STAT/BIOST 571: Homework 7

Philip Pham

March 5, 2019

Problem 1: Relationship between fixed effects, random effects, GLS, and penalized regression; confounding; model misspecification (20 points)

This is an extension of problem 3 from homework 1. There are n subjects, indexed by $i = 1, \ldots, n$, each one of which is observed at m follow-up times, indexed by $j = 1, \ldots, m$. Let Y_{ij} be the literacy score of subject i at follow-up time j, at which time the subject's age is x_{ij}. The design is fixed, meaning that the x_{ij} are all deterministic and known in advance, and the true data-generating mechanism can be written

$$Y_{ij} = f(x_{i1}) + \beta_L(x_{ij} - x_{i1}) + \epsilon_{ij}$$

with i.i.d. $\epsilon_{ij} \sim N(0, \sigma^2)$. We are interested in estimating β_L, without knowing the form of $f(x)$. For this problem, fix $n = 91$ and $m = 3$, set the initial follow-up age for subject i to be

$$x_{i1} = 1 + 0.1 * (i - 1),$$

have subsequent follow-ups evenly spaced at intervals of

$$\Delta x_i = \left(1 + \frac{10 - x_{i1}}{10}\right)^2,$$

and consider the fixed (but unknown) mean model implied by $\beta_L = 1$ and $f(x) = (10 - x)^2$. Initially fix the true variance at $\sigma^2 = 100$, but you will consider different values of σ^2 later in the problem.

(a) *Simulate a single realization of data from the model specified above and generate one or more plots that allows you to visualize the marginal and conditional trends in computer literacy.*

   Solution: A single realization is plotted in Figure 1. By looking at the data from follow period 0 and comparing subjects, one can identify the marginal trend: literacy decreases with the base age, so younger subjects generally have higher literacy. To see the conditional trend, one looks within each subject: generally, as a subject ages, their literacy increases. This conditional trend isn't very strong, however, since the variance is rather large relative to the effect size.

Based on the plot you have generated (and experience from homework 1), it should be clear that trying to estimate β_L by fitting a partitioned model in which you treat $f(x)$ as linear will not work. You explain
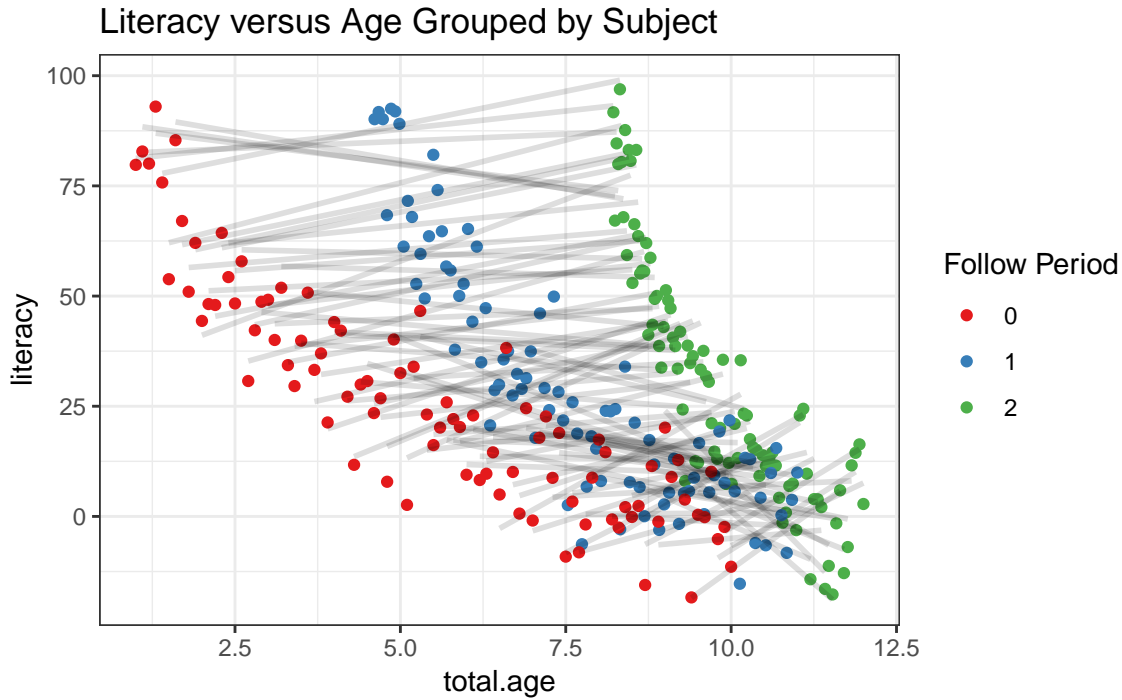
1

Figure 1: Plot of a single realization of the data with $\sigma^2 = 100$. The lines are fitted with OLS linear regression conditioned on each subject.

this to your collaborator and propose to fit a model where you include a fixed effect to stratify by subject. Your collaborator is sympathetic to the need to deal with confounding but is concerned that fitting a model with 92 parameters and only $91 \times 3 = 273$ observations is not a very good idea. He suggests, as an alternative, adjusting for subject using a random intercept in a linear mixed effects model (fit using REML, of course).

