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile from an appropriate distribution (e.g., a students- t or standard normal distribution) and $\widehat{\text{se}}[\hat{x}_0]$ is the estimated standard error of \hat{x}_0 . As with ordinary linear calibration, there is no textbook formula for the variance of \hat{x}_0 . Instead, we can approximate it using a first-order Taylor series expansion.

The variance-covariance matrix of $(\mathcal{Y}_0, \hat{\beta})$ is

$$\Sigma = \begin{bmatrix} \text{Var}[\mathcal{Y}_0] & \mathbf{0} \\ \mathbf{0} & \text{Var}[\hat{\beta}] \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \end{bmatrix}.$$

Since \mathcal{Y}_0 is independent of \mathcal{Y} , it is also independent of $\hat{\beta}$, hence the diagonal structure of Σ . Recall that our point estimate has the form $x = \mu^{-1}(y; \beta)$. Let $\mu_1^{-1}(y; \beta)$ and $\mu_2^{-1}(y; \beta)$ denote the partial derivatives of μ^{-1} with respect to the parameters y and β , respectively. Our point estimator is given by $\mu^{-1}(\mathcal{Y}_0; \hat{\beta})$, where \mathcal{Y}_0 is a new observation and $\hat{\beta}$ is the EBLUE of β . Using a first-order Taylor-series expansion, an approximate variance for \hat{x}_0 is given by

$$\begin{aligned} \text{Var}[\hat{x}_0] &= \left[\mu_1^{-1}(\mathcal{Y}_0; \hat{\beta}) \right]^2 \sigma_0^2 \\ &\quad + \left[\mu_2^{-1}(\mathcal{Y}_0; \hat{\beta}) \right]' (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \left[\mu_2^{-1}(\mathcal{Y}_0; \hat{\beta}) \right]. \end{aligned} \quad (4)$$

To obtain $\widehat{\text{Var}}[\hat{x}_0]$, we need to replace σ_0^2 and \mathbf{V} in Equation (4) with their corresponding estimates $\hat{\sigma}_0^2$ and $\hat{\mathbf{V}}$, respectively. Once we have $\widehat{\text{Var}}[\hat{x}_0]$, an approximate $1 - \alpha$ confidence interval for x_0 is given by

$$\mathcal{C}_W = \hat{x}_0 \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}[\hat{x}_0]}, \quad (5)$$

where $z_{1-\alpha/2} = \Phi(1 - \alpha/2)$ is the $1 - \alpha/2$ quantile of a standard normal distribution.

As pointed out by Davidian and Giltinan (1995, p. 283), the first term in Equation (4) usually dominates the second term, leaving the cruder approximation $\text{Var}_{\text{crude}}[\hat{x}_0] = [\mu_1^{-1}(\mathcal{Y}_0; \hat{\beta})]^2 \sigma_0^2$. For example, for the random intercept model we have

$$\begin{aligned} \mu^{-1}(y; \beta) &= (y - \beta_0) / \beta_1, \\ \frac{\partial}{\partial y} \mu^{-1}(y; \beta) &= 1 / \beta_1, \end{aligned}$$

which leads to the crude approximation $\text{Var}_{\text{crude}}[\hat{x}_0] = (\sigma_\epsilon^2 + \sigma_{\mathcal{B}}^2) / \hat{\beta}_1^2$.

In general, the approximate variance (4) can be difficult to compute by hand leading to widespread use of the cruder approximation. However, given the ease with which complex

computations can be carried out numerically, it is good practice to calculate and use both terms. In Web Appendix B, we show how easily the Wald interval (5) can be obtained in R using the `deltaMethod` function from the well-known package `car` (Fox and Weisberg, 2011). And although the Wald interval is the least accurate, it can still be used as a basis for which to compare other methods.

3.2 The Inversion Interval

Similar to how we obtained a point estimator of x_0 by inverting the mean response, we can obtain a confidence interval for the unknown x_0 by inverting an asymptotic prediction interval for \mathcal{Y}_0 . Let \mathbf{X}_0 have the same form as the i -th row of \mathbf{X} , but with x_{ij} replaced with the unknown x_0 . For example, if $\mu(x_{ij}; \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2$, then $\mathbf{X}_0 = (1, x_0, x_0^2)'$.

