Mixed Effects Models for Size-Attained Data

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Lisa M. Lendway

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Sanford Weisberg, Adviser

January 2012

© Lisa M. Lendway 2012

Acknowledgements

Thank you to my committee members – Charles Anderson, Birgit Grund, and Galin Jones – for providing useful feedback and encouragement. I would especially like to thank my advisor, Sandy Weisberg, who always made me feel confident in myself. Of course, he also taught me a great deal about statistics and writing. I cannot thank him enough.

Above all, I thank my family. Their support was undying and they helped me out in more ways than I have room to write. I love you all. Thank you.

Dedication

This dissertation is dedicated to my daughter, Adeline Mary Krieger. Throughout this process, she was always able to bring a smile to my face and remind me of the important things in life.

Abstract

It is rare to have longitudinal data on the somatic growth of fish, that is, how their body length changes over time. In most temperate fish, scales or other hard parts, like otoliths or other bones, form annual rings or increments. Growth of the hard part can be measured, thereby giving a longitudinal record of hard part growth from cross-sectional data. Methods such as back-calculation and linear mixed-effects models have used the growth of hard parts to infer somatic growth.

At times, it is not feasible to obtain the measurements of the hard part. Body length at time of capture is much easier to measure and reflects somatic growth, which is usually of more interest. In this thesis, I present a model that is based on a longitudinal approach but models length at time of capture, rather than the yearly body growth. It also allows for estimation of environmental impact on growth.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and History

Understanding fish growth is important in fisheries management. In this chapter, we give an overview of some of the most common methods used in modeling fish growth. We start with an introduction to back-calculation, a method of inferring somatic growth from growth of scales or other hard parts of fish, which has been used for over a century. We then present a few relatively new methods of modeling fish growth that use linear models, the first of which was introduced by Weisberg (1986). This method will be the basis of a new model presented in Chapter 2. We also discuss similar models used in modeling tree growth. Lastly, we present a Bayesian version of the Weisberg et al. (2010) model.

1.1 Back-calculation

For over a century the method of *back-calculation* has been used to approximate the incremental somatic growth of fish from cross-sectional data, most often measured by length (Bartlett et al., 1984; Francis, 1990; Vigliola and Meekan, 2009). Scales or other hard parts, such as otoliths or other bones, which in temperate fish form annual rings or increments, are used to infer somatic growth. Throughout this chapter, we use the word "scale" to mean any hard part. There are many back-calculation methods, each

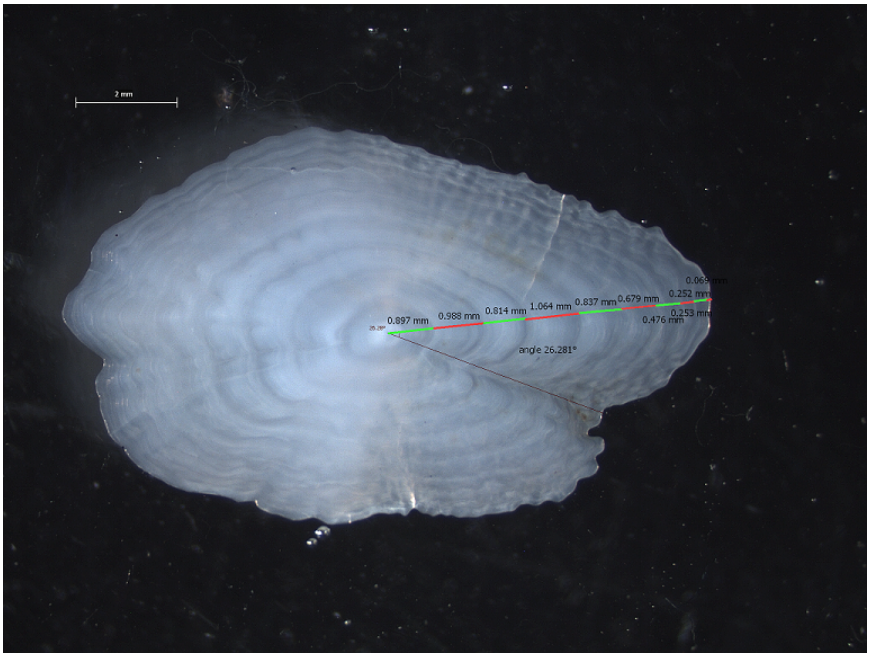


Figure 1.1: Sectioned otolith from a Pacific halibut, showing annual growth as indicated by the red and green markers (Richmond, 2009).

of which is based on assumptions about how growth of the scale and somatic growth are related. We will provide details for a few of the methods. In general, for each fish, data are collected on the age and body length at time of capture. Additionally, the radius to the ring formation is measured for all available rings, which is considered be where the scale ended at that age. They are measured in a direction fixed by a conventional protocol, since clearly they depend on the direction of measurement, as the scales are generally not circular. Figure 1.1 illustrates the ring formations on an otolith from a Pacific halibut. Letting L_k = observed body length of fish k at time of capture, S_k = scale radius of fish k at time of capture, L_{ki} = the unobserved body length of fish k at age i, and S_{ki} = observed scale radius of fish k at age i (measured by focus to ring i), the goal of back-calculation is to estimate L_{ki} from L_k, S_k, and S_{ki}.

The three methods of back-calculation we will discuss are *regression*, *Fraser-Lee*, and *modified Fry*. Each of these methods results in a *back-calculation formula* (Francis, 1990), which provides a family of lines or curves that are used to estimate the L_{ki} for each fish. A specific back-calculation line is created for each fish by inserting that fish's S_k and L_k values into the back-calculation formula. An important part of back-calculation is understanding the relationship between body length and scale radius at time of capture. Although other functions could be used, it is common to assess the relationship by either regressing S_k on L_k or regressing L_k on S_k. The regression equations are written as

$$S_k = a + bL_k + \varepsilon_k \tag{1.1}$$

$$L_k = c + dS_k + \varepsilon_k. \tag{1.2}$$

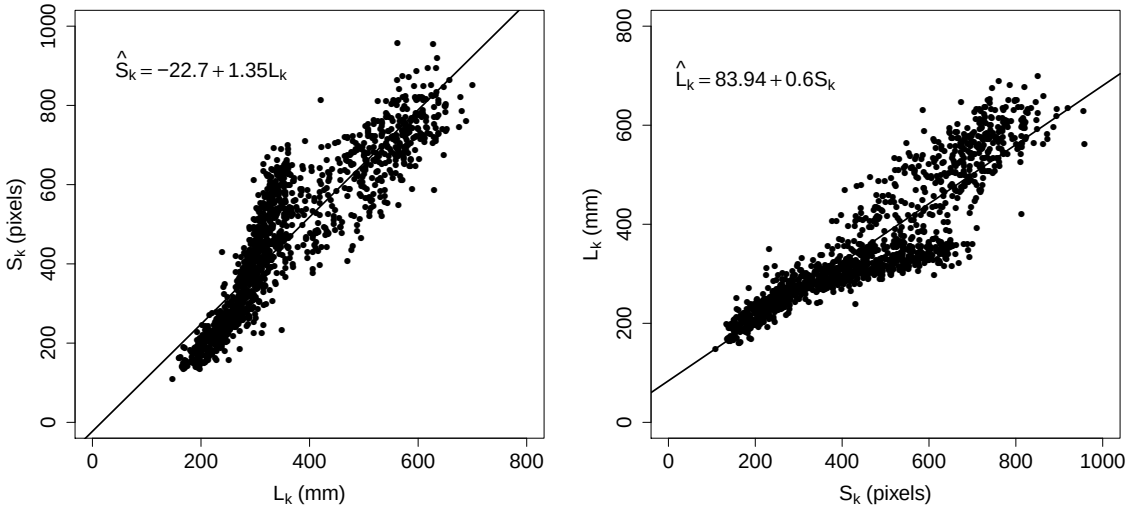
Figure 1.2 shows an example of these two regressions for a sample of 1297 freshwater drum captured from Lake Winnebago, Wisconsin (for more detail, see Chapter 5). This sample includes both males and females, which exhibit different growth patterns.

The regression method of back-calculation is the simplest but also, as Francis (1990) says, "It is hard to see any reason for the use of regression methods." It does not use any information about an individual fish's body length or scale radius at time of capture. Rather, the same back-calculation line is used for all fish. If a linear relationship between body length and scale radius is assumed, then the regression of body length on scale radius,

$$\hat{L}_k = \hat{c} + \hat{d}S_k, \tag{1.3}$$

is used as the back-calculation formula. So, L_{ki} is estimated simply by putting S_{ki} into (1.3) for S_k. An example of this regression is shown in Figure 1.2b.

In his review and critique of back-calculation, Francis (1990) found that in the 54



(a) Regression of scale radius (S_k) on body length (L_k).

(b) Regression of body length (L_k) on scale radius (S_k).

Figure 1.2: Scatterplots and regression equations used to assess the relationship between body length and scale radius.

papers he examined, the most common method of back-calculation was Fraser-Lee, which was used in 25 of the papers. The Fraser-Lee method makes the assumption that scale and body length yearly increments grow proportionally. Mathematically, this is written as

$$\frac{L_{k2} - L_{k1}}{S_{k2} - S_{k1}} = \frac{L_{k3} - L_{k2}}{S_{k3} - S_{k2}} = \cdots = \frac{L_{kn} - L_{k(n-1)}}{S_{kn} - S_{k(n-1)}}. \tag{1.4}$$

This results in the back-calculation line for a particular fish passing through the point (S_k, L_k). To define a unique line for each fish, it is also assumed that the back-calculation line passes through the point $(0, c)$, where c is the value of L_k when $S_k = 0$ from regressing L_k on S_k (as in (1.2) and shown in in Figure 1.2b). So, the back-calculation formula that results is

$$\hat{L}_{ki} = \frac{S_{ki}}{S_k}(L_k - \hat{c}) + \hat{c}. \tag{1.5}$$

Again, the back-calculation line for a particular fish is formed by putting that fish's L_k and S_k values into (1.5).

Whitney and Carlander (1956) noted that the Fraser-Lee method would follow the hypothesis that, "if the scale were 10 per cent larger when the fish was captured than the average scale for that size of fish, the scale would be 10 per cent larger than normal throughout the life" if c were the value of L_k when $S_k = 0$ from (1.1) (the L_k intercept in Figure 1.2a), rather than the c originally proposed. Carlander (1982) later suggested using standard c values for each species in (1.2), which were based on observing the actual size of young fish rather than extrapolating it from the regression of S_k on L_k or L_k on S_k. The standard c, called a biological intercept, would then be used in the Fraser-Lee formula.

Vigliola and Meekan (2009), by comparing growth curves from a visual survey to growth curves created from back-calculated lengths, found that the modified Fry

method of back-calculation led to the most accurate results. This method is a more complicated version of the biological intercept Fraser-Lee method. The hypothesis for this method, like the Fraser-Lee method, is that the scale and body length yearly increments grow proportionally, and it uses a biological intercept. The back-calculation formula is

$$\hat{L}_{ki} = \hat{a} + \exp\left\{ \log(\hat{L}_1 - \hat{a}) + \frac{\left[\log(L_k - \hat{a}) - \log(\hat{L}_1 - \hat{a})\right]\left[\log(S_{ki}) - \log(\hat{S}_1)\right]}{\log(S_k) - \log(\hat{S}_1)} \right\}. \quad (1.6)$$

The biological intercept is $(\hat{S}_1, \hat{L}_1)$, where $\hat{L}_1$ can be obtained from literature and is meant to be the mean fish length at first ring formation (Vigliola and Meekan, 2009). Then, $\hat{S}_1$ is estimated from the data by calculating the mean radius of first scale increment. That is, if N is the number of fish in the sample,

$$\hat{S}_1 = \frac{\sum_{k=1}^{N} S_{k1}}{N}.$$

The a parameter is estimated by first fitting the non-linear regression

$$L_k = \hat{L}_1 - b\hat{S}_1^c + bS_k^c + \varepsilon_k,$$

which gives estimates of b and c. Then, a is estimated by $\hat{a} = \hat{L}_1 - \hat{b}\hat{S}_1^{\hat{c}}$.

We will illustrate these methods through an example. Figure 1.3 shows the otolith radius at age, that is the S_{ki}, for one freshwater drum. For this fish, $S_k = 710$ pixels and $L_k = 480.06$ mm. Figure 1.2 shows the scatterplots of L_k on S_k and S_k on L_k and the associated regressions obtained from the entire sample of freshwater drum. Back-calculation lines using the regression, Fraser-Lee, and Modified Fry methods are shown in Figure 1.4. The vertical lines show the measurement of otolith focus at each ring formation. The y values corresponding to the intersection of the vertical lines

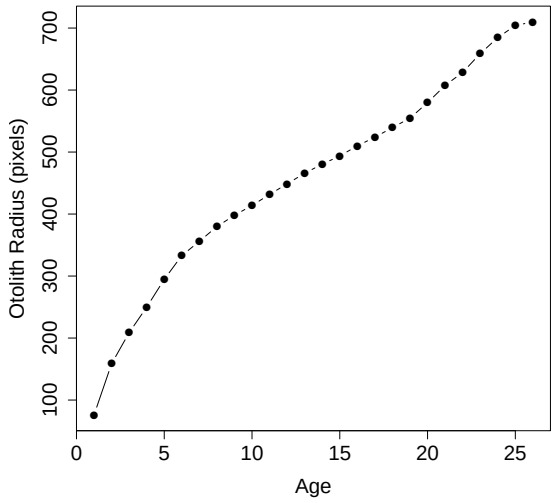


Figure 1.3: Otolith radius at age for one freshwater drum.

and the back-calculation lines are the estimated lengths at age, L_{ki}. For this fish, the Modified Fry method results in smaller estimated body lengths at the early ages, compared to the other two methods. The regression method always leads to larger estimated body lengths, compared to the other two methods.

Once the back-calculated length-at-age estimates are obtained for each fish, the length data are used to model growth and compare sub-populations. Many different methods are used, including growth curves and mixed models. None of these methods acknowledge that the back-calculated lengths have error associated with them (Francis, 1995).

Although it seems to make sense intuitively, back-calculation is rife with problems. In his review of back-calculation, Francis (1990, p.897) stated that "though the technique is widely used, it does not appear to be well understood" and is often misused or misinterpreted. He also pointed out that there is not a good way to test hypotheses to compare models. Additionally, we note that back-calculation makes no

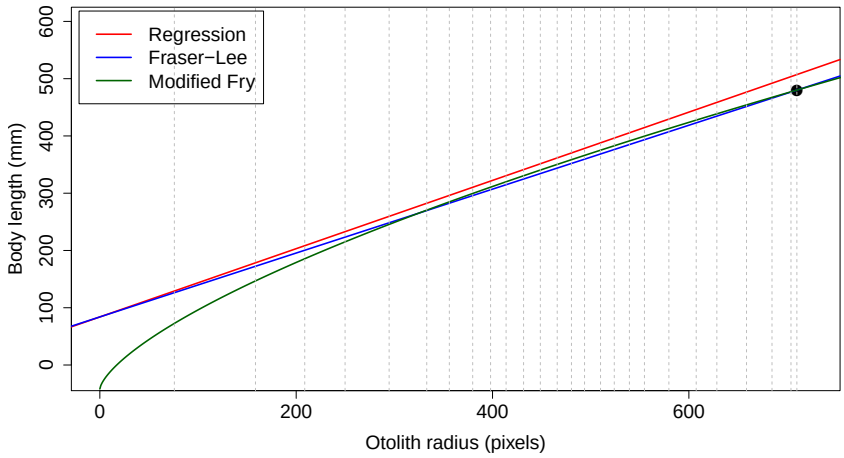


Figure 1.4: Three different back-calculation lines for a drum with $S_k = 710$ pixels and $L_k = 480.06$ mm (indicated by the black dot). The vertical lines show the measurement of otolith radius at each ring formation. The estimated lengths at age are where the vertical lines intersect the back-calculation lines.

allowance for variation in the environment. For example, a fish in the 1999 year-class has the same back-calculated values as one from 1989 if their S_k, S_{ki}, and L_k values are equal.

1.2 Fixed effects model

With a completely different methodology than back-calculation, Weisberg (1986) proposed modeling fish growth using linear models. This approach fixes some of the problems encountered in back-calculation by directly using growth of the scale radius, rather than first estimating past somatic growth. Doing this eliminates the additional variability caused by estimating lengths from the scale. A year effect is also added to model growth, which encompasses growth due to environmental effects not accounted for in age. The estimated year effects create a type of chronology that

could be compared with known environmental effects, like temperature or rainfall.