(b) *Explain to your collaborator, purely in words, why this is a bad idea.*

Solution: The random effects model assumes that the subject-specific adjustments to the intercept are normally distributed. This is not the case. See the histograms in Figure 2. In particular, look at the base literacy rate in follow period 0. This distribution is not normal and skews left. Thus, a random effects model would not properly account for the subject-specific variation.

Although you're sure that your explanation should have been persuausive, your collaborator is not yet convinced. So you conduct a simulation to further elucidate the differences between estimating β_L from models with fixed effects and random effects for subject-specific intercepts.

(c) *Write down the mathematical form of the two models you will be fitting and describe the simulation study.*

Solution: Let Y_i be the observed literacy rates for subject i. Let X_i be the covariates for subject i: an $m \times 2$ matrix of 1s in the first column and delta ages in the second. $\beta = \begin{pmatrix} \beta_0 & \beta_L \end{pmatrix}^\top$ are the fixed effect coefficients for the intercept and delta age.
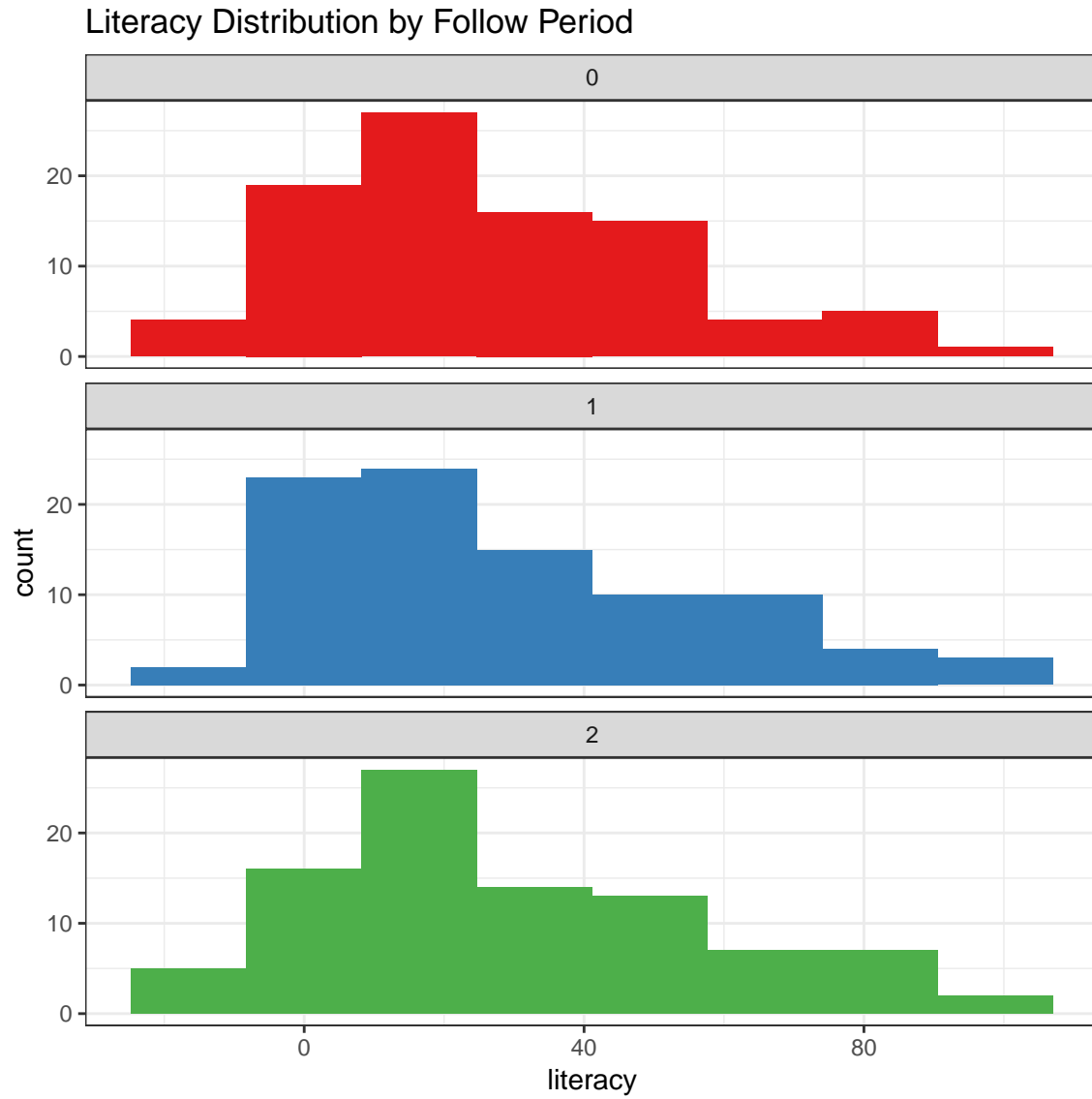
2

Figure 2: Distribution of literacy by the follow period. One sees that the distribution skews left and is not normal.

In the random effects model, we have that

$$Y_i = X_i\beta + \mathbf{1}\gamma_i + \epsilon_i \tag{1}$$
$$\gamma_i \sim \mathcal{N}\left(0, \sigma_\gamma^2\right)$$
$$\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right).$$

The γ_i induces within-subject correlation. This model will be fit with a restricted maximum likelihood (REML) procedure.