For brevity, let $\mu_0 = \mu(x_0; \boldsymbol{\beta})$, $\tilde{\mu}_0 = \mu(x_0; \tilde{\boldsymbol{\beta}})$, and $\hat{\mu}_0 = \mu(x_0; \hat{\boldsymbol{\beta}})$ where $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ denote the BLUE and EBLUE of $\boldsymbol{\beta}$, respectively. A new observation, \mathcal{Y}_0 , with unknown x_0 , is distributed as $\mathcal{N}(\mu_0, \sigma_0^2)$. Clearly, $\mathcal{Y}_0 - \tilde{\mu}_0$ is a normally distributed random variable with expectation zero (since $E[\mathcal{Y}_0] = E[\tilde{\mu}_0] = \mu_0$). Also, note that \mathcal{Y}_0 and $\tilde{\mu}_0$ are independent, hence, $\text{Cov}[\mathcal{Y}_0, \tilde{\mu}_0] = 0$. It therefore follows that the statistic

$$\mathcal{Z} = \frac{\mathcal{Y}_0 - \tilde{\mu}_0}{\sqrt{\text{Var}[\mathcal{Y}_0] + \text{Var}[\tilde{\mu}_0]}} \quad (6)$$

is a *pivotal quantity* that has a standard normal distribution or, equivalently, $\mathcal{Z}^2 \sim \chi_1^2$ (a Chi-squared distribution with a single degree of freedom). The key here is that $\tilde{\mu}_0$ is a linear function of \mathcal{Y} . For instance, in a balanced random intercept model, Equation (6) reduces to

$$\mathcal{Z} = \frac{\mathcal{Y}_0 - \tilde{\beta}_0 - \tilde{\beta}_1 x_0}{\sqrt{\sigma_\epsilon^2 + \sigma_{\mathcal{B}}^2 + \mathbf{X}_0' \left[\mathbf{X}' (\sigma_{\mathcal{B}}^2 \mathbf{I}_m \otimes \mathbf{J}_n + \sigma_\epsilon^2 \mathbf{I}_N)^{-1} \mathbf{X} \right]^{-1} \mathbf{X}_0}},$$

where $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n'$ is an $n \times n$ matrix of all ones.

In practice, σ_0^2 , \mathbf{V} , and \mathbf{X}_0 are usually unknown and need to be estimated from the data. In such cases, an approximate predictive pivot, denoted \mathcal{Q} , can be obtained by replacing σ_0^2 , \mathbf{V} , and \mathbf{X}_0 in the above equation with their respective estimates $\hat{\sigma}_0^2$, $\hat{\mathbf{V}}$ and $\hat{\mathbf{X}}_0$. This

suggests an approximate $1 - \alpha$ confidence interval for x_0 of

$$\mathcal{C}_I = \{x : z_{\alpha/2} < \mathcal{Q} \leq z_{1-\alpha/2}\}, \quad (7)$$

It is unlikely that Equation (7) will yield closed-form solutions, thus, the set must be obtained numerically. In Web Appendix B, we show how this interval can be obtained using the R software (R Core Team, 2013).

4. Using the Parametric Bootstrap (PB)

Jones and Rocke (1999) discuss how to calculate calibration intervals for the linear and nonlinear regression models based on the nonparametric bootstrap. A crucial assumption for the ordinary nonparametric bootstrap, however, is that the data are independent, but for reasons discussed earlier, this assumption is typically not valid for grouped data. Nonetheless, a different kind of bootstrap, called the *parametric bootstrap*, has shown promise as a serious inferential tool when analyzing grouped data. For examples, see McCulloch et al. (2008, pg.342) and Efron (2011). When applicable, the parametric bootstrap typically gives answers similar to a Bayesian analysis with uninformative priors, however, it is much faster than Markov Chain Monte Carlo simulations. The parametric bootstrap essentially entails sampling from the fitted model itself, rather than sampling with replacement from the data.

For linear calibration with grouped data modeled via an LMM, we propose the algorithm seen in Figure 1.

[Figure 1 about here.]