We discuss the generalized methodology of Weisberg (1993), which we call the *fixed-effects model*. This model assumes repeated measures of the growth indicator (scale increments) y_{cka} on each fish, where c is the year-class or year of birth of the fish, k is the fish number in that year, and a is the a^{th} growth increment, the increase in scale radius from age $a-1$ to a. This model equation is written as

$$y_{cka} = \iota_a + \eta_{c+a-1} + \varepsilon_{cka}. \tag{1.7}$$

The scale increment is modeled by an age (intrinsic) effect, ι_a, and a year (environmental) effect, η_{c+a-1}. Specifically, ι_a is the increase in scale radius from age $a-1$ to age a and η_{c+a-1} is the increase in scale radius from year $l = c+a-1$ to year $l = c+a$. The ε_{cka} allow for individual variability and are assumed to be independent random draws from a $N(0, \sigma^2)$. The normality and constant variance assumptions could be relaxed, which would add computational complexity.

The model is formulated in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1.8}$$

where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. The vector $\mathbf{y}$ contains the measurements of scale increments for each fish. It has length $\sum\sum p$, where p is the age at which fish k from year-class c was captured. The design matrix, $\mathbf{X}$, has $t+m$ columns, where t is the maximum age of all fish in the sample and m is the number of years in the data, calculated as ((most recent year of capture) − (earliest year-class)), and $\sum\sum p$ rows. The partial scale growth that occurred during the year of capture, that is, the scale increment from the p^{th} ring to the end of the scale, is not used since growth is not complete. $\boldsymbol{\beta}$ is a vector of length $t+m$ of coefficients, and $\boldsymbol{\varepsilon}$ is a vector of length $\sum\sum p$ of random errors.

The design matrix, $\mathbf{X}$, is composed of zeros and ones. Each row has exactly two ones. If the columns are arranged from age 1 to age t followed by the earliest year of birth to the most recent year, then there would be a 1 in the a^{th} column and the column corresponding to the year $c + a - 1$, when it was a years old, for observation y_{cka}. For example, suppose the oldest fish in the sample was four, the most recent year of capture was 2007, and the earliest year of birth was 1999. Then the $\mathbf{X}$ matrix would have $4 + (2007 - 1999) = 12$ columns. So, a 3 year old fish captured in year 2003, would have three rows of data in the $\mathbf{X}$ matrix:

$$\begin{pmatrix} \text{A1} & \text{A2} & \text{A3} & \text{A4} & \text{Y99} & \text{Y00} & \text{Y01} & \text{Y02} & \text{Y03} & \text{Y04} & \text{Y05} & \text{Y06} \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The fixed-effects model can be fit using ordinary least squares regression, so any statistical software package can be used to obtain the parameter estimates. There are some significant problems. Proponents of back-calculation argue that this method does not truly avoid back-calculation, but rather makes the same assumption as the regression method of back-calculation (Vigliola and Meekan, 2009). Further, Weisberg et al. (2010) mention two more problems. First, since $\text{rank}(\mathbf{X}) = t + m - 1$, estimates for the fixed effect parameters are not unique because they depend on which parameter is used as the baseline. That is all estimated age effects are really the difference between the effect that year and baseline year. Thus, estimates are difficult to compare across populations. Second, the error assumption is not realistic as multiple measurements from the same fish and multiple measurements from the same year are expected to be correlated.

1.3 Incremental model

In addition to pointing out faults in the fixed-effects model, Weisberg et al. (2010) proposed an improved model that treats the year effect as random, rather than fixed, and also includes a random fish effect. The new model, which we call the *incremental model*, is a mixed effects model and can be written as

$$y_{cka} = \iota_a + h_{c+a-1} + f_{ck} + \varepsilon_{cka}. \tag{1.9}$$

The response y_{cka}, the age effect ι_a, and the random error ε_{cka} are the same as in (1.7). The year effect h_{c+a-1} is now modeled as a random draw from $N(0, \sigma_h^2)$. Though the years themselves are certainly not random, it makes sense to consider how they affect growth each year as random. A random fish effect, f_{ck}, is also included, with $f_{ck} \sim N(0, \sigma_f^2)$, independently. We assume that the correlations between fish effects and year effects are zero, although this assumption can be changed.

In matrix notation, the model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}. \tag{1.10}$$

Here, $\boldsymbol{\beta}$ are the t age effects and $\mathbf{u}$ the m year effects and N fish effects. $\mathbf{X}$ is an $\sum\sum p \times t$ design matrix of mostly zeros, with a one in the a^{th} column for observation y_{cka}. The $\sum\sum p \times (m+N)$ $\mathbf{Z}$ design matrix has ones in two columns – the column corresponding to the year they were a years old and the column corresponding to their fish ID – and zeros everywhere else.

The incremental model induces a correlation structure that did not exist in the fixed effects model. We show this by computing two correlations – (a) between the yearly growth increments for two fish, say k_1 and k_2, in the same year-class and (b) between the yearly growth increments for the same fish in two different years, say a_1

and a_2.

$$\begin{aligned}
\text{(a) } \operatorname{cov}(y_{ck_1a}, y_{ck_2a}) &= \operatorname{cov}(h_{c+a-1} + f_{ck_1} + \varepsilon_{ck_1a}, h_{c+a-1} + f_{ck_1} + \varepsilon_{ck_1a}) \\
&= \operatorname{var}(h_{c+1-1}) = \sigma_h^2 \\
\text{(b) } \operatorname{cov}(y_{cka_1}, y_{cka_2}) &= \operatorname{cov}(h_{c+a_1-1} + f_{ck} + \varepsilon_{cka_1}, h_{c+a_2-1} + f_{ck} + \varepsilon_{cka_2}) \\
&= \operatorname{var}(f_{ck}) = \sigma_f^2
\end{aligned}$$

And, since $\operatorname{var}(y_{cka}) = \sigma_h^2 + \sigma_f^2 + \sigma^2$,

$$\begin{aligned}
\operatorname{corr}(y_{ck_1a}, y_{ck_2a}) &= \frac{\sigma_h^2}{\sigma_h^2 + \sigma_f^2 + \sigma^2} \\
\operatorname{corr}(y_{cka_1}, y_{cka_2}) &= \frac{\sigma_f^2}{\sigma_h^2 + \sigma_f^2 + \sigma^2}.
\end{aligned}$$

This model also creates estimates that are unique since $\operatorname{rank}(\mathbf{X}) = t$. So, estimates can be compared across different populations, thus addressing both deficiencies of the fixed-effects model.

Figure 1.5 shows the cumulative estimated age effects and predicted year effects with standard error bars from the fixed-effects model and the incremental model. It is clear that the standard errors from the incremental model are smaller than the standard errors from the fixed-effects model. The fixed-effects model uses year 1979 as a baseline, so all estimates are based on the year 1979 effect being zero. This seems to create a slight upward shift in the estimated year effects and a slight downward shift in the cumulative estimated age effects in the fixed-effects model.

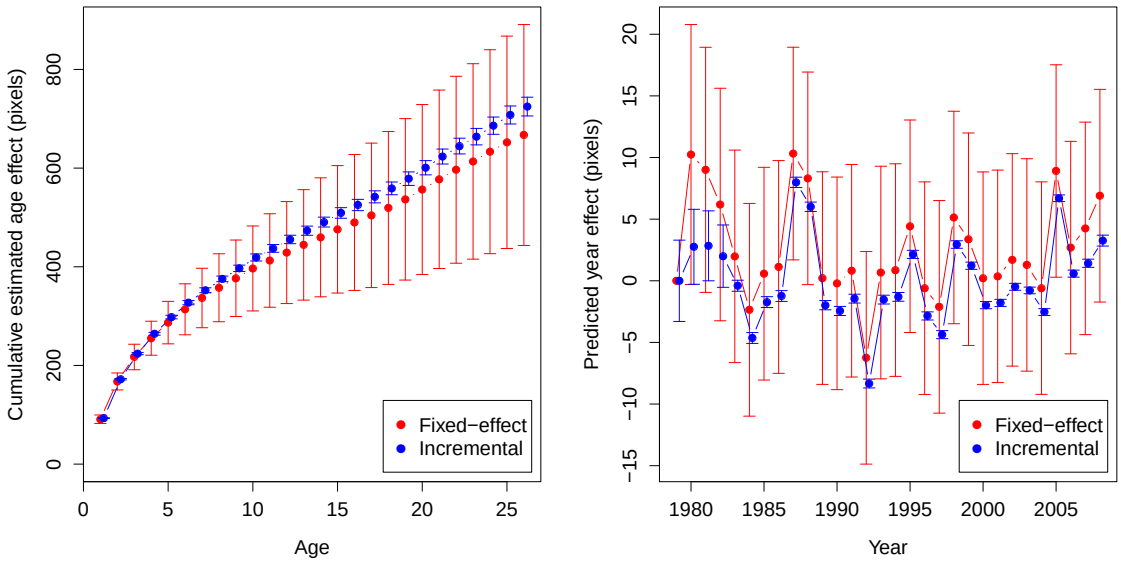


Figure 1.5: Cumulative estimated age effects and predicted year effects with standard error bars from the fixed-effects model and the incremental model.

1.4 Dendrochronology

The field of dendrochronology aims to use tree rings to reconstruct the environmental history of the area where the tree samples resided (Speer, 2010). Unlike the models that have been discussed thus far in this chapter, dendrochronologists are typically not concerned with how the average tree grows over time. Rather, they are interested in what we have been calling year-effects and what they would call the chronology.

Tree cores are used to measure the rings. They are obtained by twisting a hollow borer into the tree until just past the middle of the tree. When the borer is removed, a pencil-like sample of the tree is available. After the core is processed, the rings can be measured. The a^{th} ring width is the measurement from ring $a-1$ to ring a or the growth in radius of the tree in the a^{th} year.

Chronologies are often obtained using a method called regional curve standardization. We summarize the steps in regional curve standardization, as laid out by Briffa and Melvin (2011). First, ring widths from all samples are lined up by age. This can be visualized as plotting age on the x-axis and ring width on the y-axis. Next, average ring width is computed for each age and a smooth curve is fit to the means. Common curves include a negative exponential function or cubic smoothing spline. Then, a tree ring index is obtained by dividing each ring measurement by the fitted value of the smooth curve at that age. The process of obtaining the tree ring index is meant to have averaged out any non-environmental effects. Lastly, the tree ring indices are lined up by year (visually, that is year of formation on the x-axis and tree ring index on the y-axis) and the average tree ring index for each year forms the chronology index. The chronology is just the series of chronology indices over the years for which measurements are available. The chronology is meant to represent the environmental impact on tree growth. So, large chronology indices indicate above average growth for that year and small chronology indices indicate below average

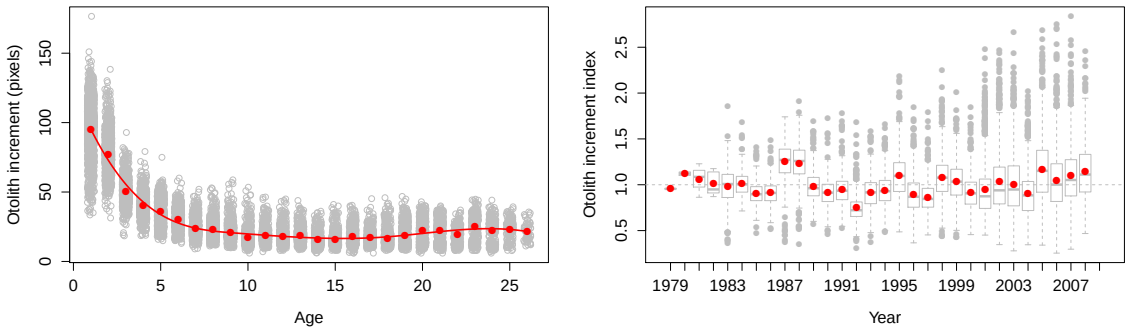
growth for that year. Often dendrochronologists will regress the chronology indices on specific environmental factors, like temperature or rainfall, to try to explain the variability in the chronology.

We applied this methodology to measurements of otolith increments, which are analogous to tree rings, from drum caught in Lake Winnebago, Wisconsin (see Chapter 5 for a detailed description of the dataset). Figure 1.6a shows a plot of otolith increment versus age. The ages were jittered so that more of the points coule be seen. The red dots are the mean otolith increments for that age. The smooth red curve is a cubic spline fit to the mean otolith increments. An otolith increment index, which is analogous to a tree ring index, is obtained by dividing each otolith increment by its fitted value on the cubic spline. Figure 1.6b shows a plot of otolith increment index versus year. The red dots are the chronology indices: the means of the otolith increment indices. We see, for example, that growth was above average in 1987 and 1988 (among other years) and below average in 1992.

Pereira et al. (1995) used a similar methodology to construct a 110-year chronology for freshwater drum obtained from Red Lakes, Minnesota. They modeled the yearly growth increment of the otolith by an exponential decrease to an asymptote. Specifically, the model

$$\log(Y_i) = \theta_0 + \theta_1 \mathrm{e}^{-\theta_2 i}, \tag{1.11}$$

where i is age and Y_i is the growth of the otolith from year $i-1$ to year i, was fit for each of the fish. This is different from the dendrochronology method where a curve was fit to the average ring widths (analogous to otolith increments). After fitting, residuals were computed and averaged over year. They called the averaged residuals growth indices, and they used the growth indices to assess environmental impact on growth. They found a correlation between the growth index and average June through September air temperature.

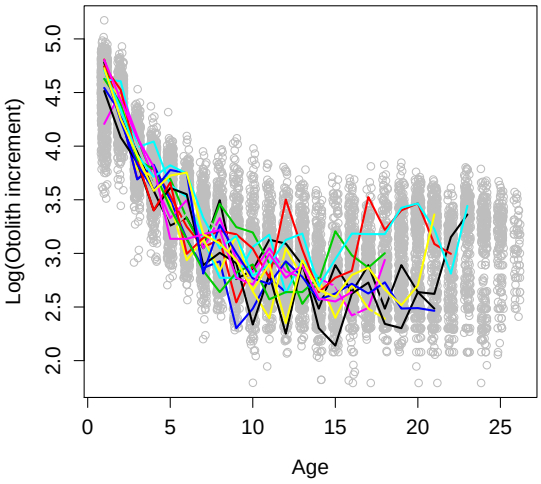


(a) Plot of otolith increment versus age. The red dots are the mean otolith increments for that age. The red curve is a cubic spline fit to the mean otolith increments.

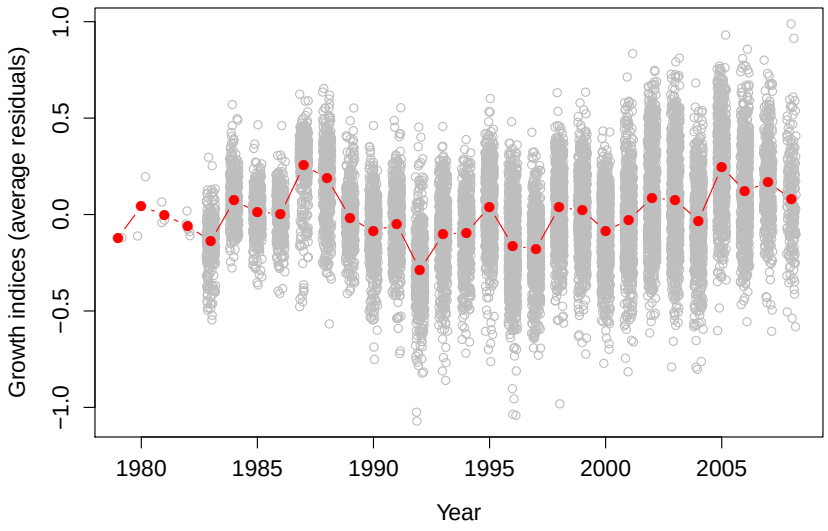
(b) Boxplots of the otolith increment indices versus year of formation of the increment. The red dots are the means of the otolith increment indices for that year, called the chronology indices.

Figure 1.6: Plots of steps taken in regional curve standardization, which is used to estimate a chronology.

The Pereira et al. (1995) method was applied to the Lake Winnebago drum data. Figure 1.7a shows a plot of the logged otolith increments versus age. A sample of fifteen individual growth trajectories from fish with at least ten years of growth are shown as the colored lines on the plot. A curve was fit to each growth trajectory using equation (1.11). Residuals were computed, then plotted versus year. Figure 1.7b shows this plot and displays the average residual by year, the growth indices, as red dots. As with the dendrochronology method, this shows above average growth in years 1987 and 1988 and below average growth in 1992.



(a) Plot of log(otolith increment) versus age. The colored lines are growth trajectories for a sample of 15 fish with at least 10 years of growth.



(b) Plot of residuals or growth indices versus year. The red dots are the average growth indices for that year.