In the fixed effects model, we have that

$$Y_i = \begin{cases} X_i\beta + \epsilon_i, & i = 1 \\ X_i\beta + \mathbf{1}\alpha_i + \epsilon_i, & i \neq 1 \end{cases} \tag{2}$$
$$\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right).$$

The α_i are subject-specific intercepts and considered fixed effects in this case. We can fit this model with ordinary least squares (OLS) regression that maximizes likelihood.

(d) *Present and explain the results of your simulation study to your collaborator, highlighting the simulation-based estimates of bias and standard errors of the $\hat{\beta}_L$ you obtain from each model and how these relate to his suggestion.*

	$\mathbb{E}\left[\hat{\beta}_L\right]$	$\mathbb{E}\left[\hat{\sigma}_{\hat{\beta}_L}\right]$	Sample $\hat{\sigma}_{\hat{\beta}_L}$	$\mathbb{E}\left[\hat{\sigma}\right]$	$\mathbb{E}\left[\hat{\sigma}_\gamma\right]$
Random Effects Intercept	1.277928	0.320807	0.327997	10.007446	24.341198
Fixed Effects Intercept	0.996969	0.321357	0.319991	9.980246	

Table 1: Results of a simulation study comparing modeling the subject-specific intercepts as a random effect or fixed effect. Parameter estimates were averaged over simulations. Standard errors for $\hat{\beta}_L$ are calculated two ways: (1) assuming the model is correct ($\mathbb{E}\left[\hat{\sigma}_{\hat{\beta}_L}\right]$), and (2) using the $\hat{\beta}_L$ samples (Sample $\hat{\sigma}_{\hat{\beta}_L}$).

Solution: 2^{13} simulations were done for each model. Equation 1 is used in the random effects intercept model, and Equation 2 is used in the fixed effects intercept model. The results can be seen in Table 1.

The random effects model has significant bias and overestimates β_L, which has true value $\beta_L = 1$, whereas the fixed effects model appears unbiased. The random effects model underestimates the subject-specific intercepts and compensates by overestimating β_L.

The standard errors of both models agree. The model-based standard errors also agree with the simulated standard errors in both models.

(e) *Summarize the estimated values of the residual and random effect variances ($\hat{\sigma}^2$ and $\hat{\sigma}_\gamma^2$, respectively) from your simulations.*

Solution: σ^2 measures the variance from the ϵ_{ij} term in the model, which is the error in each individual observation. The estimated value of the residual variance $\hat{\sigma}^2$ is very accurate in both models. See the square root in the $\mathbb{E}\left[\hat{\sigma}\right]$ column of Table 1. The actual value is 10. σ_L^2 measures the variance associated from the γ_i term in the model, which accounts for subject-level differences. It's only estimated in the random effects model by $\hat{\sigma}_\gamma^2$.

4

(f) *Although the subject-specific intercepts in this problem are not really random, there is a natural quantity that you can think of $\hat{\sigma}_\gamma^2$ as trying to estimate. Calculate this number and call it σ_γ^2. Explain how your estimates $\hat{\sigma}_\gamma^2$ compare to σ_γ^2.*

  Solution: σ_γ^2 can be thought of as the variance in the base literacy score. In this case, the base literacy score for subject i is $f(x_{i1}) = f(1 + 0.1(i-1))$.

  We have that the standard error in base literacy is

$$\sigma_\gamma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f(x_{i1}) - \frac{1}{n}\sum_{j=1}^{n}f(x_1)\right)^2} \approx 24.433055478. \tag{3}$$

  Indeed, this number is very close to our estimate in the $\mathbb{E}[\hat{\sigma}_\gamma]$ column of Table 1.

For the remainder of the problem, consider the idealized situation where REML always estimates the variances exactly, meaning $\sigma^2 = \hat{\sigma}^2 = 100$ and $\sigma_\gamma^2 = \hat{\sigma}_\gamma^2$.

(g) *Calculate the exact (to your computer's numerical precision) bias of $\hat{\beta}_L$ from the random effect model as an estimator for β_L. Report your answer to five significant digits. Do this calculation two ways, first by interpreting the estimation procedure as GLS regression and then by interpreting it as penalized OLS regression.*

  Solution: Thre random effects model described in Equation 1 estimates β_L with a bias of $\boxed{0.27288324.}$ The calculations are described below.

  The random effects model implies a subject-specific covariance structure of

$$\Sigma_i = \sigma_\gamma^2 \mathbf{1}\mathbf{1}^\intercal + \sigma^2 I, \tag{4}$$

  where we use Equation 3 to compute σ_γ^2. Using Equation 4, we can compute $\hat{\beta}$ with the generalized least squares (GLS) estimator:

$$\hat{\beta}_{\text{GLS}} = \left(\sum_{i=1}^{n}X_i^\intercal\Sigma_i^{-1}X_i\right)^{-1}\sum_{i=1}^{n}X_i^\intercal\Sigma_i^{-1}Y_i. \tag{5}$$

  Equation 5 is implemented in `expect.beta.hat.gls`.