We will take the sample $\alpha/2$ and $1 - \alpha/2$ quantiles of the bootstrap estimates $\hat{x}_{01}, \hat{x}_{02}, \dots, \hat{x}_{0R}$ as a confidence interval for x_0 . This is known as the percentile bootstrap interval and we will denote it by \mathcal{C}^* . Note that only steps (5)-(6) of the algorithm are specific to calibration. Other parametric bootstrap schemes have been proposed for LMMs. For example, we can condition on the current values of the random effects by ignoring step (1) of the algorithm

and use the current EBLUP, $\widehat{\mathcal{B}}$, in place of \mathcal{B}^* . Some semiparametric variants of steps (1)-(4) of the algorithm in Figure 1 involve sampling with replacement directly from the EBLUPs and residuals, but Morris (2002) considers this to be bad practice because it consistently underestimates the true variation in the data.

4.1 A Bootstrap Adjusted Inversion Interval

Although we favor the parametric bootstrap confidence intervals obtained directly from the R bootstrap replicates of \widehat{x}_0 , researchers are likely more familiar with the inversion and Wald intervals discussed in Section 3. These intervals, however, use the quantiles from a standard normal distribution and therefore rely on normal approximations. The parametric bootstrap can be used to improve upon these intervals by replacing the standard normal quantiles with more accurate ones. For instance, for the inversion interval, at each run of the algorithm in Figure 1, we compute

$$\mathcal{Q}^* = \frac{y_0^* - \mu(\widehat{x}_0; \widehat{\beta}^*)}{\sqrt{\widehat{\sigma}_0^{2*} + \widehat{\mathbf{X}}_0' (\mathbf{X}' \widehat{\mathbf{V}}^{*-1} \mathbf{X}')^{-1} \widehat{\mathbf{X}}_0}},$$

where $\widehat{\sigma}_0^{2*}$ is calculated at \widehat{x}_0 and not \widehat{x}_0^* . As a result, we obtain R bootstrap values \mathcal{Q}_r^* . Let $q_{\alpha/2}^*$ and $q_{1-\alpha/2}^*$ denote the sample $\alpha/2$ and $1-\alpha/2$ quantiles of \mathcal{Q}_r^* , respectively. A bootstrap adjusted inversion interval is then given by

$$\mathcal{C}_I^* = \{x : q_{\alpha/2}^* \leq \mathcal{Q} \leq q_{1-\alpha/2}^*\}. \quad (8)$$

We illustrate this on the bladder volume example in Section 6. A similar adjustment can also be made for the Wald interval. For example, we could use the parametric bootstrap to estimate $\text{Var}[\widehat{x}_0]$, instead of relying on a linear approximation. A bootstrap adjusted Wald interval is given by

$$\mathcal{C}_W^* = \widehat{x}_0 \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}^*[\widehat{x}_0]}. \quad (9)$$

Although this interval still relies on a normal approximation, it will be more accurate since the variability of the estimated variance components is accounted for in the bootstrap resampling. As a result, \mathcal{C}_W^* will usually be wider than \mathcal{C}_W , but this extra wideness is a good thing.

5. Simulation Study

To assess the empirical performance of these confidence intervals, we carried out a small Monte Carlo study. The simulations described below were conducted in R using packages **plyr** (Wickham, 2011), **nlme** (Pinheiro et al., 2013), **lme4** (Bates et al., 2014), and **boot** (Canty and Ripley, 2013).

We generated $N = 1,000$ data sets with $m = 15$ subjects and $n = 10$ observations per subject (i.e., balanced designs) from the populations defined by

$$\text{Model 1: } \mathcal{Y} = \mathcal{B}_0 + (1 + \mathcal{B}_1)x + \epsilon,$$

$$\text{Model 2: } \mathcal{Y} = \mathcal{B}_0 + (0.1 + \mathcal{B}_1)x + 0.9x^{1/3} + \epsilon,$$

where $(\mathcal{B}_0, \mathcal{B}_1)' \sim \mathcal{N}(\mathbf{0}, \text{diag}\{0.001, 0.025\})$ are random effects, and $\epsilon \sim \mathcal{N}(0, 0.001)$ are random errors independent of \mathcal{B}_0 and \mathcal{B}_1 . The design points were uniformly spaced on the interval $[0, 1]$ and the true unknowns, x_0 , were chosen to be 0, 0.25, 0.5, 0.75, and 1. A sample dataset from each model is plotted in Figure 2. We expected problems at larger values of x_0 for the data generated from Model 2 because the sampling distribution of \hat{x}_0 will be positively skewed and $\text{se}[\hat{x}_0]$ will likely be too large to produce a useful confidence interval for x_0 ; nonetheless, this extreme case illustrates an important advantage of the bootstrap and inversion intervals over the Wald intervals, in that they are not symmetric.