Figure 1.7: Plots illustrating the Pereira et al. (1995) method of estimating a chronology.

1.5 Bayesian hierarchical model

Bayesian methods are becoming more common as computer computations get faster and more efficient (Carlin and Louis, 2009). The incremental model proposed by Weisberg et al. (2010) could be analyzed in the Bayesian framework. We will illustrate how this could be done but leave out the computational detail. The actual model does not change from how it was written in (1.9). We rewrite it here:

$$y_{cka} = \iota_a + h_{c+a-1} + f_{ck} + \varepsilon_{cka}. \tag{1.12}$$

In the Bayesian framework, parameters are not assumed fixed, as they were in the incremental model (and, more generally, in the frequentist framework). Rather, they are assumed to follow what is called a prior distribution. The likelihood, which is the conditional distribution of the data given the parameters, and the priors are used to form the posterior distribution of the parameters given the data. Inference about the parameters is based on the posterior distribution. For example, the mean and standard deviation of the posterior distribution can be used as the estimates and standard deviations of the parameters, respectively.

For the model in (1.12), we will assume $\varepsilon_{cka} \sim N(0, \sigma^2)$ and independent, as we did in the formation of the incremental model. This also implies that

$$y_{cka} \sim N(\iota_a + h_{c+a-1} + f_{ck}, \sigma_2), \tag{1.13}$$

and independent. Next, prior distributions are needed for ι_a, h_{c+a-1} and f_{ck}. Although there are many choices of priors, it seems reasonable to choose priors that match assumptions we made when forming the incremental model. So, we could suppose $h_{c+a-1} \overset{iid}{\sim} N(0, \sigma_h^2)$ and $f_{ck} \overset{iid}{\sim} N(0, \sigma_f^2)$. Since the ι_a parameters represent incremental growth of the scale, it does not make sense for them to be from the same

distribution. Likely, they will be correlated. We can allow this by creating a prior for $\boldsymbol{I}$, the vector of parameters $(\iota_1, \iota_2, \ldots, \iota_t)$. A simple solution would be to let $\boldsymbol{I} \sim N_t(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

If we had previous knowledge about the prior distributions, we could potentially choose values for σ^2, σ_h^2, σ_f^2, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$. Without previous knowledge, we can further specify hyperpriors on these parameters. To make for easier computation, conjugate hyperpriors could be chosen, which are inverse gamma for the variance parameters, multivariate normal for $\boldsymbol{\mu}$, and Wishart for $\boldsymbol{\Sigma}$. Once all the priors are specified, the posterior distributions for the parameters can be found. Computational methods such as Gibb's sampling or MCMC might be needed to do this and are implemented in software such as `WinBUGS` (Lunn et al., 2000).

Chapter 2

The Size-Attained Model

The incremental model (Weisberg et al., 2010) presented in the previous chapter used yearly scale increments of fish, and ignored the more interesting somatic growth. Research has shown that there is not an exact correspondence between scale growth and somatic growth. For example, Casselman (1990, p. 676) found that in northern pike (*Esox lucius*), "as body size increased, scale growth decreased relative to growth of the body". At times, it is not feasible to obtain yearly scale increment measurements. Size attained at time of capture, most commonly quantified by body length, is much easier to measure and reflects somatic growth, which is usually of more interest to the biologists. In this chapter, we propose a model, which we call the *size-attained model*, that is based on the longitudinal approach of the incremental model but uses only size-attained at age of capture as the response variable. As in the previous model, estimates of age effects and year effects are obtained. Age at time of capture is also used in this new model, which could be determined by a hard part of the fish.

2.1 Formulation of the model

To formulate this model, let y_{cka} be the unobserved somatic growth from age $a-1$ to age a of the k^{th} fish in year-class c. If the yearly somatic growth were observed,

then the incremental model,

$$y_{cka} = \iota_a + h_{c+a-1} + f_{ck} + \varepsilon_{cka}, \tag{2.1}$$

could be used to model growth. Instead, only the size-attained at time of capture (size at capture) is observed. Letting p be the age at capture of the k^{th} fish in year-class c, size at capture of fish k in year-class c can be written as

$$y_{ck} = \sum_{a=1}^{p} y_{cka}. \tag{2.2}$$

It follows that

$$y_{ck} = \sum_{a=1}^{p} \iota_a + \sum_{a=1}^{p} h_{c+a-1} + \sum_{a=1}^{p} \varepsilon_{cka}. \tag{2.3}$$

This is the size-attained model. Since there is just one observation per fish, there is no longer a fish effect in the model.

Letting $\sum_{a=1}^{p} \varepsilon_{cka} = \delta_{ck}$, (2.3) can be written as

$$y_{ck} = \mathbf{x}_{ck}^T \boldsymbol{\beta} + \mathbf{z}_{ck}^T \mathbf{u} + \delta_{ck}. \tag{2.4}$$

In this formation of the size-attained model, $\mathbf{x}_{ck}$ is a vector of length t composed of zeros and ones. It has ones in the first to the a^{th} positions, for a fish that was a years old at capture, and zeros the remainder of the vector. The vector $\mathbf{z}_{ck}$ is of length m and is also composed of zeros and ones. It has ones in a consecutive positions, corresponding to the years from when the fish was born to when it was captured, and zeros otherwise. Again, t is the maximum age of a fish in the sample and m is the number of years in the data ((most recent year of capture) − (earliest year-class)). In the previous chapter, when scale increments were used to infer size, the partial scale growth that occurred during the year of capture was not used in the analysis. With

lengths, this is not possible. An easy resolution is to assume that all fish captured after a certain date, which could change depending on the species of fish, have finished growing for the year and relabel them as being a year older, both in age and year of capture. A more complex resolution is discussed in Section 3.5.

Since the ε_{cka} were assumed to be independent random draws from $N(0, \sigma^2)$, the δ_{ck} are independent $N(0, p\sigma^2)$, as

$$\begin{aligned} \text{var}(\delta_{ck}) &= \text{var}(\sum_{a=1}^{p} \varepsilon_{cka}) \\ &= \sum_{a=1}^{p} \text{var}(\varepsilon_{cka}) \\ &= p\sigma^2. \end{aligned}$$

As noted earlier, the distributional assumptions of ε_{cka} could be changed, which would affect δ_{ck} as well. It is also assumed

$$\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_m) \tag{2.5}$$

and $\delta_{ck} \perp \mathbf{u} \; \forall \; c, k$.

The size-attained model can be written in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\delta}. \tag{2.6}$$

Here, $\mathbf{y}$ is a vector of length N containing the size-attained measurements for each fish, $\mathbf{X}$ is an $N \times t$ matrix where $\mathbf{x}_{ck}^T$ are the rows, $\mathbf{Z}$ is an $N \times m$ matrix where $\mathbf{z}_{ck}^T$ are the rows, $\boldsymbol{\beta}$ is a vector of coefficients for the fixed effects, and $\mathbf{u}$ is a vector of coefficients for the random effects. For this model, it is important that fish are sampled from at least two different years or else the age and year effects would be

completely confounded. Using the example from earlier, again suppose the oldest fish in the sample was four, the most recent year of capture was 2007, and the earliest year of birth was 1999. Then the $\mathbf{X}$ matrix would have four columns and the $\mathbf{Z}$ matrix would have eight columns. Each fish has one row of data in each of the $\mathbf{X}$ and $\mathbf{Z}$ matrices. For example, a three year old fish caught in year 2003, would have a row $(1, 1, 1, 0)$ in the $\mathbf{X}$ matrix and a row $(0, 1, 1, 1, 0, 0, 0, 0)$ in the $\mathbf{Z}$ matrix.

The goal is to estimate $\boldsymbol{\beta}$, σ, and σ_u and to predict the random effect vector $\mathbf{u}$. In order to make inferences, standard errors of the estimates and predictions must also be obtained.

2.2 Fixed effects

2.2.1 Estimation

Maximum likelihood is used to find estimators of $\boldsymbol{\beta}$, σ, and σ_u. Let $\mathbf{D}$ be a $N \times N$ diagonal matrix with age at capture of fish k in position (k, k) and everywhere else zeros. Then, $\mathrm{E}(\mathbf{y}) = \mathrm{E}(\mathbf{X}\boldsymbol{\beta}) + \mathrm{E}(\mathbf{Zu}) + \mathrm{E}(\boldsymbol{\delta}) = \mathbf{X}\boldsymbol{\beta}$ and

$$\begin{aligned} \mathrm{var}(\mathbf{y}) &= \mathrm{var}(\mathbf{Zu} + \boldsymbol{\delta}) \\ &= \mathrm{var}(\mathbf{Zu}) + \mathrm{var}(\boldsymbol{\delta}) \\ &= \sigma_u^2 \mathbf{Z}\mathbf{Z}^T + \sigma^2 \mathbf{D}. \end{aligned}$$

Thus

$$\mathbf{y} \ \sim \ N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{D} + \sigma_u^2 \mathbf{Z}\mathbf{Z}^T).$$

Letting $\boldsymbol{\Sigma} = \sigma^2 \mathbf{D} + \sigma_u^2 \mathbf{Z}\mathbf{Z}^T$, the log-likelihood, $\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, can be written as

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{2.7}$$

Maximum likelihood estimators are obtained by equating to zero the first derivative of the log-likelihood function with respect to each of the parameters. The following four identities (Petersen and Pedersen, 2008) are used in computing the derivatives of the log-likelihood function.

1. For matrices $\mathbf{A}$ and $\mathbf{B}$ that are functions of $\mathbf{t}$,

$$\frac{\partial(\mathbf{AB})}{\partial \mathbf{t}} = \Big(\frac{\partial \mathbf{A}}{\partial \mathbf{t}}\Big)\mathbf{B} + \mathbf{A}\Big(\frac{\partial \mathbf{B}}{\partial \mathbf{t}}\Big). \tag{2.8}$$

2. For a matrix $\mathbf{A}$ that is a function of the scalar t,

$$\frac{\partial \mathbf{A}^{-1}}{\partial t} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial t}\mathbf{A}^{-1}. \tag{2.9}$$

3. For a matrix $\mathbf{A}$ that is a function of the scalar t,

$$\frac{\partial \log(|\mathbf{A}|)}{\partial t} = \mathrm{Tr}(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial t}). \tag{2.10}$$

4. For a matrix $\mathbf{A}$ and a vector $\mathbf{t}$,

$$\frac{\partial(\mathbf{t}^T\mathbf{A}\mathbf{t})}{\partial \mathbf{t}} = \frac{\partial \mathbf{t}^T}{\partial \mathbf{t}}\mathbf{A}\mathbf{t} + \frac{\partial(\mathbf{A}\mathbf{t})^T}{\partial \mathbf{t}}\mathbf{t}. \tag{2.11}$$

Differentiating (2.7) with respect to $\boldsymbol{\beta}$ gives

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta}} &= -\frac{1}{2}\left[\frac{\partial(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T}{\partial \boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \frac{\partial(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\beta}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \\
&= \frac{1}{2}\Big[\mathbf{X}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \mathbf{X}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\Big] \\
&= \mathbf{X}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} - \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta}.
\end{aligned} \tag{2.12}$$

Setting this equal to zero gives $\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} = \mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta}$ and thus, if $\boldsymbol{\Sigma}$ were known,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \tag{2.13}$$

would be the maximum likelihood estimator of $\boldsymbol{\beta}$. With $\boldsymbol{\Sigma}$ unknown, the maximum likelihood estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}, \tag{2.14}$$

where $\hat{\boldsymbol{\Sigma}}$ contains the maximum likelihood estimators of σ and σ_u, rather than the parameters σ and σ_u.

Let $\boldsymbol{\theta}^T = (\sigma, \sigma_u)$. Then, taking the first derivative of (2.7) with respect to θ_k (which is either σ or σ_u) gives

$$\frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \theta_k} = -\frac{1}{2}\Big[\mathrm{Tr}\Big(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\Big) - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\Big]. \tag{2.15}$$

In certain situations with balanced data and additional assumptions made on the variance components, this can be solved algebraically (Searle et al., 1992). The case we are dealing with is far too complicated. Instead, substituting the maximum likelihood estimator $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\boldsymbol{\Sigma})$, into $\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ for $\boldsymbol{\beta}$ gives the profile log-likelihood,

$$\ell(\boldsymbol{\beta}(\boldsymbol{\Sigma}), \boldsymbol{\Sigma}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\boldsymbol{\Sigma}))^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\boldsymbol{\Sigma})). \tag{2.16}$$

Then, we just need to find $\boldsymbol{\Sigma}$ that maximizes the profile log-likelihood. There are only two unknowns in $\boldsymbol{\Sigma}$, namely σ and σ_u. With only two unknowns, it is a fairly simple numerical problem. Details are shown in Chapter 3. Once we find the maximum likelihood estimators for σ and σ_u, say $\hat{\sigma}$ and $\hat{\sigma}_u$, we can find $\hat{\boldsymbol{\beta}}$ from the equation $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}$.

2.2.2 Standard errors

The covariance matrix of the maximum likelihood estimators is approximated by the inverse in the expected information (McCulloch et al., 2008; Searle et al., 1992). The square root of the diagonal elements of this matrix gives estimates of the standard errors of $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}$, and $\hat{\sigma_u}$. The expected information is

$$\mathrm{I}(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\mathrm{E}\begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\beta},\boldsymbol{\Sigma})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell(\boldsymbol{\beta},\boldsymbol{\Sigma})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}^T} \\ \\ \frac{\partial^2 \ell(\boldsymbol{\beta},\boldsymbol{\Sigma})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell(\boldsymbol{\beta},\boldsymbol{\Sigma})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \end{bmatrix}. \tag{2.17}$$

Each of the four pieces of $\mathrm{I}(\boldsymbol{\beta}, \boldsymbol{\theta})$ are computed. Taking the transpose of (2.12) and differentiating it with respect to $\boldsymbol{\beta}$ gives

$$\frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X},$$

and

$$-\mathrm{E}(-\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}) = \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}, \tag{2.18}$$

the top left block of the information matrix. Taking the transpose of (2.12) and differentiating it with respect to θ_k gives

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \theta_k \partial \boldsymbol{\beta}^T} &= \mathbf{y}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k} \boldsymbol{\Sigma}^{-1} \mathbf{X} - (\mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k} \boldsymbol{\Sigma}^{-1} \mathbf{X} \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k} \boldsymbol{\Sigma}^{-1} \mathbf{X}, \end{aligned}$$

and the negative expected value is

$$-\mathrm{E}((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\mathbf{X}) = \mathrm{E}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\mathbf{X} \tag{2.19}$$

$$= \mathbf{0}. \tag{2.20}$$

So, the upper right and lower left blocks of the expected information matrix are both $\mathbf{0}$. Lastly, differentiating (2.15) with respect to θ_s gives

$$\frac{\partial^2\ell(\boldsymbol{\beta},\boldsymbol{\Sigma})}{\partial\theta_s\partial\theta_k} = -\frac{1}{2}\Bigg\{\mathrm{Tr}\Big[\frac{\partial}{\partial\theta_s}\Big(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\Big)\Big] - (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\frac{\partial}{\partial\theta_s}\Big(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\Big)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\Bigg\}. \tag{2.21}$$

Note that

$$\begin{aligned}\frac{\partial}{\partial\theta_s}\Big(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\Big) &= \Big(\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\theta_s}\Big)\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k} + \boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_s\partial\theta_k}\\ &= -\boldsymbol{\Sigma}^{-1}\Big(\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_s}\Big)\boldsymbol{\Sigma}^{-1}\Big(\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\Big) + \boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_s\partial\theta_k},\end{aligned} \tag{2.22}$$

and

$$\begin{aligned}\frac{\partial}{\partial\theta_s}\Big(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\Big) &= \Big(\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\theta_s}\Big)\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\frac{\partial}{\partial\theta_s}\Big(\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\Big)\\ &= -\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_s}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Big(\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_s\partial\theta_k}\boldsymbol{\Sigma}^{-1} + \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\theta_s}\Big)\\ &= -\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_s}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Big(\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_s\partial\theta_k}\boldsymbol{\Sigma}^{-1} - \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_s}\boldsymbol{\Sigma}^{-1}\Big)\\ &= -\boldsymbol{\Sigma}^{-1}\Big(\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_s}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k} - \frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_s\partial\theta_k} + \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_s}\Big)\boldsymbol{\Sigma}^{-1}.\end{aligned} \tag{2.23}$$

Putting (2.22) and (2.23) back into (2.21) gives

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \theta_k \partial \theta_s} = &-\frac{1}{2}\Bigg\{ \mathrm{Tr}\Big[-\boldsymbol{\Sigma}^{-1}\Big(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_s}\Big)\boldsymbol{\Sigma}^{-1}\Big(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\Big) + \boldsymbol{\Sigma}^{-1}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_s \partial \theta_k}\Big] \\ &+ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\Big[\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_s}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_s \partial \theta_k}\boldsymbol{\Sigma}^{-1} \\ &+ \boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_s}\boldsymbol{\Sigma}^{-1}\Big](\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\Bigg\} \end{aligned} \tag{2.24}$$

To calculate the negative expected value of (2.24), we rely on the fact that (McCulloch et al., 2008, p.181), if $\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu}$ and $\mathrm{var}(\mathbf{y}) = \mathbf{V}$, then for any matrix $\mathbf{A}$,

$$\mathrm{E}[(\mathbf{y} - \boldsymbol{\mu})^T\mathbf{A}(\mathbf{y} - \boldsymbol{\mu})] = \mathrm{Tr}\{\mathbf{A}\mathrm{E}[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T] = \mathrm{Tr}(\mathbf{AV})\}.$$

Therefore,

$$\begin{aligned}-\mathrm{E}\Big(\frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \theta_k \partial \theta_s}\Big) &= \frac{1}{2}\Bigg\{ \mathrm{Tr}\Big[-\boldsymbol{\Sigma}^{-1}\Big(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_s}\Big)\boldsymbol{\Sigma}^{-1}\Big(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\Big) + \boldsymbol{\Sigma}^{-1}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_s \partial \theta_k}\Big] \\ &\quad + \mathrm{Tr}\Big[\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_s}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^{-1}\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_s \partial \theta_k}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma} \\ &\quad + \boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_s}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\Big]\Bigg\} \\ &= \frac{1}{2}\mathrm{Tr}\Big[\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_s}\Big] \end{aligned} \tag{2.25}$$

is the (k, s) element of the 2×2 lower right block of the expected information matrix. The maximum likelihood estimators $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}$, and $\hat{\sigma}_u$ are substituted in the expected information matrix for $\boldsymbol{\beta}$, σ, and σ_u, giving an approximation to the expected information matrix. Taking the square root of the diagonal elements of the inverse of the approximated expected information matrix provides estimated standard errors of the fixed effects.