  Let Y be the concatenation of the Y_i. Let X be the $nm \times (n+2)$ design matrix for an overspecified OLS problem. The first two columns are a result of stacking the X_i and the last $j+2$ column represents an indicator for subject j. The first 2 coefficients are the fixed effects, and the remaining n are random effects, which we want to penalize. In penalized OLS regression, we are minimizing the mean squared error with an additional penalty term:

$$\hat{\beta}_{\text{OLS}} = \arg\min_{\beta}\left[(Y - X\beta)^\intercal(Y - X\beta) + \beta^\intercal Q\beta\right] \tag{6}$$

$$\text{where } Q = \begin{pmatrix} 0 \cdot I_2 & \mathbf{0} \\ \mathbf{0} & \frac{\sigma^2}{\sigma_\gamma^2} \cdot I_n \end{pmatrix},$$

which means the fixed effects are not penalized, and the random effects are penalized with weight $\frac{\sigma^2}{\sigma_\gamma^2}$ as in slide 3.70 of the course notes.

This is just ridge regression and has the closed-formed solution:

$$\hat{\beta}_{\mathrm{OLS}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_L \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_n \end{pmatrix} = (X^\intercal X + Q)^{-1} X^\intercal Y. \tag{7}$$

Equation 7 is implemented in `expect.beta.hat.ridge`.

Applying Equation 7 means solving a system of equations with $O(n)$ variables, which has complexity $O\left(n^3\right)$. If this is too expensive, gradient-based methods applied to Equation 6 can be used. The gradient-based method is implemented in `expect.beta.hat.penalized` and relies on the `penalized` package.

For both Equations 6 and 7, we can use linearity of expectation to calculate the expected value, $\mathbb{E}\left[Y_{ij}\right] = f\left(x_{i1}\right) + \beta_L x_{ij}$.

With $\sigma^2 = 100$, all three functions produce $\mathbb{E}\left[\hat{\beta}_L\right] = 1.27288324$, so there is bias of $\boxed{0.27288324}$ in the random effects model. This agrees with the results of the simulation study in Table 1.

It turns out that the closed form solution to ridge regression in Equation 7 is fastest.

(h) *Using the calculation from part (g) of this problem, plot the bias of $\hat{\beta}_L$ as a function of σ^2, over an interesting range of choices for σ^2 (continue to use the same fixed value of σ_γ^2 you estimated from the data).*

Solution: See Figure 3. The bias increases with σ^2.

(i) *Explain your plot from part (h) of this problem to your collaborator. Offer two explanations for the pattern in the plot, one based on the clustered WLS (equivalently, GLS) intepretation and the other based on the penalized OLS interpretation of the linear mixed effects model estimation algorithm.*

Solution: Recall the model in Equation 1. Consider the random vectors $Y_i - X_i\beta = \mathbf{1}\gamma_i + \epsilon_i$ obtained from subtracting out the fixed effects. The residuals can be decomposed into a within-subject and independent component.

GLS: From the GLS perspective, the random vector has distribution $\mathcal{N}\left(\mathbf{0}, \Sigma_i\right)$, where Σ_i takes the form in Equation 4, and we're maximizing likelihood.

Small values of σ^2 mean that the entries of the random vector are highly correlated, so each entry of the random vector should be almost identical. In this way, the model is encouraged to find the subject effect $f\left(x_{i1}\right)$ of the true model. The model weights the $\mathbf{1}\gamma_i$ term appropriately, and the bias of $\hat{\beta}_L$ is small.

Large values of σ^2 mean the entries of the random vector are uncorrelated, so there is no need to choose $\hat{\beta}$ such that the entries of the random vector are similar. The model is not able to disentangle the subject effect $f\left(x_{i1}\right)$ from the noise random vector ϵ_i. The
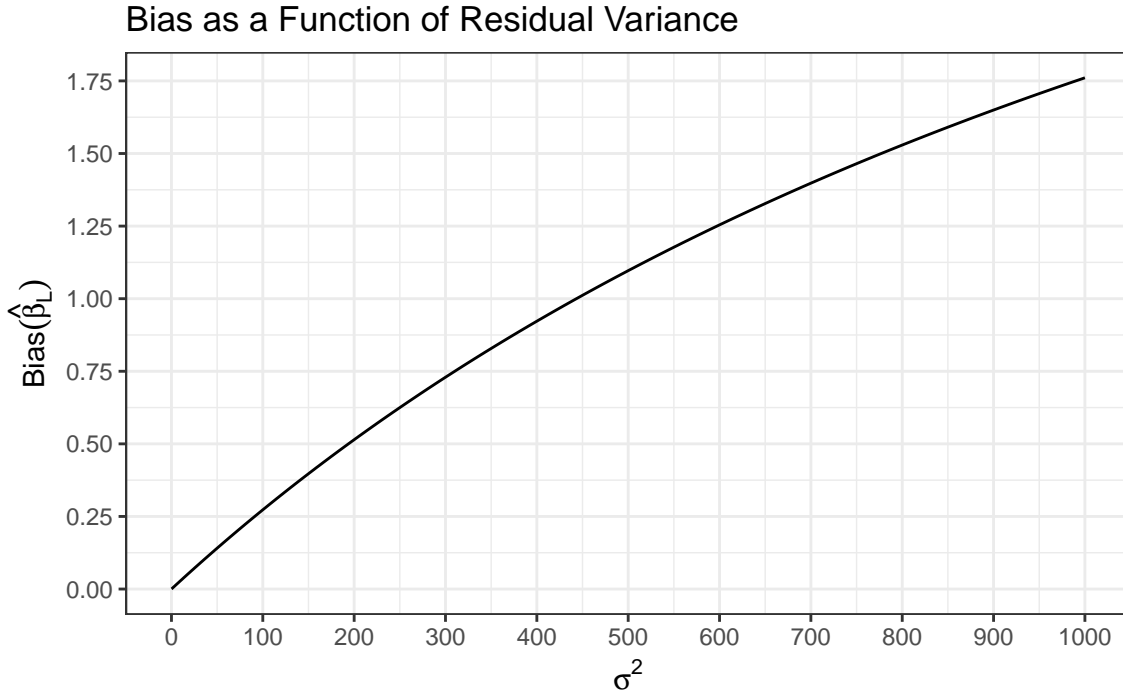
6