[Figure 2 about here.]

For each case, we computed \mathcal{C}_W , \mathcal{C}_I , \mathcal{C}_W^* , \mathcal{C}_I^* , and \mathcal{C}^* at the 0.95 level of confidence. The estimated coverage probabilities and lengths are reported in Tables 1 and 2. The standard deviation of the coverage estimates is approximately $\sqrt{0.95(1 - 0.95)/1000} = 0.007$.

We see from the Table 1 that the coverage is fairly close to the nominal level of 0.95 for all five methods, though there is a slight problem with under-coverage at larger values x_0 where the variation in the response is highest. The average lengths of the intervals is also comparable across all five methods. From Table 2, we see that the Wald intervals \mathcal{C}_W and \mathcal{C}_W^* performed rather poorly with respect to coverage when compared to the inversion and percentile bootstrap intervals as x_0 is increased, though the coverage for all five intervals decreased for large values of x_0 (as expected). These results indicate to us that it is safest to use the inversion and percentile bootstrap intervals when there is curvature in the population mean response. Even when all the intervals perform similarly — as in Table 1 — the parametric bootstrap procedure provides an estimate of the entire sampling distribution of \hat{x}_0 , rather than just a standard error and confidence interval for x_0 . And while the inversion and bootstrap adjusted inversion intervals seem to perform well, they may not always produce a finite interval, such is the case for \mathcal{C}_I and \mathcal{C}_I^* in the last four columns of Table 2. The percentile interval, on the other hand, always exists as long as the point estimate exists (i.e., when the population mean response is monotonic in the region of interest).

Although we favor the percentile interval (\mathcal{C}^*), there are many other types of bootstrap confidence intervals that may be worth considering; see for example, Davison and Hinkley (1997, chap. 5). Bootstrap confidence intervals are typically calculated using as large a bootstrap sample size as possible. However, due to limited time and computational power, the number of bootstrap samples used in this simulation was $R = 999$. While this is a modest number to use, we suspect that the coverage of the parametric bootstrap intervals may improve, albeit only slightly, if R was on the order of 9,999 or so. Secondly, a number of other important factors were held fixed in this study (e.g., the sample sizes and variance components). To investigate how these parameters affect the length and coverage estimates of the intervals under question, a much larger Monte Carlo study is required.

[Table 1 about here.]

[Table 2 about here.]

6. Bladder Volume Data

In this section, we illustrate the techniques described in the previous section on a real dataset taken from Brown (1993). In Brown's words

"A series of 23 women patients attending a urodynamic clinic were recruited for the study. After successful voiding of the bladder, sterile water was introduced in additions of 10, 15, and then 25 ml increments up to a final cumulative total of 175 ml. At each volume a measure of height (H) in mm and depth (D) in mm of largest ultrasound bladder images were taken. The product $H \times D$ was taken as a measure of liquid volume..."

The product of H and D is plotted against true volume in Figure 3 for each of the 23 subjects. As can be seen from this plot, each subject has a slightly nonlinear trend and there are some missing observations (i.e., the data are unbalanced). Both Brown (1993) and Oman (1998) suggested transforming the data first. For instance, a simple *log-log* transformation produces a relatively linear trend for each subject. However, not all data can be transformed as successfully, thus, we will leave the data on the original scale and simply model the nonlinearity with a quadratic term. Let HD_{ij} be the j -th measurement for the i -th subject. The model we use is a second-degree polynomial with uncorrelated random effects for the intercept and linear terms:

$$HD_{ij} = \beta_0 + \mathcal{B}_{0i} + (\beta_1 + \mathcal{B}_{1i}) \text{volume}_{ij} + \beta_3 \text{volume}_{ij}^2 + \epsilon_{ij},$$

where $\mathcal{B}_{0i} \sim \mathcal{N}(0, \tau_0^2)$ independent of $\mathcal{B}_{1i} \sim \mathcal{N}(0, \tau_1^2)$ and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

[Figure 3 about here.]