2.3 Random effects

2.3.1 Prediction

Another method is used to predict the random effects. In general, for a vector of random variables $\mathbf{w}$, and its predictor $\tilde{\mathbf{w}}$, the mean squared error of prediction is

$$\mathrm{E}[(\tilde{\mathbf{w}} - \mathbf{w})^T \mathbf{A} (\tilde{\mathbf{w}} - \mathbf{w})], \tag{2.26}$$

where $\mathbf{A}$ is a positive definite symmetric matrix. The *best predictor* is obtained by finding the predictor that minimizes (2.26). Searle et al. (1992, p. 262) showed that the best predictor is

$$\tilde{\mathbf{w}} = \mathrm{E}(\mathbf{w}|\mathbf{y}).$$

Following their proof, let $\mathbf{w}_0 = \mathrm{E}(\mathbf{w}|\mathbf{y})$, and we show that $\tilde{\mathbf{w}} = \mathbf{w}_0$.

$$\begin{aligned}
\mathrm{E}(\tilde{\mathbf{w}} - \mathbf{w})^T \mathbf{A} (\tilde{\mathbf{w}} - \mathbf{w}) &= \mathrm{E}(\tilde{\mathbf{w}} - \mathbf{w}_0 + \mathbf{w}_0 - \mathbf{w})^T \mathbf{A} (\tilde{\mathbf{w}} - \mathbf{w}_0 + \mathbf{w}_0 - \mathbf{w}) \\
&= \mathrm{E}(\tilde{\mathbf{w}} - \mathbf{w}_0)^T \mathbf{A} (\tilde{\mathbf{w}} - \mathbf{w}_0) + 2\mathrm{E}(\tilde{\mathbf{w}} - \mathbf{w}_0)^T \mathbf{A} (\mathbf{w}_0 - \mathbf{w}) \\
&\quad + \mathrm{E}(\mathbf{w}_0 - \mathbf{w})^T \mathbf{A} (\mathbf{w}_0 - \mathbf{w}) \\
&= \mathrm{E}(\tilde{\mathbf{w}} - \mathbf{w}_0)^T \mathbf{A} (\tilde{\mathbf{w}} - \mathbf{w}_0) + 2\mathrm{E}_y\{\mathrm{E}_w[(\tilde{\mathbf{w}} - \mathbf{w}_0)^T \mathbf{A} (\mathbf{w}_0 - \mathbf{w})|y]\} \\
&\quad + \mathrm{E}(\mathbf{w}_0 - \mathbf{w})^T \mathbf{A} (\mathbf{w}_0 - \mathbf{w}) \\
&= \mathrm{E}(\tilde{\mathbf{w}} - \mathbf{w}_0)^T \mathbf{A} (\tilde{\mathbf{w}} - \mathbf{w}_0) + 2\mathrm{E}_y\{[\tilde{\mathbf{w}} - \mathbf{w}_0]^T \mathbf{A} \mathrm{E}[(\mathbf{w}_0 - \mathbf{w})|y]\} \\
&\quad + \mathrm{E}(\mathbf{w}_0 - \mathbf{w})^T \mathbf{A} (\mathbf{w}_0 - \mathbf{w}) \\
&= \mathrm{E}(\tilde{\mathbf{w}} - \mathbf{w}_0)^T \mathbf{A} (\tilde{\mathbf{w}} - \mathbf{w}_0) + 0 + \mathrm{E}(\mathbf{w}_0 - \mathbf{w})^T \mathbf{A} (\mathbf{w}_0 - \mathbf{w})
\end{aligned}$$

The first term is minimized by choosing $\tilde{\mathbf{w}} = \mathrm{E}(\mathbf{w}|\mathbf{y})$, and the last term is constant for any estimate.

For the size-attained model, to find the best predictor $\mathrm{E}(\mathbf{u}|\mathbf{y})$, the additional assumption is made that

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{X}\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \mathbf{I}_m & \mathbf{C}^T \\ \mathbf{C} & \boldsymbol{\Sigma} \end{bmatrix} \right),$$

where

$$\begin{aligned} \mathbf{C} &= \mathrm{cov}(\mathbf{y}, \mathbf{u}^T) \\ &= \mathrm{cov}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \mathbf{u}^T) \\ &= \mathrm{cov}(\mathbf{X}\boldsymbol{\beta}, \mathbf{u}^T) + \mathrm{cov}(\mathbf{Z}\mathbf{u}, \mathbf{u}^T) + \mathrm{cov}(\boldsymbol{\varepsilon}, \mathbf{u}^T) \\ &= 0 + \mathbf{Z}\mathrm{var}(\mathbf{u}) + 0 = \sigma_u^2 \mathbf{Z}. \end{aligned} \tag{2.27}$$

For a Normal random vector

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right],$$

then

$$\mathbf{X}_1|\mathbf{X}_2 \sim N(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}^{-1}{}_{22}\boldsymbol{\Sigma}_{21}).$$

Thus,

$$\begin{aligned} \mathrm{E}(\mathbf{u}|\mathbf{y}) &= \mathrm{E}(\mathbf{u}) + \mathbf{C}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathrm{E}(\mathbf{y})) \\ &= \sigma_u^2 \mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned} \tag{2.28}$$

is the best predictor of $\mathbf{u}$. Replacing $\boldsymbol{\beta}$ with its conditional maximum likelihood

estimator given $\boldsymbol{\Sigma}$, $\hat{\boldsymbol{\beta}}$ from (2.13), gives the *best linear unbiased predictor* (BLUP) of $\mathbf{u}$,

$$\begin{aligned}\text{BLUP}(\mathbf{u}) \equiv \tilde{\mathbf{u}} &= \sigma_u^2 \mathbf{Z}^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}) \\ &= \sigma_u^2 \mathbf{Z}^T \mathbf{Q}\mathbf{y}, \end{aligned} \tag{2.29}$$

where

$$\mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}. \tag{2.30}$$

This equation contains unknown parameters, σ_u and σ. It is common to replace these parameters with their maximum likelihood estimates, giving what McCulloch et al. (2008) call the *estimated best predictor* and what Harville and Jeske (1992) call the *empirical BLUP*,

$$\hat{\mathbf{u}} = \hat{\sigma}_u^2 \mathbf{Z}^T \hat{\mathbf{Q}}\mathbf{y}. \tag{2.31}$$

It is this predictor that we will use to predict the random effect vector $\mathbf{u}$ in the size-attained model.

2.3.2 Standard error of prediction

We are also interested in calculating standard errors for the predicted random effects. Letting u_i be the i^{th} element of $\mathbf{u}$, we will show that

$$\text{SE}(\hat{u}_i - u_i) \approx \left\{ \text{var}(\tilde{u}_i - u_i) + \text{Tr}\Big[\text{var}\Big[\frac{\partial \tilde{u}_i}{\partial \boldsymbol{\theta}}\Big]\ E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\Big] \right\}^{1/2}, \tag{2.32}$$

where $\boldsymbol{\theta}^T = (\sigma, \sigma_u)$.

If $\boldsymbol{\Sigma}$ were known and $\tilde{\mathbf{u}}$ could be used to predict $\mathbf{u}$, we might consider using the square root of the diagonal elements of the matrix $\text{var}(\tilde{\mathbf{u}})$ as the standard errors. But, this does not take into account that the $\mathbf{u}$ are unobservable quantities, that is,

that they are random effects (Littell et al., 2006). What we are actually interested in is the *mean squared error of prediction* (MSE). For an unbiased predictor, where $\mathrm{E}(\tilde{\mathbf{w}}) = \mathrm{E}(\mathbf{w})$,

$$\mathrm{MSE}(\tilde{\mathbf{w}}) = \mathrm{E}[(\tilde{\mathbf{w}} - \mathbf{w})(\tilde{\mathbf{w}} - \mathbf{w})^T] = \mathrm{var}(\tilde{\mathbf{w}} - \mathbf{w}) = \mathrm{var}(\tilde{\mathbf{w}}) + \mathrm{var}(\mathbf{w}) - 2\mathrm{cov}(\tilde{\mathbf{w}}, \mathbf{w}).$$

Taking the positive square root of the diagonal elements of the MSE matrix gives the *standard errors of prediction* (Kackar and Harville, 1984). If $\mathbf{w}$ were fixed, this would just reduce to $\mathrm{MSE}(\tilde{\mathbf{w}}) = \mathrm{var}(\tilde{\mathbf{w}})$.

Since $\tilde{\mathbf{u}}$ is an unbiased predictor of $\mathbf{u}$, as

$$\begin{aligned}
\mathrm{E}(\tilde{\mathbf{u}}) &= \mathrm{E}[\sigma_u^2 \mathbf{Z}^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y})] \\
&= \sigma_u^2 \mathbf{Z}^T \boldsymbol{\Sigma}^{-1}[\mathrm{E}(\mathbf{y}) - \mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathrm{E}(\mathbf{y})] \\
&= \sigma_u^2 \mathbf{Z}^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= \sigma_u^2 \mathbf{Z}^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0} = \mathrm{E}(\mathbf{u}),
\end{aligned}$$

then the MSE of $\tilde{\mathbf{u}}$ is computed as follows

$$\begin{aligned}
\mathrm{MSE}(\tilde{\mathbf{u}}) &= \mathrm{var}(\tilde{\mathbf{u}}) + \mathrm{var}(\mathbf{u}) - 2\mathrm{cov}(\tilde{\mathbf{u}}, \mathbf{u}) \\
&= \mathrm{var}[\sigma_u^2 \mathbf{Z}^T \mathbf{Q}\mathbf{y}] + \sigma_u^2 \mathbf{I}_m - 2\mathrm{cov}(\sigma_u^2 \mathbf{Z}^T \mathbf{Q}\mathbf{y}, \mathbf{u}) \\
&= \sigma_u^4 \mathbf{Z}^T \mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}\mathbf{Z} + \sigma_u^2 \mathbf{I}_m - 2\mathrm{cov}(\sigma_u^2 \mathbf{Z}^T \mathbf{Q}\mathbf{Z}\mathbf{u}, \mathbf{u}) \\
&= \sigma_u^4 \mathbf{Z}^T \mathbf{Q}\mathbf{Z} + \sigma_u^2 \mathbf{I}_m - 2\sigma_u^2 \mathbf{Z}^T \mathbf{Q}\mathbf{Z}\mathrm{var}(\mathbf{u}) \\
&= \sigma_u^2 \mathbf{I}_m - \sigma_u^4 \mathbf{Z}^T \mathbf{Q}\mathbf{Z}. \qquad (2.33)
\end{aligned}$$

The second to the last step uses the fact that $\mathbf{Q\Sigma Q} = \mathbf{Q}$, which we verify,

$$\begin{aligned}
\mathbf{Q\Sigma Q} &= [\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}]\mathbf{\Sigma}^{-1}[\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}] \\
&= [\mathbf{I}_N - \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T][\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}] \\
&= \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} - \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} \\
&\quad + \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} \\
&= \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} = \mathbf{Q}. \qquad (2.34)
\end{aligned}$$

Taking the square root of the diagonal elements of (2.33) would provide accurate standard errors of $\mathbf{u}$, if $\mathbf{\Sigma}$ were known. As discussed in section Section 2.3.1, since σ and σ_u are unknown, $\hat{\mathbf{u}}$ must be used to estimate $\mathbf{u}$, rather than $\tilde{\mathbf{u}}$. So, we are actually interested in MSE($\hat{\mathbf{u}} - \mathbf{u}$), which is sometimes estimated by replacing σ and σ_u with their maximum likelihood estimates, $\hat{\sigma}$ and $\hat{\sigma}_u$, in (2.33). This estimate can be exact, but when it isn't, it usually underestimates the variability of $\hat{\mathbf{u}}$ because no account is made for the uncertainty in estimating σ and σ_u (Littell et al., 2006).

Many inflation factors have been proposed to account for the underestimation. Harville and Jeske (1992) discuss inflation factors for $w = \boldsymbol{\gamma}^T\boldsymbol{\beta} + \boldsymbol{\delta}^T\mathbf{u}$, a linear combination of fixed and random effects. We discuss the case when $\boldsymbol{\gamma} = \mathbf{0}$, a vector of zeros of length t, and $\boldsymbol{\delta}_i = (0, ..., 0, 1, 0, ..., 0)^T$, a vector of length m of all zeros except a 1 in the i^{th} position. So, the w we are interested in is just the i^{th} element of the random vector $\mathbf{u}$, therefore we label it u_i.

Analogously, define the i^{th} element of $\tilde{\mathbf{u}}$ as

$$\tilde{u}_i(\boldsymbol{\theta}) = \sigma_u^2\mathbf{Z}_i^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{y}, \qquad (2.35)$$

and the i^{th} diagonal element of MSE($\tilde{\mathbf{u}}$) as

$$m_1(\boldsymbol{\theta}) = \sigma_u^2 - \sigma_u^4 \mathbf{Z}_i^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{Z}_i, \tag{2.36}$$

where $\mathbf{Q}(\boldsymbol{\theta})$ is written for $\mathbf{Q}$ as a reminder that $\mathbf{Q}$ is a function of $\boldsymbol{\theta}^T = (\sigma, \sigma_u)$. So, if $\boldsymbol{\theta}$ were known, then $\tilde{u}_i(\boldsymbol{\theta})$ would be a BLUP of u_i, with MSE $m_1(\boldsymbol{\theta})$. As mentioned in section Section 2.3.1, $\tilde{\mathbf{u}}$ is estimated by $\hat{\mathbf{u}}$. Likewise, $\tilde{u}_i(\boldsymbol{\theta})$ is estimated by $\tilde{u}_i(\hat{\boldsymbol{\theta}})$, the i^{th} element of $\hat{\mathbf{u}}$.

Define $m(\boldsymbol{\theta}) \equiv \text{MSE}[\tilde{u}_i(\hat{\boldsymbol{\theta}})] = \text{E}[(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - u_i)^2]$. Kackar and Harville (1981) showed that $(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - u_i)$ is distributed symmetrically around 0 and that, therefore, $\text{E}(\tilde{u}_i(\hat{\boldsymbol{\theta}})) = \text{E}(u_i)$. Thus, $\tilde{u}_i(\hat{\boldsymbol{\theta}})$ is an unbiased predictor of u_i. They later showed that $m(\boldsymbol{\theta}) \geq m_1(\boldsymbol{\theta})$ and typically $m(\boldsymbol{\theta}) > m_1(\boldsymbol{\theta})$ (Kackar and Harville, 1984). Using common practice and acting as though $m(\boldsymbol{\theta}) = m_1(\boldsymbol{\theta})$ and estimating $m(\boldsymbol{\theta})$ by $m_1(\hat{\boldsymbol{\theta}})$ thus often underestimates the actual variability of $\tilde{u}_i(\hat{\boldsymbol{\theta}})$. The estimate of the standard error of $\tilde{u}_i(\hat{\boldsymbol{\theta}})$ can be improved by working directly with $m(\boldsymbol{\theta})$.