Bias as a Function of Residual Variance

Figure 3: The bias of $\hat{\beta}_L$ is calculated with Equation 7 varying σ^2 from 10^{-1} to 10^3.

difference between a subject's base literacy and β_0 is considered noise. Base literacy skews left (see Figure 2), so β_0 is underestimated. The model compensates by overestimating β_L. As $\sigma^2 \to \infty$, the model approaches the OLS estimates.

Penalized OLS: From the penalized OLS perspective, there is no within-subject correlation structure. The γ_i are just additional fixed effects that are encouraged to be close to 0 by the objective function in Equation 6. σ^2 is proportional to the penalty.

When the penalty is small, γ_i are allowed to vary freely. Since the true model will minimize the expected mean squared error, $\gamma_i \to f(x_{i1}) - \beta_0$ as $\sigma^2 \to 0$. Thus, we estimate β_L with small bias.

When the penalty is large, γ_i are pushed towards 0. As σ^2 gets large, the penalized OLS estimate approaches the estimate of OLS with just two coefficients β_0 and β_L. Without these additional effects, the left skew of base literacy results in underestimating β_0 (see Figure 2). To fit the data better in this situation, β_L is overestimated.

(j) *Your collaborator is nearly convinced, but not quite. He reminds you that you once told him mixed model effect estimates can be thought of as solutions to a GLS problem and that any old GLS estimator will do (regardless of correlation structure) to consistently estimate the coefficients in a linear regression model. Explain why this argument doesn't apply in the present situation.*

Solution: Consider the GLS estimator in Equation 5. We have

$$\mathbb{E}\left[\hat{\beta}_{\text{GLS}}\right] = \left(\sum_{i=1}^{n} X_i^{\intercal}\Sigma_i^{-1}X_i\right)^{-1}\sum_{i=1}^{n} X_i^{\intercal}\Sigma_i^{-1}\mathbb{E}\left[Y_i\right] \tag{8}$$

$$= \left(\sum_{i=1}^{n} X_i^{\intercal}\Sigma_i^{-1}X_i\right)^{-1}\sum_{i=1}^{n} X_i^{\intercal}\Sigma_i^{-1}\mathbb{E}\left[X_i\beta + \mathbf{1}\gamma_i + \epsilon_i\right]$$

$$= \beta + \left(\sum_{i=1}^{n} X_i^{\intercal}\Sigma_i^{-1}X_i\right)^{-1}\sum_{i=1}^{n} X_i^{\intercal}\Sigma_i^{-1}\mathbf{1}\mathbb{E}\left[\gamma_i\right].$$

The GLS estimator is only unbiased if and only if $\left(\sum_{i=1}^{n} X_i^{\intercal}\Sigma_i^{-1}X_i\right)^{-1}\sum_{i=1}^{n} X_i^{\intercal}\Sigma_i^{-1}\mathbf{1}\mathbb{E}\left[\gamma_i\right] = \mathbf{0}$.
In the true model, the γ_i are not random. They are fixed $\gamma_i = f\left(x_{i1}\right) - \beta_0$, where $\beta_0 = \frac{1}{n}\sum_{i=1}^{n} f\left(x_{i1}\right)$. A realization of the $f\left(x_{i1}\right)$ with noise can be seen in Figure 2 in the follow period 0 histogram. The distribution skews left and is not normal. Morever, in Figure 1, one sees that it's value is correlated with x_{i1}.

GLS is fitted by maximizing likelihood and assumes normality, so it adjusts the estimate of $\hat{\beta}$ so that the distribution of the $\hat{\gamma}_i = Y_{ij} - \hat{\beta}_0 - \hat{\beta}_L x_{ij}$ fits a normal distribution with mean 0. γ_i is the source of within-subject correlation. In this way, there is a tension between correctly accounting for the within-subject correlation and having the residuals $Y_i - X_i\beta$ maximize the multivariate normal likelihood model.

Given that the true distribution is not normal, the fitting procedure must choose biased values for $\hat{\beta}$. In particular, the left skew results in an underestimate of β_0. To better fit the data, β_L must then be overestimated.

Appendix

The following pages contain code to reproduce tables and plots. Functions to computed the expected value of $\hat{\beta}$ are implemented.