Suppose we obtain a new observation, $HD_0 = 85 \text{ mm}^2$, for which the true volume is unknown. To estimate the unknown volume, denoted v_0 , we solve the equation

$$\hat{\mu}(v_0) = \hat{\beta}_0 + \hat{\beta}_1 v_0 + \hat{\beta}_2 v_0^2 = 85 \text{ mm}^2$$

for v_0 using the quadratic formula. The point estimate obtained is $\hat{v}_0 = 120.5289$ ml (see Figure 4). The corresponding R code is given in R Example 2 in Web Appendix B.

R Example 3 from Web Appendix B shows how to use the R package **car** to compute the approximate standard error based on the first-order Taylor series estimate given in Equation (4). Note that our model has $\text{Var}[\mathcal{Y}_0] = \tau_0^2 + x_0^2\tau_1^2 + \sigma_\epsilon^2$ where τ_0^2 and τ_1^2 are the variances of the random intercept and linear term, respectively. The estimated standard error of \hat{v}_0 is 26.1981 which yields an approximate 95% Wald-based confidence interval for v_0 of (69.1815, 171.8763).

Obtaining the inversion interval (7) is less straightforward; in R, we have to write our own prediction function that will return the standard errors of the fitted values. R Example 4 from Web Appendix B shows the minimal R code necessary to obtain \mathcal{C}_I for the bladder volume example. The resulting interval, (75.3908, 184.7025), is only slightly larger than the Wald interval. Notice the inversion interval is not symmetric about the inverse estimate $\hat{v}_0 = 120.5289$ ml which, given the nonlinear trend and increasing variation in the data, is more realistic. In addition, we may also obtain the bootstrap adjusted inversion interval (8) described in Section 4.1 — see R example 5 in Web Appendix B which uses the new **bootMer** function in the R package **lme4** to perform the parametric bootstraps outlined in Section 4.. A histogram of the 9,999 bootstrap replicates of the predictive pivot \mathcal{Q} is also given in the appendix. The estimated sampling distribution of \mathcal{Q} seems reasonably normal. The necessary quantiles to compute \mathcal{C}_I^* are $q_{0.025}^* = -2.0031$ and $q_{0.975}^* = 1.9804$ (comparable to the corresponding quantiles from a standard normal distribution; i.e., ± 1.96). Using these quantiles, a 95% bootstrap adjusted inversion interval for x_0 is $\mathcal{C}_I^* = (57.94, 138.16)$, which is only slightly wider than the unadjusted inversion interval based on the normal approximation. In short, the normal approximation seems to work quite well for these data.

Finally, we compute the bootstrap distribution of \hat{v}_0 directly using the algorithm in Figure 1 — see R Example 5 in Web Appendix B. Our results are based on a bootstrap simulation of size $R = 9,999$. Since we do not expect the sampling distribution of \hat{v}_0 to be symmetric, we provide only a 95% percentile bootstrap interval; that is, the interval obtained by taking the lower 2.5 and upper 97.5 percentiles of the resulting bootstrap distribution. The computed interval, (58.37, 136.02), is quite close to the inversion interval previously obtained. A histogram of 9,999 bootstrap replicates of \hat{v}_0 is given in Figure 4. The histogram estimate suggests that the sampling distribution of \hat{v}_0 is positively skewed (the estimated skewness is 0.3011), providing further support for the inversion and percentile intervals over the symmetric Wald interval. The results for this example are summarized in Table 3.

[Figure 4 about here.]

[Table 3 about here.]

7. Discussion

We have described a number of techniques for controlled calibration in a linear mixed model setting. The Wald interval is the simplest, but relies on the asymptotic normality of \hat{x}_0 along with a Taylor-series approximation of its variance. Perhaps the biggest drawback to using a Wald calibration interval is that it is always symmetric about \hat{x}_0 . While this is appealing to many researchers, it is not very realistic in standard situations where, say, the data exhibit nonlinear behavior (e.g., horizontal asymptotes) and nonconstant variance. The asymptotic normality of \hat{x}_0 may also be questioned when it is not the ML estimate, as in the bladder volume example. This is akin to using the Wald method in nonlinear calibration problems (the software JMP does this). The Wald approach may, however, produce reliable inference when the sampling distribution of \hat{x}_0 is reasonably symmetric. Although more difficult to obtain, the inversion interval and its bootstrap adjustment appear more accurate