First, write $(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - u_i)$ as $(\tilde{u}_i(\boldsymbol{\theta}) - u_i) + (\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))$. Then,

$$\begin{aligned}
m(\boldsymbol{\theta}) &= \text{E}[(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - u_i)^2] \\
&= \text{E}[(\tilde{u}_i(\boldsymbol{\theta}) - u_i) + (\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))]^2 \\
&= \text{E}[(\tilde{u}_i(\boldsymbol{\theta}) - u_i)]^2 + \text{E}(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))^2 + 2\text{E}[(\tilde{u}_i(\boldsymbol{\theta}) - u_i)(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))] \\
&= m_1(\boldsymbol{\theta}) + \text{E}(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))^2 + 2\text{E}[(\tilde{u}_i(\boldsymbol{\theta}) - u_i)(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))].
\end{aligned}$$

Kackar and Harville (1984) showed that if $\hat{\boldsymbol{\theta}}$ is even, that is $\hat{\boldsymbol{\theta}}(-\mathbf{y}) = \hat{\boldsymbol{\theta}}(\mathbf{y})$ for all $\mathbf{y}$, and translation invariant, that is $\hat{\boldsymbol{\theta}}(\mathbf{y} + \mathbf{X}\boldsymbol{\alpha}) = \hat{\boldsymbol{\theta}}(\mathbf{y})$ for all $\boldsymbol{\alpha}$ and $\mathbf{y}$, then $(\tilde{u}_i(\boldsymbol{\theta}) - u_i)$ and $(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))$ are distributed independently. They also noted that maximum likelihood estimators of $\boldsymbol{\theta}$, like the $\hat{\boldsymbol{\theta}}$ we are using, are even and translation invariant.

Thus, $(\tilde{u}_i(\boldsymbol{\theta}) - u_i)$ and $(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))$ are distributed independently. So,

$$\begin{aligned}\mathrm{E}[(\tilde{u}_i(\boldsymbol{\theta}) - u_i)(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))] &= \mathrm{E}(\tilde{u}_i(\boldsymbol{\theta}) - u_i)\mathrm{E}(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta})) \\ &= 0,\end{aligned}$$

and therefore

$$m(\boldsymbol{\theta}) = m_1(\boldsymbol{\theta}) + \mathrm{E}(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))^2. \tag{2.37}$$

The second term in (2.37) cannot usually be calculated exactly. But, letting $f(\hat{\boldsymbol{\theta}}) = \tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta})$, and doing a Taylor approximation of $f(\hat{\boldsymbol{\theta}})$ in $\hat{\boldsymbol{\theta}}$ about $\boldsymbol{\theta}$ gives

$$\begin{aligned}f(\hat{\boldsymbol{\theta}}) &\approx f(\boldsymbol{\theta}) + \left(\frac{\partial f(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}(\boldsymbol{\theta})\right)^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= \tilde{u}_i(\boldsymbol{\theta}) - \tilde{u}_i(\boldsymbol{\theta}) + \left(\frac{\partial \tilde{u}_i(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}(\boldsymbol{\theta}) - \frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \hat{\boldsymbol{\theta}}}(\boldsymbol{\theta})\right)^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= \left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}).\end{aligned}$$

Then, the second term in (2.37) can be approximated by

$$\mathrm{E}(\tilde{u}_i(\hat{\boldsymbol{\theta}}) - \tilde{u}_i(\boldsymbol{\theta}))^2 \approx \mathrm{E}\left[\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right]^2. \tag{2.38}$$

This approximation is an equality if $\tilde{u}_i(\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$. The right side of (2.38) can

be further simplified by writing

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right]^2 &= \mathrm{E}\left[\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right] \\
&= \mathrm{E}\left\{\mathrm{Tr}\left[\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right]\right\} \\
&= \mathrm{E}\left\{\mathrm{Tr}\left[\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]\right\} \\
&= \mathrm{Tr}\left\{\mathrm{E}\left[\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]\right\}.
\end{aligned} \tag{2.39}$$

To simplify (2.39) further, note that if

$$\mathrm{cov}\left[\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \theta_k}\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \theta_j}, (\hat{\theta}_k-\theta_k)(\hat{\theta}_j-\theta_j)\right] = 0, \tag{2.40}$$

then

$$\begin{aligned}
&\mathrm{Tr}\left\{\mathrm{E}\left[\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]\right\} \\
&\qquad = \mathrm{Tr}\left[\mathrm{E}\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T\right)\mathrm{E}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right].
\end{aligned} \tag{2.41}$$

One example of when (2.40) is true is if the estimator $\hat{\boldsymbol{\theta}}$ and $\tilde{u}_i(\boldsymbol{\theta})$ are obtained from two independently distributed data sets (Kackar and Harville, 1984). If (2.40) is not

true, then (2.41) is treated as an approximation, rather than an equality. That is

$$\begin{aligned}\mathrm{Tr}\left\{\mathrm{E}\left[\left(\frac{\partial\tilde{u}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\tilde{u}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]\right\} \\ \approx \mathrm{Tr}\left[\mathrm{E}\left(\frac{\partial\tilde{u}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\tilde{u}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}^T\right)\mathrm{E}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right].\end{aligned} \tag{2.42}$$

Using this approximation gives

$$\mathrm{E}\left[\left(\frac{\partial\tilde{u}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right]^2 \approx \mathrm{Tr}\left[\mathrm{E}\left(\frac{\partial\tilde{u}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\tilde{u}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}^T\right)\mathrm{E}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]. \tag{2.43}$$

At this point, it is useful to compute $\partial\tilde{u}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Recall from (2.35), and writing out $\mathbf{Q}(\boldsymbol{\theta})$ in detail, that

$$\tilde{u}_i(\boldsymbol{\theta}) = \sigma_u^2\mathbf{Z}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{y} - \sigma_u^2\mathbf{Z}_i^T\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}. \tag{2.44}$$

Letting

$$F = \sigma_u^2\mathbf{Z}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

and

$$G = \sigma_u^2\mathbf{Z}_i^T\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y},$$

then $\tilde{u}_i(\boldsymbol{\theta}) = F - G$. To make for easier reading, $\partial F/\partial\theta_i$ and $\partial G/\partial\theta_i$ are computed separately. The identities (2.8) and (2.9) are used in the calculations.

$$\begin{aligned}\frac{\partial F}{\partial\theta_i} &= \sigma_u^2\mathbf{Z}_i^T\Big(\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\theta_i}\Big)\mathbf{y} + \Big(\frac{\partial}{\partial\theta_i}\sigma_u^2\mathbf{Z}_i^T\Big)\boldsymbol{\Sigma}^{-1}\mathbf{y} \\ &= -\sigma_u^2\mathbf{Z}_i^T\boldsymbol{\Sigma}^{-1}\Big(\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\Big)\boldsymbol{\Sigma}^{-1}\mathbf{y} + \Big(\frac{\partial}{\partial\theta_i}\sigma_u^2\mathbf{Z}_i^T\Big)\boldsymbol{\Sigma}^{-1}\mathbf{y}\end{aligned} \tag{2.45}$$

$$\frac{\partial G}{\partial \theta_i} = \sigma_u^2 \mathbf{Z}_i^T \Big[\frac{\partial}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \Big] \mathbf{y} + \Big(\frac{\partial}{\partial \theta_i} \sigma_u^2 \mathbf{Z}_i^T \Big) \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{y} \tag{2.46}$$

Letting $G_1 = \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1}$, the piece within square brackets in (2.46) can be computed as

$$\begin{aligned}
\frac{\partial G_1}{\partial \theta_i} &= \boldsymbol{\Sigma}^{-1} \mathbf{X} \Big(\frac{\partial}{\partial \theta_i} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \Big) + \Big(\frac{\partial}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{X} \Big) (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \\
&= \boldsymbol{\Sigma}^{-1} \mathbf{X} \Big[(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \theta_i} + \Big(\frac{\partial}{\partial \theta_i} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \Big) \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \Big] \\
&\quad - \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \\
&= \boldsymbol{\Sigma}^{-1} \mathbf{X} \Big[-(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \\
&\qquad - (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \Big(\frac{\partial}{\partial \theta_i} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}) \Big) (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \Big] \\
&\quad - \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \\
&= -\boldsymbol{\Sigma}^{-1} \mathbf{X} \Big[(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \\
&\qquad - (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \Big] \\
&\quad - \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \\
&= -\boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \\
&\quad + \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \\
&\quad - \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \\
&= -\Big[\boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \mathbf{Q} + \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \Big].
\end{aligned} \tag{2.47}$$

Combining (2.45), (2.46), and (2.47) gives

$$\begin{aligned}
\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \theta_i} &= \frac{\partial F}{\partial \theta_i} - \frac{\partial G}{\partial \theta_i} \\
&= \Big[-\sigma_u^2 \mathbf{Z}_i^T \boldsymbol{\Sigma}^{-1}\Big(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\Big)\boldsymbol{\Sigma}^{-1}\mathbf{y} + \Big(\frac{\partial}{\partial \theta_i}\sigma_u^2 \mathbf{Z}_i^T\Big)\boldsymbol{\Sigma}^{-1}\mathbf{y}\Big] \\
&\quad - \Big\{-\sigma_u^2 \mathbf{Z}_i^T \Big[\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\mathbf{Q} \\
&\qquad + \boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\Big]\mathbf{y} \\
&\quad + \Big(\frac{\partial}{\partial \theta_i}\sigma_u^2 \mathbf{Z}_i^T\Big)\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}\Big\} \\
&= \Big(\frac{\partial}{\partial \theta_i}\sigma_u^2 \mathbf{Z}_i^T\Big)\boldsymbol{\Sigma}^{-1}\mathbf{y} - \Big(\frac{\partial}{\partial \theta_i}\sigma_u^2 \mathbf{Z}_i^T\Big)\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \\
&\quad - \sigma_u^2 \mathbf{Z}_i^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\mathbf{y} + \sigma_u^2 \mathbf{Z}_i^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \\
&\quad + \sigma_u^2 \mathbf{Z}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\mathbf{Q}\mathbf{y} \\
&= \Big(\frac{\partial}{\partial \theta_i}\sigma_u^2 \mathbf{Z}_i^T\Big)\mathbf{Q}\mathbf{y} - \sigma_u^2 \mathbf{Z}_i^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\mathbf{Q}\mathbf{y} \\
&\quad + \sigma_u^2 \mathbf{Z}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\mathbf{Q}\mathbf{y} \\
&= \Big(\frac{\partial}{\partial \theta_i}\sigma_u^2 \mathbf{Z}_i^T\Big)\mathbf{Q}\mathbf{y} - \sigma_u^2 \mathbf{Z}_i^T\mathbf{Q}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\mathbf{Q}\mathbf{y} \qquad (2.48)
\end{aligned}$$

Letting

$$\mathbf{c}_i^T = \frac{\partial \sigma_u^2 \mathbf{Z}_i^T}{\partial \theta_i} - \sigma_u^2 \mathbf{Z}_i^T\mathbf{Q}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}, \qquad (2.49)$$

allows (2.48) to be written more concisely as

$$\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \theta_i} = \mathbf{c}_i^T\mathbf{Q}\mathbf{y}. \qquad (2.50)$$

With $\mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2]$, we can write

$$\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{c}^T\mathbf{Q}\mathbf{y}. \qquad (2.51)$$

Returning to (2.43), in order to simplify the right-hand side, note that

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) &= \mathrm{E}(\mathbf{cQy}) \\
&= \mathbf{c}^T\mathbf{Q}\mathrm{E}(\mathbf{y}) \\
&= \mathbf{c}^T\mathbf{QX}\boldsymbol{\beta} \\
&= \mathbf{c}^T(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1})\mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{c}^T(\boldsymbol{\Sigma}^{-1}\mathbf{X} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})\boldsymbol{\beta} \\
&= \mathbf{0}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T\right) &= \mathrm{var}\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) - \mathrm{E}\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\mathrm{E}\left(\frac{\partial \tilde{u}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \qquad (2.52) \\
&= \mathrm{var}(\mathbf{c}^T\mathbf{Qy}) - \mathbf{0} \\
&= \mathbf{c}^T\mathbf{Q}\mathrm{var}(\mathbf{y})(\mathbf{c}^T\mathbf{Q})^T \\
&= \mathbf{c}^T\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^T\mathbf{c} \\
&= \mathbf{c}^T\mathbf{Qc}, \qquad (2.53)
\end{aligned}$$

where the last step uses the facts that $\mathbf{Q} = \mathbf{Q}^T$ and $\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q} = \mathbf{Q}$, which was verified in (2.34).

The second expectation in (2.43), $\mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]$, is equal to the covariance matrix of $\hat{\boldsymbol{\theta}}$ if $\hat{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$ (Kackar and Harville, 1984). So, we use the $\boldsymbol{\theta}$ portion of the inverse information matrix discussed in Section 2.2.2 as an approximation. That is,

$$\mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \approx \left\{ \underset{m}{\frac{1}{2}\mathrm{Tr}}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_s}\right] \right\}^{-1}. \qquad (2.54)$$

The notation $\{_m \quad \}$ is used to represent the right-hand side of (2.54) as a matrix with the (k, s) element shown in brackets.

Thus, a sometimes exact approximation of $m(\boldsymbol{\theta})$ is

$$\tilde{m}(\boldsymbol{\theta}) \approx m_1(\boldsymbol{\theta}) + \mathrm{Tr}\left[\mathbf{c}^T\mathbf{Q}\mathbf{c}\left\{_m \frac{1}{2}\mathrm{Tr}\left[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_s}\right]\right\}^{-1}\right]. \tag{2.55}$$

We further estimate this by replacing $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$.

Chapter 3

Computing

This chapter provides detail about the R (R Development Core Team, 2011) code used to find estimates and predictions for the size-attained model parameters and their associated standard errors. It discusses a solution to finding estimates and predictions when the sample size is large, which typically requires inverting large matrices. Lastly, it describes some options for residual analysis and how to add more variables to the model.

3.1 The data

In order to use the functions discussed in this chapter, three variables are needed: size at capture, age at capture, and year of capture. These are called `lengthvar`, `agevar`, and `yrvar` in the R code. The age and year variables are assumed to be already adjusted for any partial growth (see Section 2.1). In Section 3.5 we discuss other variables, besides age and year, that could be incorporated into the model. Ultimately, `lengthvar`, `agevar`, and `yrvar` are used as arguments in a function called `sizeatt()`, which returns parameter estimates and predictions, standard errors of the estimates and prediction, the variance-covariance matrix of the estimates of the fixed effects, fitted values, residuals, and the value of the log-likelihood evaluated at the parameter

```
createX <- function(agecap){
   maxage <- max(agecap)
   n <- length(agecap)
   ages <- matrix(0, nrow = n, ncol= maxage)
   for (i in 1:n) {
      a <- agecap[i]
      ages[i, 1:a] <- 1
   }
   colnames(ages) <- paste("A", 1:maxage, sep = "")
   ages
}
```

Figure 3.1: R function used to create the model matrix, $\mathbf{X}$.

estimates. Before fitting the model, the input variables must be manipulated. We first discuss how that is done.

The model matrix for the age effects, $\mathbf{X}$, is created by using a function named `createX()`. The R code for this function is shown in Figure 3.1. As discussed at the end of Section 2.1, there is a row of data for each fish, and the number of columns is equal to the maximum age of the fish in the sample. There are ones in the first to the a^{th} columns, for a fish that was a years old at capture, and zeros the remainder of the row.

The model matrix for the random effects, $\mathbf{Z}$ is created in a similar fashion using the function `createZ()`. It takes two arguments `agecap` and `yearcap`, which are the age at capture and year of capture variables. The function returns a matrix with a row for each fish and number of columns equal to the number of years of data, calculated as ((most recent year of capture) − (earliest year-class)). As detailed in Section 2.1, each row of the matrix has ones in a consecutive positions, corresponding to the years from when the fish was born to when it was captured, and zeros otherwise. The code is shown in Figure 3.2.