Relationship between fixed effects, random effects, GLS, and penalized regression; confounding; model misspecification

```
In [1]:  library(data.table)
         library(ggplot2)
         library(lme4)
         library(parallel)
         library(penalized)
         library(xtable)

         options(repr.plot.width=6, repr.plot.height=4)
         theme_set(theme_bw())
```

```
Loading required package: Matrix
Loading required package: survival
Welcome to penalized. For extended examples, see vignette("penalized").
```

Prepare Data

```
In [2]:  literacy.covariates.data <- data.table(subject=c(1:91))
         literacy.covariates.data[,base.age:=1 + 0.1*(subject - 1)]
         literacy.covariates.data[,`0`:=base.age]
         literacy.covariates.data[,`1`:=`0` + (1 + (10 - base.age)/10)^2]
         literacy.covariates.data[,`2`:=`1` + (1 + (10 - base.age)/10)^2]
         literacy.covariates.data[,subject:=factor(subject)]
         literacy.covariates.data <- melt(literacy.covariates.data, id.vars=c('subject', 'base.age'),
                                          value.name='total.age', variable.name='follow.per')
         literacy.covariates.data[,follow.per:=factor(as.numeric(follow.per) - 1, ordered=TRUE)]
         setkey(literacy.covariates.data, subject, follow.per)
         literacy.covariates.data[,delta.age:=total.age - base.age]
         head(literacy.covariates.data)
```

subject	base.age	follow.per	total.age	delta.age
1	1.0	0	1.0000	0.0000
1	1.0	1	4.6100	3.6100
1	1.0	2	8.2200	7.2200
2	1.1	0	1.1000	0.0000
2	1.1	1	4.6721	3.5721
2	1.1	2	8.2442	7.1442

Response

```
In [3]:  make.base.literacy <- function(base.age) {
             (10 - base.age)^2
         }

         make.response <- function(covariates, variance=100) {
             make.base.literacy(covariates$base.age) +
             covariates$delta.age +
             rnorm(n=nrow(covariates), sd=sqrt(variance))
         }
```

Visualization

```
In [4]: set.seed(2021)
        literacy.data <- copy(literacy.covariates.data)
        literacy.data[,literacy:=make.response(literacy.data)]

        p <- ggplot(literacy.data, aes(x=total.age, y=literacy)) +
          geom_point(aes(color=follow.per)) +
          scale_color_brewer('Follow Period', palette='Set1') +
          geom_smooth(aes(group=subject), method='lm', se=FALSE, color='#00000022') +
          ggtitle('Literacy versus Age Grouped by Subject')

        pdf('literacy_versus_age.pdf', width=6, height=3.75)
        p
        dev.off()
        p
```

png: 2

Literacy versus Age Grouped by Subject
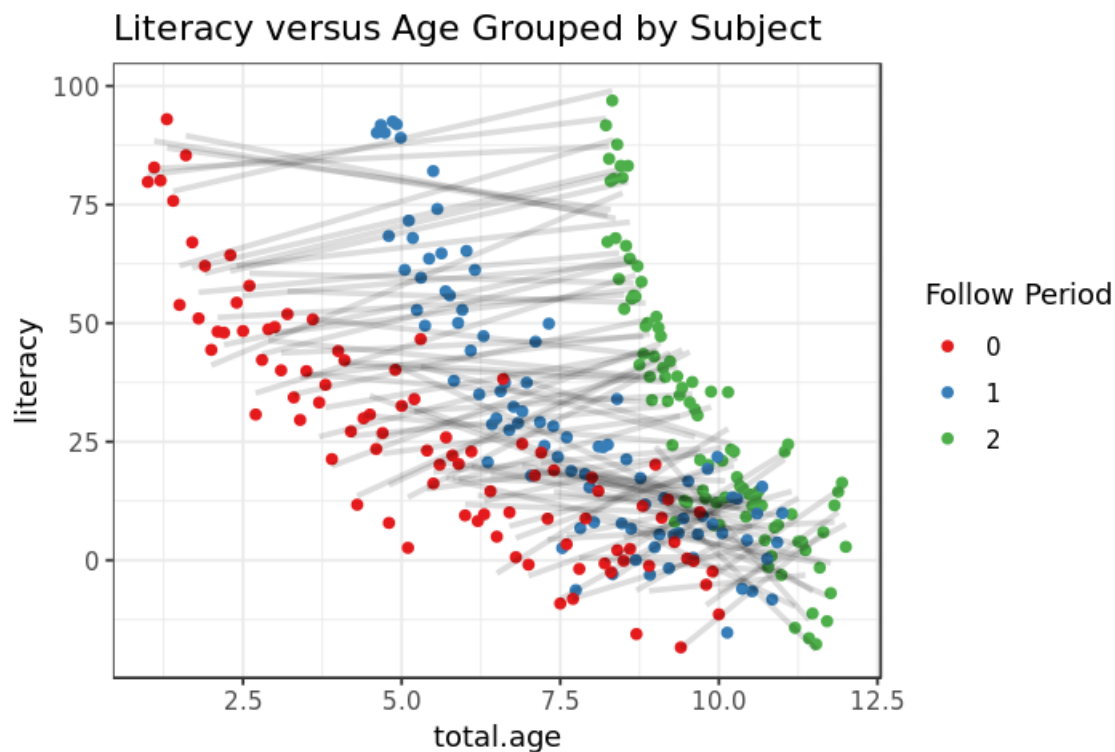
```
In [5]:  p <- ggplot(literacy.data, aes(x=literacy, fill=follow.per)) +
            geom_histogram(bins=8) +
            scale_fill_brewer('Follow Period', palette='Set1', guide=FALSE) +
            facet_wrap(~follow.per, ncol=1) +
            ggtitle('Literacy Distribution by Follow Period')

         pdf('literacy_by_follow_period.pdf', width=6, height=6)
         p
         dev.off()
         p
```
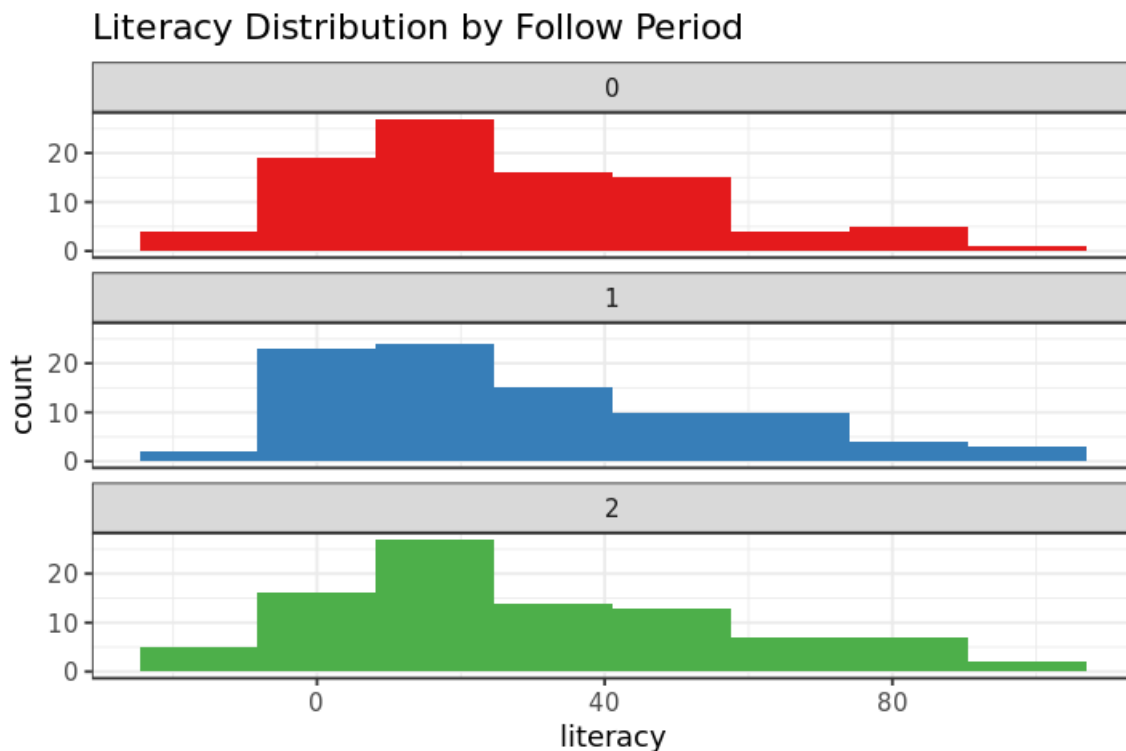
png: 2