than the Wald intervals, especially when the mean response curve is not linear. However, the inversion interval may not exist for a given observation y_0 at the specified confidence level $1 - \alpha$. The percentile interval based on the parametric bootstrap algorithm given in this paper is especially promising in that it has comparable coverage probability and length to the inversion intervals; however, unlike the inversion intervals, the percentile parametric bootstrap method will always provides finite confidence limits as long as the fitted population mean response is monotonic in the region of interest. Finally, although we give snippets of R code for implementing the methods described in this paper for the bladder volume data in Web Appendix B, it is likely that a future release of the R package **investr** (Greenwell, 2013) will contain this functionality.

ACKNOWLEDGEMENTS

This work was supported by the AFOSR/RTA [grant F1ATA03039J001]

SUPPLEMENTARY MATERIALS

Web Appendices A and B referenced in Sections 2, 3, 4, and 6 are available with this paper at the Biometrics website on Wiley Online Library.

REFERENCES

- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-5.
- Brown, P. (1993). *Measurement, Regression, and Calibration*. Clarendon Press Oxford University Press.
- Canty, A. and Ripley, B. (2013). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-9.
- Concordet, D. and Nunez, O. G. (2000). Calibration for nonlinear mixed effects models: An application to the withdrawal time prediction. *Biometrics* **56**, 1040–1046.

- Davidian, M. and Giltinan, D. (1995). *Nonlinear Models for Repeated Measurement Data*. Taylor & Francis.
- Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press.
- Efron, B. (2011). The bootstrap and markov-chain monte carlo. *Journal of Biopharmaceutical Statistics* **21**, 1052–1062.
- Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. Sage, second edition.
- Graybill, F. A. (1976). *Theory and Application of the Linear Model*. Duxbury.
- Greenwell, B. M. (2013). *investr: Inverse Estimation with Linear and Nonlinear Regression Models in R*. R package version 1.0.
- Jiang, J. and Zhang, W. (2002). Distribution-free prediction intervals in mixed linear models. *Statistica Sinica* **12**, 537–553.
- Jones, G. and Rocke, D. M. (1999). Bootstrapping in controlled calibration experiments. *Technometrics* **41**, 224–233.
- McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. Wiley.
- Morris, J. S. (2002). The blups are not "best" when it comes to bootstrapping. *Statistics & Probability Letters* **56**, 425–430.
- Oman, S. D. (1998). Calibration with random slopes. *Biometrika* **85**, 439–449.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2013). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-110.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Version 3.0.1.
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of*

Statistical Software **40**, 1–29. R package version 1.8.1.

Received February 2014. Revised June 2014. Accepted August 2014.

For $r = 1, 2, \dots, R$ do:

- (1) generate q new values of the random effects, denoted \mathcal{B}_r^* , from a $\mathcal{N}(\mathbf{0}, \widehat{\mathbf{G}})$ distribution;
- (2) generate N new errors, denoted $\boldsymbol{\epsilon}_r^*$, from a $\mathcal{N}(\mathbf{0}, \widehat{\sigma}_\epsilon^2 \mathbf{I})$ distribution;
- (3) set $\mathbf{y}_r^* = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\mathcal{B}_r^* + \boldsymbol{\epsilon}_r^*$;
- (4) update the original model using \mathbf{y}_r^* as the response vector to obtain $\widehat{\boldsymbol{\beta}}_r^*$;
- (5) generate y_{0r}^* from a $\mathcal{N}(y_0, \widehat{\sigma}_0^2)$ distribution;
- (6) set $\widehat{x}_{0r}^* = \mu^{-1}(y_{0r}^*; \widehat{\boldsymbol{\beta}}_r^*)$.

Figure 1. Parametric bootstrap algorithm for controlled calibration in LMMs.