The first step of the `sizeatt()` function simply creates the $\mathbf{X}$ and $\mathbf{Z}$ matrices,

```
createZ <- function(agecap, yearcap){
   maxage <- max(agecap)
   firstyear <- min(yearcap) - maxage
   lastyear <- max(yearcap) - 1
   nyears <- lastyear - firstyear + 1
   n <- length(yearcap)
   years <- matrix(0, nrow = n, ncol = nyears)
   for (i in 1:n) {
      a <- agecap[i]
      y <- yearcap[i]
      years[i, (y-a):(y-1) - (firstyear) + 1] <- 1
   }
   yobs <- apply(years, 2, function(x) max(x) - min(x) > 0)
   years <- years[, yobs]
   colnames(years) <- paste("GY",
     (firstyear:lastyear) - 1900, sep = "")[yobs]
   years
}
```

Figure 3.2: R function used to create the model matrix, **Z**.

called `X2` and `Z2`, respectively, and defines the `lengthvar` as `Y2` and the `agevar` as `D2`. This is shown in Figure 3.3. The next step optimizes the log-likelihood function and is detailed in the next section.

```
sizeatt <- function(lengthvar, agevar, yrvar){
  D2 <- as.vector(agevar)
  Y2 <- as.vector(lengthvar)
  X2 <- as.matrix(createX(agevar))
  Z2 <- as.matrix(createZ(agevar, yrvar))
  .
  .
  .
}
```

Figure 3.3: Defining vectors and matrices used in the `sizeatt()` function. The `sizeatt()` function is shown in its entirety in Appendix A.

3.2 Estimates, predictions, and standard errors

As discussed in Section 2.2.1, no algebraic solution is available to find maximum likelihood estimators of σ and σ_u. Rather, we must numerically maximize the profile log-likelihood, which substitutes $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\boldsymbol{\Sigma})$ for $\boldsymbol{\beta}$ in the log-likelihood, over σ and σ_u. The profile log-likelihood is

$$\ell(\boldsymbol{\beta}(\boldsymbol{\Sigma}), \boldsymbol{\Sigma}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\boldsymbol{\Sigma}))^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\boldsymbol{\Sigma})). \quad (3.1)$$

Before the optimization is presented, we make some modifications to the profile log-likelihood. First, let

$$\begin{aligned}\boldsymbol{\Sigma} &= \sigma^2\mathbf{D} + \sigma_u^2\mathbf{Z}\mathbf{Z}^T \\ &= \sigma^2\mathbf{D}(\mathbf{I} + \lambda\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^T) \\ &= e^{2\eta}\mathbf{D}(\mathbf{I} + e^{\rho}\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^T), \end{aligned} \quad (3.2)$$

where $e^{\rho} = \lambda = \sigma_u^2/\sigma^2$ and $e^{\eta} = \sigma$. The substitution of e^{ρ} for λ and e^{η} for σ are done so that σ_u and σ are bounded below by zero. The matrix $\mathbf{D}$ is factored out because if it were not, it can lead to $|\boldsymbol{\Sigma}|$ being numerically equivalent to infinity, which results

in the log-likelihood being infinity. Factoring it out gives

$$\begin{aligned}
\log|\boldsymbol{\Sigma}| &= \log|e^{2\eta}\mathbf{D}(\mathbf{I} + e^{\rho}\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^T)| \\
&= \log(|e^{2\eta}\mathbf{D}||\mathbf{I} + e^{\rho}\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^T|) \\
&= \log(e^{2N\eta}|\mathbf{D}||\mathbf{I} + e^{\rho}\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^T|) \\
&= \log\Big[e^{2N\eta}\left(\prod_{i=1}^{N} d_i\right)|\mathbf{I} + e^{\rho}\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^T|\Big] \\
&= 2N\eta + \sum_{i=1}^{N}\log(d_i) + \log|\mathbf{I} + e^{\rho}\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^T| \qquad (3.3)
\end{aligned}$$

Also, $\boldsymbol{\Sigma}^{-1}$ is written as

$$\begin{aligned}
\boldsymbol{\Sigma}^{-1} &= [e^{2\eta}\mathbf{D}(\mathbf{I} + e^{\rho}\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^T)]^{-1} \\
&= e^{-2\eta}(\mathbf{I} + e^{\rho}\mathbf{D}^{-1}\mathbf{Z}\mathbf{Z}^T)^{-1}\mathbf{D}^{-1}. \qquad (3.4)
\end{aligned}$$

Since $\mathbf{D}$ is a diagonal matrix with elements d_i, its inverse is also a diagonal matrix with elements $1/d_i$.

In R, the `optim()` function is used to find the the values of η and ρ that maximize the profile log-likelihood. The arguments in `optim()` that will be used are `par`, the parameters over which the function is optimized; `fn`, the function to be minimized; `method`, the method of optimization used; and `lower` and `upper`, which give the lower and upper bounds for the parameters over which the function is optimized. Since `optim()` performs minimization, rather than maximization, the negative of the profile log-likelihood is the function that will be minimized.

Figure 3.4 shows an R function, `negloglkhd()`, that will be minimized using `optim()`. It returns the value of the negative profile log-likelihood evaluated at `eta` (η) and `rho` (ρ) and takes five arguments: `parms` is a vector of length two containing the values of `eta` and `rho`, `Y` is the response vector, `X` is the model matrix for the

```
negloglkhd <- function(parms, Y, X, D, Z){
  eta <- parms[1]
  rho <- parms[2]
  N <- length(Y)
  Siginv <- (1/exp(2*eta)) * solve(diag(rep(1,N)) +
    exp(rho)* diag(1/D) %*% tcrossprod(Z), diag(1/D))
  B <- ginv(crossprod(X, Siginv %*% X))  %*%
    crossprod(X, Siginv %*% Y)
  nll <- (N/2)*log(2*pi) + (1/2)*(2*N*eta + sum(log(D)) +
    log(det(diag(rep(1,N)) + exp(rho) *
    diag(1/D) %*% tcrossprod(Z)))) +
    (1/2) * crossprod((Y - X %*% B), Siginv %*% (Y - X %*% B))
  nll
}
```

Figure 3.4: The negative profile log-likelihood function that is minimized using the `optim()` function in R.

fixed effects, `D` is a vector of ages, and `Z` is the model matrix for the random effects.

The matrix $\mathbf{\Sigma}^{-1}$ is evaluated in `Siginv` using the formula in (3.4). The coefficient estimate, $\hat{\boldsymbol{\beta}}$, is evaluated in `B` using the function `ginv()` (Venables and Ripley, 2002), which gives the Moore-Penrose generalized inverse. If there were no fish captured from an age group for which an estimated age effect is desired, for example if there were no fish captured at age one, then the model matrix for the fixed effects would no longer have full rank since the columns for age one and age two in the model matrix $\mathbf{X}$ would be exactly the same. So $(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$ would not have a regular inverse. In this case, a generalized inverse can be obtained, and the Moore-Penrose inverse provides a unique solution. When the $\mathbf{X}$ matrix is of full rank, `ginv()` will give the regular inverse.

The `optim()` function in R finds values of `eta` and `rho` that minimize the `negloglkhd()` function. To obtain initial values for `eta` and `rho`, length at capture (`lengthvar`) is regressed on age (`agevar`) and year (`yrvar`), which assumes length at capture is linearly related to age and year(see Figure 3.5). The initial value for `eta` is the log

```
p <- c(log(summary(lm(lengthvar ~ agevar + yrvar))$sigma), log(.25))
opt <- optim(p, negloglkhd, method = "L-BFGS-B", lower=c(-Inf,-Inf),
  upper=c(Inf,10), Y=Y2, X=X2, D=D2, Z=Z2)
```

Figure 3.5: Using the `optim()` function in R to find values of η and ρ that minimize the negative log-likelihood.

of the estimated σ from this regression. The initial value for `rho` is $\log(.25)$, which makes σ_u half the size of σ. These initial values are put into the `par` argument in the `optim()` function.

The method used in the optimization is `L-BFGS-B`, which is a quasi-Newton method that allows for lower and upper bounds for each of the parameters (Nocedal and Wright, 1999). An upper bound is occasionally needed for the `rho` parameter. The parameter ρ is defined as $e^{\rho} = \lambda = \sigma_u^2/\sigma^2$. If σ is close enough to zero and σ_u is large enough, λ is numerically equivalent to infinity, which also makes ρ numerically equivalent to infinity. To prevent this, the maximum value for ρ is set to 10. So, the maximum value for $\lambda = \sigma_u^2/\sigma^2$ is e^{10}, which allows σ_u to be $e^5 \approx 149$ times larger than σ. A situation where the upper bound is utilized would be extremely rare.

Arguments `Y`, `X`, `D`, and `Z` are also entered in the `optim()` function to be passed to the `negloglkhd()` function. These are set to `Y2`, `X2`, `D2`, and `Z2`, respectively, which were defined in Figure 3.3.

Estimates obtained for η and ρ from the `optim()` function, say $\hat{\eta}$ and $\hat{\rho}$, minimize the negative log-likelihood function, which is equivalent to maximizing the log-likelihood. Since σ and σ_u are monotone functions of η and ρ, the values of σ and σ_u that maximize the log-likelihood (with the constraints that $\sigma > 0$, $\sigma_u > 0$, and $\sigma_u/\sigma < e^5$) are

$$\hat{\sigma} = e^{\hat{\eta}} \tag{3.5}$$

and

$$\hat{\sigma}_u = \hat{\sigma}\sqrt{\hat{\lambda}} = \hat{\sigma}e^{\hat{\rho}/2}. \tag{3.6}$$

```
sig <- exp(opt$par[1])
lambda <- exp(opt$par[2])
sigu <- sig * sqrt(lambda)
N <- length(Y2)
Siginv <- (1/sig^2) * solve(diag(rep(1,N)) +
  lambda * diag(1/D2) %*% tcrossprod(Z2), diag(1/D2))
Beta <- ginv(crossprod(X2, Siginv %*% X2)) %*%
  crossprod(X2, Siginv %*% Y2)
Q <- Siginv - Siginv %*% X2 %*%
  ginv(crossprod(X2, Siginv %*% X2)) %*% crossprod(X2, Siginv)
predran <- diag(rep(sigu^2, ncol(Z2))) %*% crossprod(Z2, Q) %*% Y2
```

Figure 3.6: R code used to obtain $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{u}}$.

These are computed as `sig` and `sigu`, respectively, in the R code shown in Figure 3.6.

With these maximum likelihood estimators $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{u}}$, and their standard errors can be computed by translating the formulas derived in Chapter 2 into R code. Figure 3.6 shows this R code for obtaining $\hat{\boldsymbol{\beta}}$ (`Beta`) and $\hat{\mathbf{u}}$ (`predran`). The R code for obtaining the standard errors is detailed in Appendix A. The standard errors of the fixed effects are computed by taking the square root of the diagonal elements of the inverse of the expected information matrix, and the standard errors of the random effects are computed using (2.55). The output of the code was verified by comparing to results obtained from fitting the same model using SAS software ®.

3.3 Modifications for large sample size

There will be problems computing $\boldsymbol{\Sigma}^{-1}$, an $N \times N$ matrix, if the sample size N is large. One way to alleviate this problem is to model the mean size of fish in unique yearclass/age-at-capture combination, which we call a (c, p) class, rather than modeling the size of a single fish in yearclass c. This model is called the *model for*

means. Using the following definitions:

$$
\begin{aligned}
p &= \text{age at capture of fish } k \text{ in yearclass } c \\
r_c &= \text{number of unique } (c,p) \text{ combinations in the sample} \\
n_{cp} &= \text{number of fish in year class } c \text{ captured at age } p,
\end{aligned}
$$

the size-attained model from (2.3) can be written as

$$
y_{ckp} = \sum_{a=1}^{p} \iota_a + \sum_{a=1}^{p} h_{c+a-1} + \delta_{ckp}. \tag{3.7}
$$

The model for means is

$$
\begin{aligned}
\bar{y}_{c\cdot p} &= \frac{1}{n_{cp}} \sum_{k=1}^{n_{cp}} y_{ckp} \\
&= \sum_{a=1}^{p} \iota_a + \sum_{a=1}^{p} h_{c+a-1} + \frac{1}{n_{cp}} \sum_{k=1}^{n_{cp}} \delta_{ckp} \\
&= \sum_{a=1}^{p} \iota_a + \sum_{a=1}^{p} h_{c+a-1} + \bar{\delta}_{c\cdot p},
\end{aligned} \tag{3.8}
$$

where

$$
\begin{aligned}
\operatorname{var}(\bar{\delta}_{c\cdot p}) &= \operatorname{var}\Big(\frac{1}{n_{cp}} \sum_{k=1}^{n_{cp}} \delta_{ckp} \Big) \\
&= \frac{1}{n_{ac}^2} \sum_{k=1}^{n_{cp}} \operatorname{var}(\delta_{ckp}) \\
&= \frac{p\sigma^2}{n_{cp}}.
\end{aligned}
$$

The only difference between this model and the size-attained model is the covariance matrix. So, it could be fit using the same techniques. The only thing that would

need to change in the R code is that an element of `D2`, which was age at capture in the size attained model, is age of fish in the (c, p) class divided by number of fish in the (c, p) class.

When the model for means was run just making the change to `D2`, simulations showed that estimates of σ from the model for means were not similar to estimates of σ from the size-attained model (see Figure 4.12). In an attempt to address these large differences, we proposed another estimator for σ, which we call the *pooled estimator* of σ, or $\hat{\sigma}_p$. Let

$$s^2_{cp} = \begin{cases} \frac{1}{p(n_{cp}-1)} \sum_{k=1}^{n_{cp}} (y_{ckp} - \bar{y}_{c\cdot p})^2, & n_{cp} > 1 \\ 0, & n_{cp} = 1 \end{cases} \tag{3.9}$$

be the within (c, p) class variance in length. Then write

$$\begin{aligned} s^2_{cp} &= \frac{1}{p(n_{cp}-1)} \sum_{k=1}^{n_{cp}} \left[\left(\sum_{a=1}^{p} \iota_a + \sum_{a=1}^{p} h_{c+a-1} + \delta_{ckp} \right) - \left(\sum_{a=1}^{p} \iota_a + \sum_{a=1}^{p} h_{c+a-1} + \bar{\delta}_{c\cdot p} \right) \right]^2 \\ &= \frac{1}{p(n_{cp}-1)} \sum_{k=1}^{n_{cp}} (\delta_{ckp} - \bar{\delta}_{c\cdot p})^2. \end{aligned} \tag{3.10}$$

It can be shown that $\mathrm{E}(s_{cp}^2) = \sigma^2$.

$$\begin{aligned}
\mathrm{E}(s_{cp}^2) &= \frac{1}{p(n_{cp}-1)}\mathrm{E}\left[\sum_{k=1}^{n_{cp}}(\delta_{ckp}-\bar{\delta}_{c\cdot p})^2\right] \\
&= \frac{1}{p(n_{cp}-1)}\mathrm{E}\left(\sum_{k=1}^{n_{cp}}\delta_{ckp}^2 - 2\bar{\delta}_{c\cdot p}\sum_{k=1}^{n_{cp}}\delta_{ckp} + \sum_{k=1}^{n_{cp}}\bar{\delta}_{c\cdot p}^2\right) \\
&= \frac{1}{p(n_{cp}-1)}\mathrm{E}\left(\sum_{k=1}^{n_{cp}}\delta_{ckp}^2 - 2n_{cp}\bar{\delta}_{c\cdot p}^2 + n_{cp}\bar{\delta}_{c\cdot p}^2\right) \\
&= \frac{1}{p(n_{cp}-1)}\left[\sum_{k=1}^{n_{cp}}\mathrm{E}(\delta_{ckp}^2) - n_{cp}\mathrm{E}(\bar{\delta}_{c\cdot p}^2)\right] \\
&= \frac{1}{p(n_{cp}-1)}\left[\sum_{k=1}^{n_{cp}}\mathrm{var}(\delta_{ckp}^2) - n_{cp}\mathrm{var}(\bar{\delta}_{c\cdot p})\right] \\
&= \frac{1}{p(n_{cp}-1)}\left(n_{cp}p\sigma^2 - \frac{n_{cp}p\sigma^2}{n_{cp}}\right) \\
&= \frac{1}{p(n_{cp}-1)}(n_{cp}-1)p\sigma^2 = \sigma^2
\end{aligned}$$

Let

$$\hat{\sigma}_p^2 = \frac{\sum_1^{r_c}(n_{cp}-1)s_{cp}^2}{\sum_1^{r_c}(n_{cp}-1)} \tag{3.11}$$

be the pooled estimator of σ^2. It is an unbiased estimator since

$$\begin{aligned}
\mathrm{E}(\hat{\sigma}_p^2) &= \frac{\sum_1^{r_c}(n_{cp}-1)\mathrm{E}(s_{cp}^2)}{\sum_1^{r_c}(n_{cp}-1)} \\
&= \frac{\sum_1^{r_c}(n_{cp}-1)\sigma^2}{\sum_1^{r_c}(n_{cp}-1)} = \sigma^2.
\end{aligned}$$

Simulations showed that $\hat{\sigma}_p$ was similar to the estimate of σ from the size-attained model (see Figure 4.14). Once the pooled estimator is calculated, it is substituted into the likelihood equation for σ in (3.1), and an estimate of σ_u is obtained by maximizing the likelihood over σ_u. Simulations showed that these estimates of σ_u were similar to the $\hat{\sigma}_u$ obtained from the size-attained model (see Figure 4.15).

```
createAVG <- function(lengthvar, agevar, yrvar){
  avgbyyrage <- tapply(lengthvar, list(agevar, yrvar), mean)
  sdbyyrage <- tapply(lengthvar, list(agevar, yrvar),
    function(x){ifelse(length(x) == 1, 0, sd(x))})
  nbyyrage <- tapply(lengthvar, list(agevar, yrvar), length)
  avg.dat <- na.omit(data.frame(
    yearcap = rep(as.numeric(colnames(avgbyyrage)),
      each=length(row.names(avgbyyrage))),
    agecap = rep(as.numeric(row.names(avgbyyrage)),
      length(colnames(avgbyyrage))),
    y = as.vector(avgbyyrage),
    sds = as.vector(sdbyyrage),
    ns = as.vector(nbyyrage)))
  avg.dat[order(avg.dat$agecap, avg.dat$yearcap),]
}
```

Figure 3.7: R function used to summarize the data at the unique yearclass/age-at-capture level.