Simulation Study: Random versus Fixed Effect Intercept

```
In [6]:  lmer.simulation.samples <- do.call(
            rbind, mclapply(replicate(8192, literacy.covariates.data, simplify=FALSE),
                            function(covariates) {
            literacy <- make.response(covariates)
            model <- lmer(literacy ~ delta.age + (1|subject), data=covariates, REML=TRUE)
            var.corr <- VarCorr(model)
            data.frame(
                delta.age=fixef(model)[['delta.age']],
                delta.age.std.error=summary(model)$coefficients['delta.age', 'Std. Error'],
                sigma=attributes(var.corr)$sc,
                sigma.random=attributes(var.corr$subject)$stddev[['(Intercept)']])
         }, mc.cores=4))
```

```
In [7]:  lm.simulation.samples <- do.call(
             rbind, mclapply(replicate(8192, literacy.covariates.data, simplify=FALSE),
                             function(covariates) {
             literacy <- make.response(covariates)
             model <- lm(literacy ~ subject + delta.age, data=covariates)
             model.summary <- summary(model)
             data.frame(
                 delta.age=model$coefficients[['delta.age']],
                 delta.age.std.error=model.summary$coefficients['delta.age', 'Std. Error'],
                 sigma=model.summary$sigma)

         }, mc.cores=4))
```

```
In [8]:  summarize.simulation <- function(data) {
             l <- list(
                 `$\\mathbb{E}\\left[\\hat{\\beta}_L\\right]$`=mean(data$delta.age),
                 `$\\mathbb{E}\\left[\\hat{\\sigma}_{\\hat{\\beta}_L}\\right]$`=mean(data$delta.age.std.
         error),
                 `Sample $\\hat{\\sigma}_{\\hat{\\beta}_L}$`=sd(data$delta.age),
                 `$\\mathbb{E}\\left[\\hat{\\sigma}\\right]$`=mean(data$sigma))
             l[['$\\mathbb{E}\\left[\\hat{\\sigma}_{\\gamma}\\right]$']] <- if ('sigma.random' %in% name
         s(data)) {
                 mean(data$sigma.random)
             } else {
                 NA
             }
             l
         }
```

```
In [9]:  simulation.comparison <- as.data.frame(rbind(
             data.table(lmer.simulation.samples)[,summarize.simulation(.SD)],
             data.table(lm.simulation.samples)[,summarize.simulation(.SD)]))
         row.names(simulation.comparison) <- c('Random Effects Intercept', 'Fixed Effects Intercept')
         simulation.comparison
```