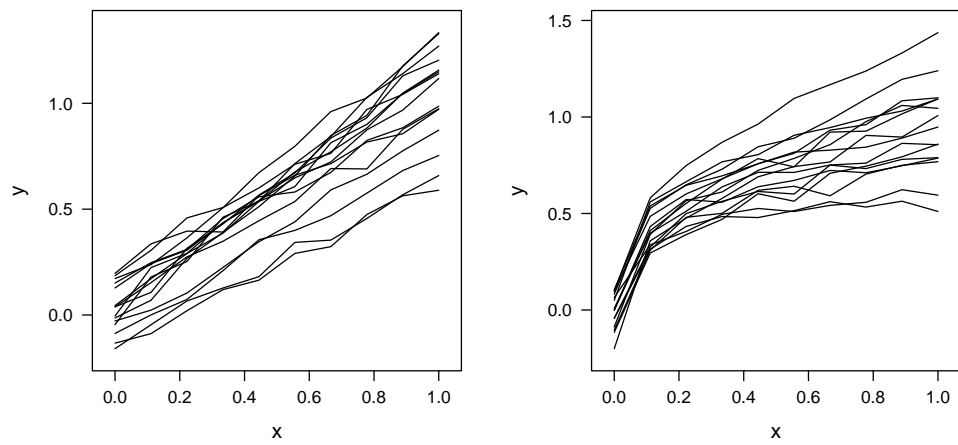


Figure 2. Spaghetti plots of simulated data. *Left:* Data simulated from Model (1). *Right:* Data simulated from Model (2).

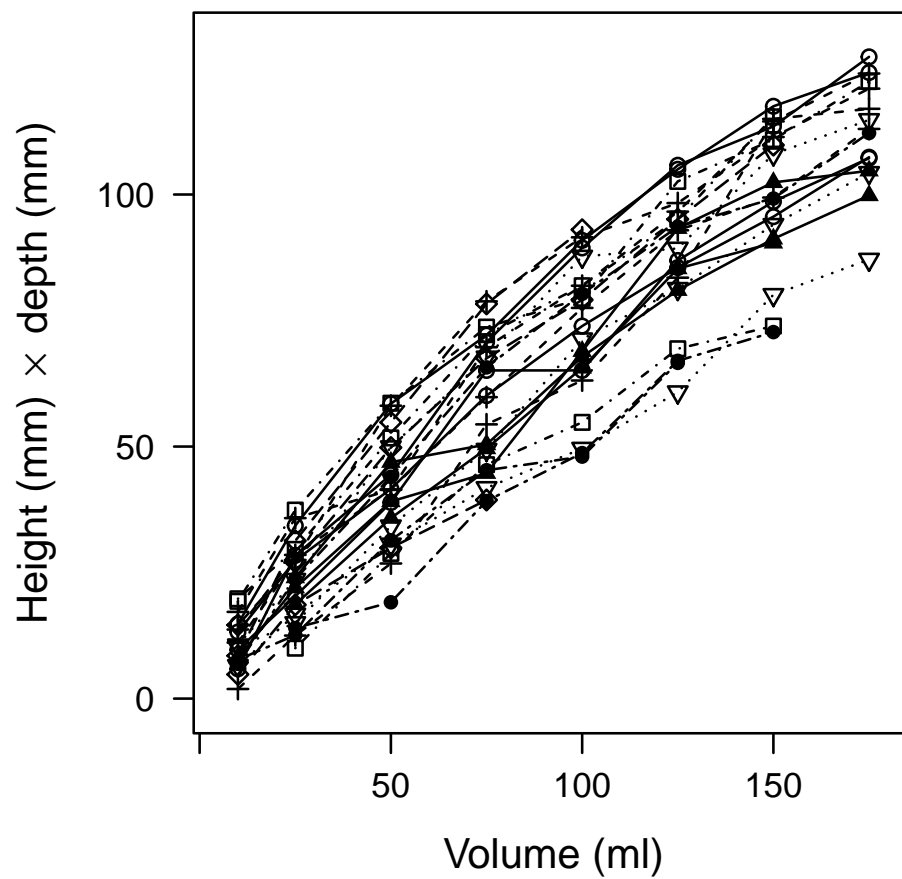


Figure 3. Scatterplot of the bladder volume data. Lines connect measurements belonging to the same person.