In order to run this model in R, we use a new function called `modelmeans()`. Like the `sizeatt()` function, it takes as its arguments `lengthvar`, `agevar`, and `yrvar` and returns parameter estimates and predictions, standard errors of the estimates and prediction, the variance-covariance matrix of the estimates of the fixed effects, fitted values, residuals, and value of the log-likelihood evaluated at the parameter estimates. The first step in the function is to summarize the data at the (c, p) class level. This is done using the `createAVG()` function, which is shown in Figure 3.7.

Average lengths, standard deviation of the lengths, and number of fish for each $(c, p$ class are computed using the `tapply()` function. The standard deviation is the square root of the within (c, p) class variance defined in (3.10). It is set to zero if there is only one fish in the class. A `data.frame` is created with the following variables: the year of capture (`yearcap`), age at capture (`agecap`), average length(`y`), within (c, p) class standard deviation (`sds`), and number of fish (`ns`). There is a row for each (c, p) class that contains at least one fish.

```
modmeans <- function(lengthvar, agevar, yrvar){
# Create averaged data
  mmdat <- createAVG(lengthvar, agevar, yrvar)
# Pooled estimator of sigma (sigma_p)
  sig <- sqrt(sum((mmdat$ns - 1) * mmdat$sds^2 / mmdat$agecap) /
    sum(mmdat$ns - 1))
# Create y vector, X and Z matrices, and diagonal of D matrix
  D2 <- as.vector(mmdat$agecap/mmdat$ns)
  Y2 <- as.vector(mmdat$y)
  X2 <- as.matrix(createX(mmdat$agecap))
  Z2 <- as.matrix(createZ(mmdat$agecap, mmdat$yearcap))
  .
  .
  .
}
```

Figure 3.8: First steps of the `modelmeans()` function where the data are summarized by unique yearclass/age-at-capture, σ_p is computed, and matrices and vectors to be used later in the function are created or defined.

Inside the `modelmeans()` function, this `data.frame` is stored as `mmdat`. Variables from `mmdat` are used in finding $\hat{\sigma}_p$ following (3.11). They are also used in creating $\mathbf{X}$ and $\mathbf{Z}$ and defining the response vector $\mathbf{y}$ and the diagonal elements of $\mathbf{D}$. This R code is shown in Figure 3.8.

The negative log-likelihood need only be minimized over one parameter, ρ, after $\hat{\sigma}_p$ is put into the profile log-likelihood for σ. The `negloglkhdmeans()` function, shown in Figure 3.9, is the function to be minimized in `optim()`. It is similar to `negloglkhd()`, but with only one unknown parameter, ρ, rather than two.

In `optim()`, the initial value of ρ is set at $\log(.25)$, which makes σ_u half the size of σ. The `L-BFGS-B` method of optimization is used again. Arguments `Y`, `X`, `D`, `Z`, and `sigma` are also entered in the `optim()` function to be passed to the `negloglkhdmeans()` function. These are set to `Y2`, `X2`, `D2`, `Z2`, and `sig` respectively, which were defined in Figure 3.8. Once an estimate of ρ is obtained, $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ are estimated using formulas

```
negloglkhdmeans <- function(parm, Y, X, D, Z, sigma){
  rho <- parm
  eta <- log(sigma)
  N <- length(Y)
  Siginv <- (1/exp(2*eta)) * solve(diag(rep(1,N)) +
    exp(rho)* diag(1/D) %*% tcrossprod(Z), diag(1/D))
  B <- ginv(crossprod(X, Siginv %*% X)) %*% crossprod(X, Siginv %*% Y)
  nll <- (N/2)*log(2*pi) + (1/2)*(2*N*eta + sum(log(D)) +
    log(det(diag(rep(1,N)) + exp(rho)*diag(1/D) %*% tcrossprod(Z)))) +
    (1/2) * crossprod((Y - X %*% B), Siginv %*% (Y - X %*% B))
  nll
}
```

Figure 3.9: Function to be minimized in `optim()`.

derived in Section 2.1. R code for doing this is shown in Figure 3.10: $\hat{\boldsymbol{\beta}}$ is called `Beta` and $\hat{\mathbf{u}}$ is called `predran`. The remainder of the code is the same as that used in the `sizeatt()` function and is shown in Appendix A.

```
opt <- optim(log(.25), negloglkhdmeans, method = "L-BFGS-B",
  upper=c(Inf,10), Y=Y2, X=X2, D=D2, Z=Z2, sigma=sig)
lambda <- exp(opt$par[1])
sigu <- sig * sqrt(lambda)
N <- length(Y2)
Siginv <- (1/sig^2) * solve(diag(rep(1,N)) +
  lambda * solve(diag(D2), tcrossprod(Z2)), diag(1/D2))
Beta <- ginv(crossprod(X2, Siginv %*% X2)) %*%
  crossprod(X2, Siginv %*% Y2)
Q <- Siginv - Siginv %*% X2 %*%
  ginv(crossprod(X2, Siginv %*% X2)) %*% crossprod(X2, Siginv)
predran <- diag(rep(sigu^2, ncol(Z2))) %*% crossprod(Z2, Q) %*% Y2
```

Figure 3.10: Function to be minimized in `optim()`.

3.4 Residual analysis

In this section, we discuss graphical methods of residual analysis to evaluate model assumptions. We focus on doing this for the size-attained model, but the same methods could be applied to the model for means. The `sizeatt()` and `modmeans()` both return the fitted fixed effects ($\mathbf{X}\hat{\boldsymbol{\beta}}$) and the predicted random effects ($\mathbf{Z}\hat{\mathbf{u}}$), which are used to form different types of fitted values, and the two types of standardized residuals discussed.

The marginal residuals, $\hat{\boldsymbol{\xi}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, are used to check linearity. The marginal residuals predict the marginal errors $\boldsymbol{\xi} = \mathbf{y} - \mathrm{E}(\mathbf{y}) = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\delta}$. Scatterplots of the marginal residuals versus fitted fixed effects ($\mathbf{X}\hat{\boldsymbol{\beta}}$) or predictor variables should show random spread around zero if the linearity condition holds (Santos Nobre and da Motta Singer, 2007).

Pinheiro and Bates (2000) suggest using standardized conditional residuals to assess model assumptions. A conditional residual is defined as $\hat{\boldsymbol{\delta}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}}$,

and the k^{th} standardized conditional residual is

$$\hat{\boldsymbol{\delta}}_k^s = \frac{\hat{\boldsymbol{\delta}}}{\hat{\sigma}\sqrt{p}}. \tag{3.12}$$

They suggest using a scatterplot of $\hat{\boldsymbol{\delta}}_k^s$ versus fitted values ($\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$) to assess the heteroscedastic fit. Although the model does not assume heteroscedasticity of the errors, since $\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma^2\mathbf{D})$, the standardized conditional residuals should show random spread around zero since they are divided by ($\hat{\sigma}\sqrt{p}$). Q-Q plots of the standardized conditional residuals are used to check for normality of $\boldsymbol{\delta}$.

Santos Nobre and da Motta Singer (2007) proposed a another standardized version of the conditional residuals, which we call *studentized conditional residuals* since they are analogous to the studentized residuals advocated by Cook and Weisberg (1982) for standard linear models. The k^{th} studentized conditional residual is defined as

$$\hat{\boldsymbol{\delta}}_k^* = \frac{\hat{\boldsymbol{\delta}}_k}{\hat{\sigma}\sqrt{\hat{m}_{kk}}}, \tag{3.13}$$

where m_{kk} is the k^{th} diagonal element of $(\sigma^2\mathbf{D})\mathbf{Q}(\sigma^2\mathbf{D})$. These are plotted versus the observation number to identify outlying points. They can also be plotted versus fitted values to assess the heteroscedastic fit and used in a Q-Q plot to check for normality of $\boldsymbol{\delta}$. Santos Nobre and da Motta Singer (2007) warn that $\hat{\boldsymbol{\delta}}$ is confounded with $\mathbf{u}$. So, if $\mathbf{u}$ is non-normal, $\hat{\boldsymbol{\delta}}$ may be non-normal, even if $\boldsymbol{\delta}$ is normal.

To assess assumptions on the random effects, we use two plots. A plot of $\mathbf{Z}\hat{\mathbf{u}}$ versus observation number can show outlying observations (Santos Nobre and da Motta Singer, 2007). Q-Q plots of $\hat{\mathbf{u}}$ are used to assess the normality assumption of the random effects (Pinheiro and Bates, 2000).

In the size-attained model it is assumed $\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma^2\mathbf{D})$. One way to check that the assumed covariance structure fits is by examining the within (c, p) class variances

in size, s^2_{cp}, as defined in (3.10). A scatterplot of ps^2_{cp} on the y-axis and age at capture, p, on the x-axis should show a linear relationship between ps^2_{cp} and p since $\text{var}(\delta_{ckp}) = p\sigma^2$. The line with intercept zero and slope $\hat{\sigma}^2$ can be plotted to see if the data fit to a reasonable approximation.

Examples of using these plots to assess model assumptions are shown in Chapter 5.

3.5 Adding variables to the model

Adding fixed effects to the size-attained model that are at the fish-level or higher can easily be achieved. For example, a sex effect can be added so that the average size-attained is different for males and females. The size-attained model is then written as

$$y_{ck} = \sum_{a=1}^{p} \iota_a + \sum_{a=1}^{p} h_{c+a-1} + \alpha s_{ck} + \delta_{ck}, \tag{3.14}$$

where α is a parameter to be estimated and s_{ck} is a 1 if sex is male and 0 if sex is female. In general, the number of additional parameters to be estimated is equal to the number of levels of the variable minus one; so, one in this case. A model like this, though possible, may not be very interesting. For example, when adding a fixed sex effect a male fish will always have a predicted size-attained $\hat{\alpha}$ greater (or less) than a female fish, regardless of age. This would only be expected if males and females followed parallel growth patterns, meaning the difference in average size-attained for males and females at age one is the same as the difference at all ages for which estimated effects are obtained.

A more complex model could be fit to allow for each sex to have a different estimated age effect. It is written as

$$y_{ck} = \sum_{a=1}^{p} \iota_a + \sum_{a=1}^{p} h_{c+a-1} + \alpha \sum_{a=1}^{p} (\iota s)_{a,ck} + \delta_{ck}. \tag{3.15}$$

The predicted year effects would still be the same for males and females. This model has an additional (number of age effects) $\times$ (number of levels of categorical variable -1) parameters to be estimated, compared to the original size-attained model with only fixed age effects. Other fixed categorical variables could be added, such as gear type used to capture the fish.

Fixed quantitative variables can be added to the model in a similar fashion. A variable that records the number of days since January 1st that the fish was captured would be one useful continuous variable that could be incorporated into the model. So, for example, a fish captured on January 8th would have a value of 7 for this variable. Then, a model that includes this variable interacting with the age effects gives a way to account for fish being captured on different dates. Because an interaction with the age effects is included, the affect of date of capture could be different for different ages. R code for adding categorical and quantitative variables to the size-attained model is currently specialized to each situation. Examples are shown in Chapter 5.

Variables lower than the fish-level, which most often would be at the fish/year level, have difficult and often useless interpretations in the size-attained model. For example, a variable for average temperature for the year could easily be added to and interpreted in the incremental model. It is written as

$$y_{cka} = \iota_a + h_{c+a-1} + f_{ck} + \gamma t_{c+a-1,k} + \varepsilon_{cka}, \tag{3.16}$$

where γ is the parameter to be estimated and $t_{c+a-1,k}$ is the average temperature for fish k in year $c + a - 1$. Then, the size-attained model could be written as

$$y_{ck} = \sum_{a=1}^{p} \iota_a + \sum_{a=1}^{p} h_{c+a-1} + \tau \sum_{a=1}^{p} t_{c+a-1,k} + \delta_{ck}, \tag{3.17}$$

where τ is the parameter to be estimated for the sum of the average temperatures

from age one to age at capture. The temperature variable in this model is diluted because high temperature and low temperature years could be added together.

Chapter 4

Simulation Results

In this chapter, we use simulation to assess how well the size-attained model (2.6) performs. Two data sets were used in the simulations. The West Bearskin Lake data set includes records on 420 smallmouth bass captured in 1988, 1989, and 1990. Their ages at capture range from one to seven years. The Drum data set is much larger, containing records on 1297 drum all captured from Lake Winnebago, Wisconsin. The years of capture range from 2003 to 2009 and ages at capture from one to twenty-six.

4.1 The accuracy of estimation and prediction

Using simulation, we explore how accurately σ, σ_u, and $\boldsymbol{\beta}$ are estimated in the size-attained model (2.6). First, the West Bearskin Lake data were used to form the $\mathbf{X}$ and $\mathbf{Z}$ matrices. Values used for $\boldsymbol{\beta}$ are shown in Table 4.1. For each simulated data set, the random year effects were independent random draws from $N(0, \sigma_u^2)$. The year effects are the same across all fish. The random error for each fish was an independent random draw from $N(0, p\sigma^2)$, where p is the age of the fish at time of capture.

We are concerned with how the ratio of σ_u to σ affects the accuracy of the estimates. We simulated data sets using nine different combinations of σ and σ_u values. These are shown in Table 4.2. For each of the combinations, 500 data sets were

Age	1	2	3	4	5	6	7
Value	65.27	53.55	43.92	36.03	29.56	24.25	19.89

Table 4.1: Values of $\boldsymbol{\beta}$ parameter used in the simulations.

σ	1	2	5	5	10	10	10	10	20
σ_u	5	20	10	20	0	5	10	20	5

Table 4.2: Combinations of σ and σ_u used in the simulations.

simulated and the size-attained model was fit to the simulated data.

4.1.1 σ and σ_u

Histograms of $\hat{\sigma}$ from the simulations are shown in Figure 4.1, and histograms of $\hat{\sigma}_u$ are shown in Figure 4.2. The method appears to produce fairly accurate estimates of σ, as the sample means of the $\hat{\sigma}$ are close to the actual parameter values, although all sample means are slightly less than the actual parameter value in all cases. The distribution of $\hat{\sigma}$ appears approximately Normal for all combinations of σ and σ_u.

The accuracy of $\hat{\sigma}_u$ appears to be affected by the ratio of σ to σ_u. In general, these graphs show that when σ_u is less than σ, then σ_u is often underestimated. The way the estimates are obtained (see Chapter 3) guarantees that $\hat{\sigma}_u$ will always be greater than zero, but it appears to often be close to zero in these cases. For example, the distribution of $\hat{\sigma}_u$ when $\sigma = 10$ and $\sigma_u = 5$ shows that many values of $\hat{\sigma}_u$ are close to zero, with over 50% being between zero and one. The sample mean of the estimates is about 1.6. But, when $\sigma = 1$ and $\sigma_u = 5$, fewer estimates are close to zero and the distribution of $\hat{\sigma_u}$ appears fairly Normal. The sample mean of the estimates is about 4.4, which is much closer to the actual parameter value than in the previous case.