	$\mathbb{E}\left[\hat{\beta}_L\right]$	$\mathbb{E}\left[\hat{\sigma}_{\hat{\beta}_L}\right]$	Sample $\hat{\sigma}_{\hat{\beta}_L}$	$\mathbb{E}\left[\hat{\sigma}\right]$	$\mathbb{E}\left[\hat{\sigma}_{\gamma}\right]$
Random Effects Intercept	1.2779277	0.3208065	0.3279972	10.007446	24.3412
Fixed Effects Intercept	0.9969691	0.3213570	0.3199906	9.980246	NA

```
In [10]:  print(xtable(simulation.comparison,
                   caption=paste('\\small Results of a simulation study comparing modeling the ',
                               'subject-specific intercepts as a random effect or fixed effect.',
                               'Parameter estimates were averaged over simulations. Standard error
          s',
                               'for $\\hat{\\beta_L}$ are calculated two ways:',
                               '(1) assuming the model is correct',
                               '($\\mathbb{E}\\left[\\hat{\\sigma}_{\\hat{\\beta}_L}\\right]$),',
                               'and (2) using the $\\hat{\\beta_L}$ samples',
                               '(Sample $\\hat{\\sigma}_{\\hat{\\beta}_L}$).'),
                   label='tab:simulation_comparison',
                   digits=c(0, 6, 6, 6, 6, 6)),
              booktabs=TRUE,
              sanitize.colnames.function=identity,
              sanitize.rownames.function=identity,
              size='small',
              file='simulation_comparison.tex')
```

σ_γ Calculation

```
In [11]:  base.literacy <- make.base.literacy(literacy.covariates.data[follow.per==0]$base.age)
          variance.random.intercept <- mean((base.literacy - mean(base.literacy))^2)
          sqrt(variance.random.intercept)
```

24.4330554781836

Expected $\hat{\beta}$

GLS

```
In [12]:  expect.beta.hat.gls <- function(residual.variance) {
              subject.covariance.inv <- chol2inv(chol(
                  as.matrix(nlme::pdCompSymm(variance.random.intercept + diag(3)*residual.variance))))
              projected.data <- lapply(levels(literacy.covariates.data$subject), function(i) {
                  data <- literacy.covariates.data[J(i)]
                  X <- cbind(1, data$delta.age)
                  tX <- t(X)
                  list(X=tX %*% subject.covariance.inv %*% X,
                       y=tX %*% subject.covariance.inv %*% make.response(data, 0))
              })
              projected.X <- Reduce(
                  `+`, lapply(projected.data, function(projection) { projection$X }))
              projected.y <- Reduce(
                  `+`, lapply(projected.data, function(projection) { projection$y }))
              r <- chol(projected.X)   # Upper trianglar
              beta.hat <- as.vector(backsolve(r, forwardsolve(r, projected.y, upper.tri=TRUE, transpose=T
          RUE)))
              names(beta.hat) <- c('(Intercept)', 'delta.age')
              beta.hat
          }
```

Ridge Regression

```
In [13]:  expect.beta.hat.ridge <- function(residual.variance) {
              X <- cbind(1, model.matrix(~ 0 + delta.age + subject, literacy.covariates.data))
              y <- make.response(literacy.covariates.data, 0)
              # Penalize only random effects.
              Q <- diag(ncol(X))
              Q[1, 1] <- 0
              Q[2, 2] <- 0
              # Estimate beta
              tX <- t(X)
              r <- chol(tX %*% X + residual.variance/variance.random.intercept*Q)
              beta.hat <- as.vector(backsolve(r, forwardsolve(r, tX %*% y, upper.tri=TRUE, transpose=TRUE
          )))
              names(beta.hat) <- colnames(X)
              names(beta.hat)[1] <- '(Intercept)'
              beta.hat
          }
```

Penalized Regression

```
In [14]:  expect.beta.hat.penalized <- function(residual.variance) {
              coef(penalized(make.response(literacy.covariates.data, 0),
                             unpenalized=~delta.age,
                             penalized=~subject,
                             data=literacy.covariates.data,
                             lambda2=residual.variance/variance.random.intercept))
          }
```

```
In [15]:  expect.beta.hat.gls(100)['delta.age']
          expect.beta.hat.ridge(100)['delta.age']
          expect.beta.hat.penalized(100)['delta.age']
```

delta.age: 1.27288323972858

delta.age: 1.27288323972858

delta.age: 1.27288323972857

Bias as a Function of σ^2

```
In [16]:  bias.data <- data.table(residual.variance=c(seq(0.1, by=0.1, length.out=9), 1:1000))
          bias.data[,bias:=do.call(c, mclapply(
              residual.variance,
              function(residual.variance) { expect.beta.hat.ridge(residual.variance)[['delta.age']] },
              mc.cores=4)) - 1]

          p <- ggplot(bias.data, aes(x=residual.variance, y=bias)) +
            geom_line() +
            scale_x_continuous(expression(sigma^2), breaks=seq(0, 1000, by=100)) +
            scale_y_continuous(expression(Bias(hat(beta)[L])), breaks=seq(0, 1.75, by=0.25)) +
            ggtitle('Bias as a Function of Residual Variance')

          pdf('bias_variance_plot.pdf', width=6, height=3.75)
          p
          dev.off()
          p
```

png: 2

Bias as a Function of Residual Variance