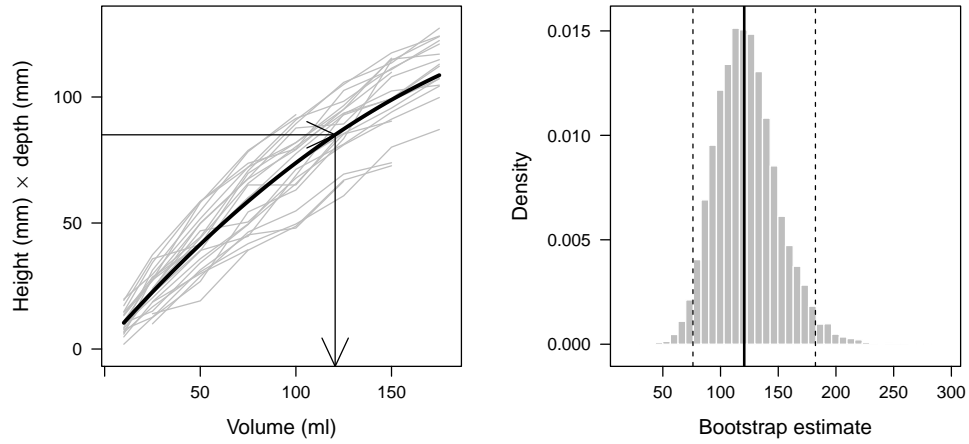


Figure 4. Bladder data results. *Left:* Spaghetti plot of the bladder data with fitted mean response. The horizontal arrow indicates the position of $\text{HD}_0 = 85 \text{ mm}^2$ and the vertical arrow indicates the position of the inverse estimate, $\hat{v}_0 = 120.53 \text{ ml}$. *Right:* Histogram of the 9,999 bootstrap replicates of \hat{v}_0 . The solid line indicates the position of the original estimate and the dotted lines represent the sample 0.025 and 0.975 quantiles of the bootstrap distribution of \hat{v}_0 .

Table 1
Coverage and length of 95% confidence intervals for data generated by Model 1.

		$x_0 =$	0	0.25	0.5	0.75	1
Coverage	\mathcal{C}_W		0.97	0.95	0.93	0.93	0.92
	\mathcal{C}_I		0.95	0.95	0.93	0.92	0.92
	\mathcal{C}_W^*		0.97	0.95	0.93	0.91	0.92
	\mathcal{C}_I^*		0.97	0.96	0.94	0.92	0.92
	\mathcal{C}^*		0.96	0.95	0.93	0.92	0.93
Length	\mathcal{C}_W		0.43	0.49	0.62	0.80	1.00
	\mathcal{C}_I		0.44	0.49	0.63	0.81	1.02
	\mathcal{C}_W^*		0.43	0.49	0.62	0.81	1.02
	\mathcal{C}_I^*		0.46	0.51	0.65	0.85	1.08
	\mathcal{C}^*		0.44	0.49	0.63	0.81	1.02

Table 2

Coverage and length of 95% confidence intervals for data generated by Model 2. An average length of ∞ indicates that the upper confidence limit was positive infinite.

	$x_0 =$	0	0.25	0.5	0.75	1
Coverage	\mathcal{C}_W	1.00	0.91	0.89	0.85	0.83
	\mathcal{C}_I	0.94	0.96	0.92	0.90	0.87
	\mathcal{C}_W^*	1.00	0.89	0.88	0.88	0.84
	\mathcal{C}_I^*	0.94	0.96	0.93	0.93	0.90
	\mathcal{C}^*	0.95	0.95	0.93	0.93	0.90
Length	\mathcal{C}_W	0.02	0.64	1.34	2.39	4.46
	\mathcal{C}_I	0.05	∞	∞	∞	∞
	\mathcal{C}_W^*	0.05	1.13	3.33	5.75	7.20
	\mathcal{C}_I^*	0.05	∞	∞	∞	∞
	\mathcal{C}^*	0.05	0.70	2.04	4.18	5.66

Table 3
Summary of results for the bladder volume example.

Interval	Estimate	SE	95% confidence limits
\mathcal{C}_W	120.53	26.20	(69.18, 171.88)
\mathcal{C}_I	120.53	—	(75.39, 184.70)
\mathcal{C}_W^*	120.53	27.40	(64.54, 171.96)
\mathcal{C}_I^*	120.53	27.40	(74.97, 186.62)
\mathcal{C}^*	120.53	27.40	(76.05, 182.24)