In order to see if this variability and pattern in the estimation of σ_u is typical, a

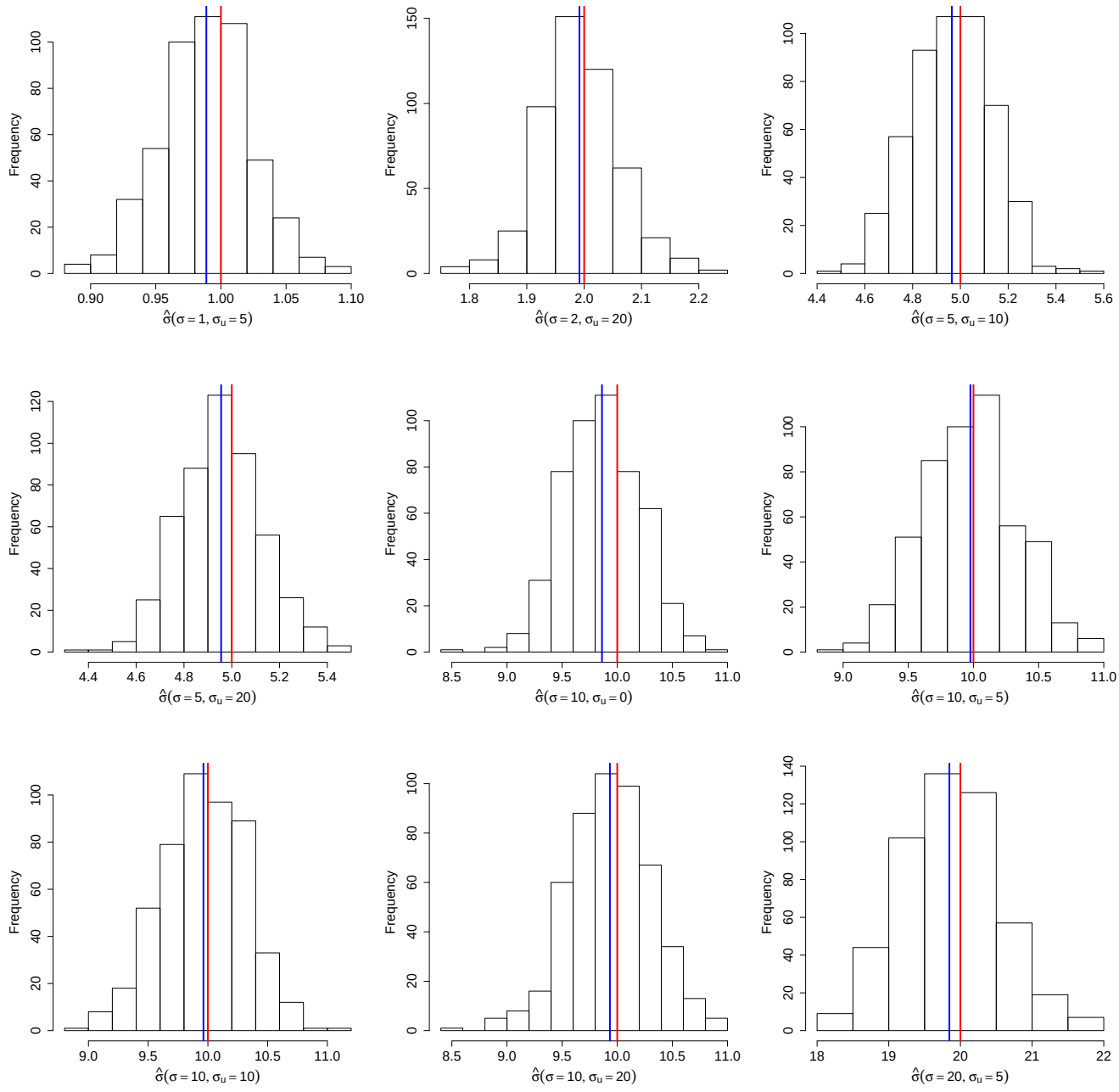


Figure 4.1: Histograms of $\hat{\sigma}$ from 500 simulations. The vertical red line indicates the true parameter value. The vertical blue line is the sample mean of the estimates from the 500 simulations.

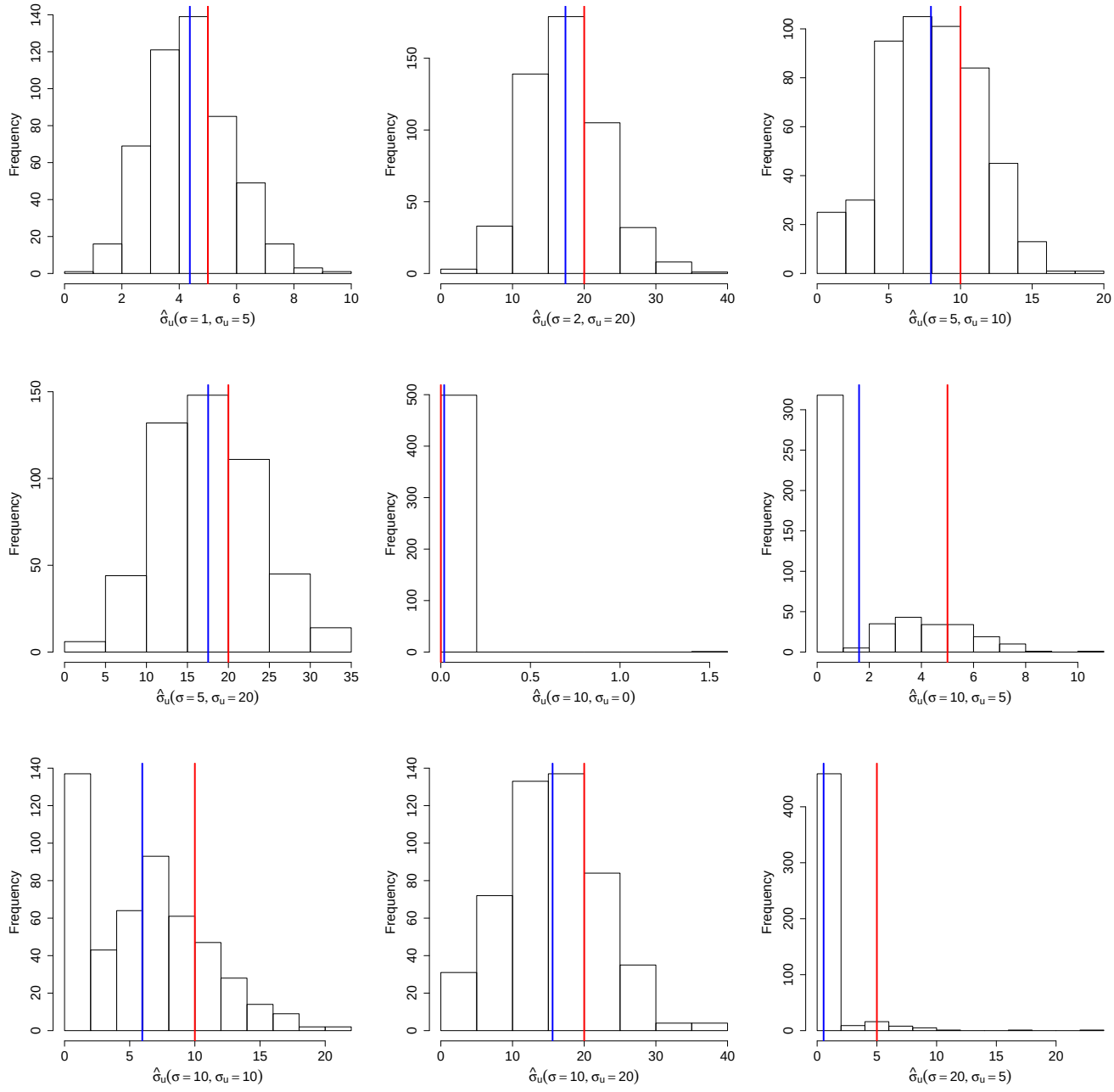


Figure 4.2: Histograms of $\hat{\sigma}_u$ from 500 simulations for different values of σ and σ_u. The vertical red line indicates the true parameter value. The vertical blue line is the sample mean of the estimates from the 500 simulations.

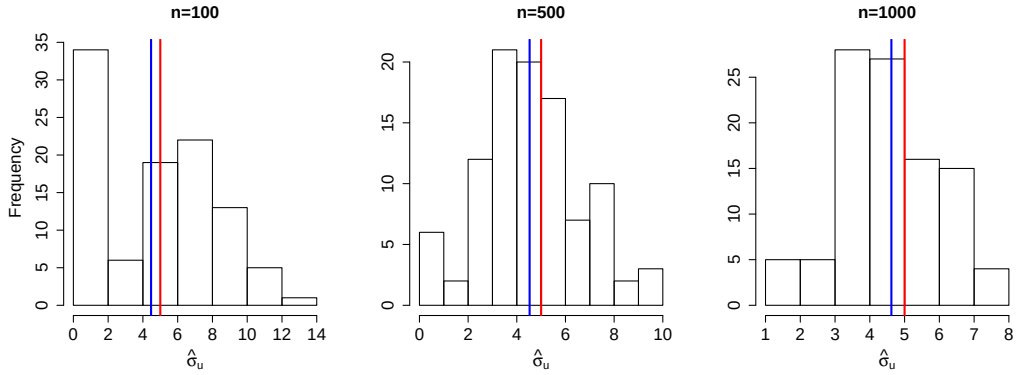


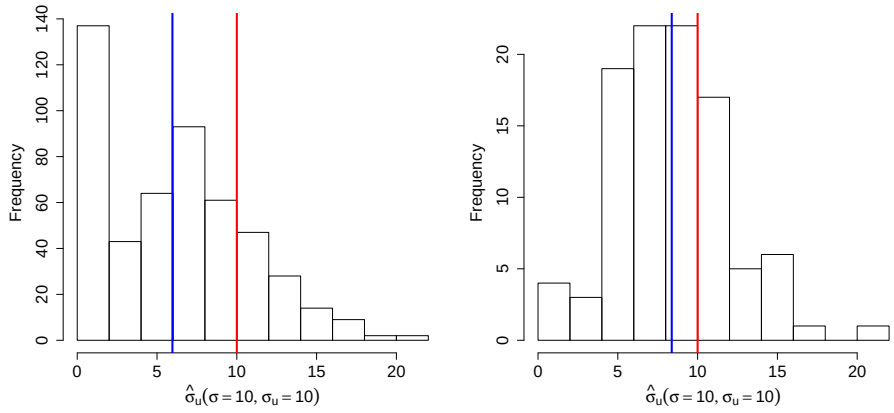
Figure 4.3: Histograms of $\hat{\sigma}_u$ from the simple random effects model (4.1) fit to 100 simulated data sets with sample sizes of $n = 100$, $n = 500$, and $n = 1000$.

simpler one-way random effects model was examined:

$$y_{ij} = \mu + r_i + \varepsilon_{ij}, \tag{4.1}$$

where $\mu = 0$, the r_i were random draws from $N(0, \sigma_u^2)$, the ε_{ij} were random draws from $N(0, \sigma^2)$, and the r_i were independent of the ε_{ij} for all i and j. In this model, we let $\sigma = 20$ and $\sigma_u = 5$ and created 500 simulated data sets for each sample size of 100, 500, and 1000. Histograms of $\hat{\sigma_u}$ are shown in Figure 4.3. We see that $\hat{\sigma_u}$ is close to zero quite often when $n = 100$, but this rarely occurs when $n = 500$ and is even more rare when $n = 1000$. This suggests that estimates of σ_u from the size-attained model might also improve with increased sample size.

The sample size was artificially increased by tripling each record in the West Bearskin Lake data set. This larger data set has 1260 records. One hundred simulated data sets were created using $\sigma = 10$, $\sigma_u = 10$, and the $\boldsymbol{\beta}$ values shown in Table 4.1, and the size-attained model was fit to each data set. In order to assess the effect of sample size, results from these simulations can be compared to the results from the simulations using the smaller data set with $\sigma = 10$ and $\sigma_u = 10$. The distribution of



(a) Small data sets (b) Large data sets

Figure 4.4: Histograms of $\hat{\sigma}_u$ from simulations using two different sample sizes. The vertical red line indicates the true parameter value. The vertical blue line is the sample mean of the estimates from the simulations.

$\hat{\sigma}_u$ from the simulations using the smaller data set and larger data set are shown in Figure 4.4. The estimation of σ_u when the large data sets were used is more accurate than when the small data sets were used. The sample mean of $\hat{\sigma}_u$ from when the small data sets were used is about 5.9 whereas it is about 8.4 when the large data sets were used. Also, over one third of the estimates are between zero and two when the small data sets were used, but less than 5% of the estimates are between zero and two when the large data sets were used.

The standard errors were evaluated by comparing the simulated standard error, which is the sample standard deviation of the estimates from the simulations, and the sample mean of the standard errors of the estimates from the simulations to the theoretical standard error of the estimates. The theoretical standard error of the estimates is computed by putting the actual parameter values in the negative expected information matrix (see Section 2.2.2). For each set of five hundred simulations, these three standard errors were computed for $\hat{\sigma}$ and $\hat{\sigma_u}$. Then, ratios of the simulated

		A		B	
σ	σ_u	$\hat{\sigma}_u$	$\hat{\sigma}$	$\hat{\sigma}_u$	$\hat{\sigma}$
1	5	1.03	0.99	0.88	0.99
2	20	0.96	1.01	0.87	1.00
5	10	1.17	0.96	1.31	0.99
5	20	1.11	0.99	0.93	0.99
10	0	∞	1.00	∞	0.99
10	5	1.25	1.03	24.05	0.99
10	10	1.50	1.00	6.08	0.99
10	20	1.20	1.08	1.47	0.99
20	5	0.85	0.99	45.60	0.99

Table 4.3: Ratios of the simulated standard error to the theoretical standard error (A) and the sample mean of the standard errors of the estimates from the simulations to the theoretical standard error (B) of $\hat{\sigma}_u$ and $\hat{\sigma}$. The simulated standard error of the estimates is computed as the sample standard deviation of the estimates from the simulations. The theoretical standard errors of the estimates are obtained by putting the actual parameter values in the negative expected information matrix and taking the square root of the diagonal elements of the inverse of that matrix.

standard error to the theoretical standard error and the sample mean of the model standard errors to the theoretical standard error were calculated. These are displayed in Table 4.3.

The ratios of the simulated standard error to the theoretical standard error of $\hat{\sigma}$ are close to one for all combinations of σ and σ_u, suggesting that the actual variability in the simulated estimates matches closely with the theoretical variation. Since the ratios of the sample mean of the standard errors from the simulations to the theoretical standard error of $\hat{\sigma}$ are also close to one for all combinations of σ and σ_u, the standard error of $\hat{\sigma}$ seems to be accurate.

For $\hat{\sigma}_u$, the case when $\sigma_u = 0$ leads to both ratios being infinity because the theoretical standard error is zero. Otherwise, the ratios of the simulated standard error to the theoretical standard error range from 0.85 when $\sigma = 20$ and $\sigma_u = 5$ to 1.50 when $\sigma = 10$ and $\sigma_u = 10$. The ratios of the sample mean of the standard

errors from the simulations to the theoretical standard error range from 0.87 when $\sigma = 2$ and $\sigma_u = 20$ to 45.60 when $\sigma = 20$ and $\sigma_u = 5$. The larger ratios suggest that either the standard error of $\hat{\sigma}_u$ is often overestimated, or that using the inverse of the expected information matrix underestimates the standard error of $\hat{\sigma}_u$, or a combination of both.

Overall, we saw that the $\hat{\sigma}$ were only slightly underestimated, regardless of the values of σ and σ_u and the standard errors of $\hat{\sigma}$ were quite accurate. The parameter σ_u was often underestimated when σ_u was smaller than σ, but this improved with increased sample size. Even when σ_u was larger than σ, the sample means of the $\hat{\sigma_u}$ were always less than σ_u. The standard errors of $\hat{\sigma}_u$ were often under or overestimated.

This agrees with previous results. For example Searle et al. (1992) showed that even for the simple random model with unbalanced data, like that in (4.1), the estimates of the variance components are biased, although the exact bias is not easily derived. They also stated that using estimates of the variance components in the calculation of the standard errors "may lead to under-estimation of variances of the estimators" (Searle et al., 1992, p. 240).

4.1.2 $\boldsymbol{\beta}$

Figure 4.5 shows $\hat{\boldsymbol{\beta}}$ from the size-attained model fit to each of the 500 simulated data sets for four of the combinations of σ and σ_u. The true parameter values for $\boldsymbol{\beta}$ are indicated by black dots. Boxplots show the range and give an idea of the variability of $\hat{\boldsymbol{\beta}}$. For all combinations of σ and σ_u, including those run but not shown, the sample means of the estimates from the simulations, indicated by red stars, are close to the true parameter values, suggesting that estimates of $\boldsymbol{\beta}$ from the size-attained model are unbiased.

As was done with the standard errors of $\hat{\sigma}$ and $\hat{\sigma}_u$, the standard errors of $\hat{\boldsymbol{\beta}}$ were evaluated by comparing the simulated standard error of the estimated age effects

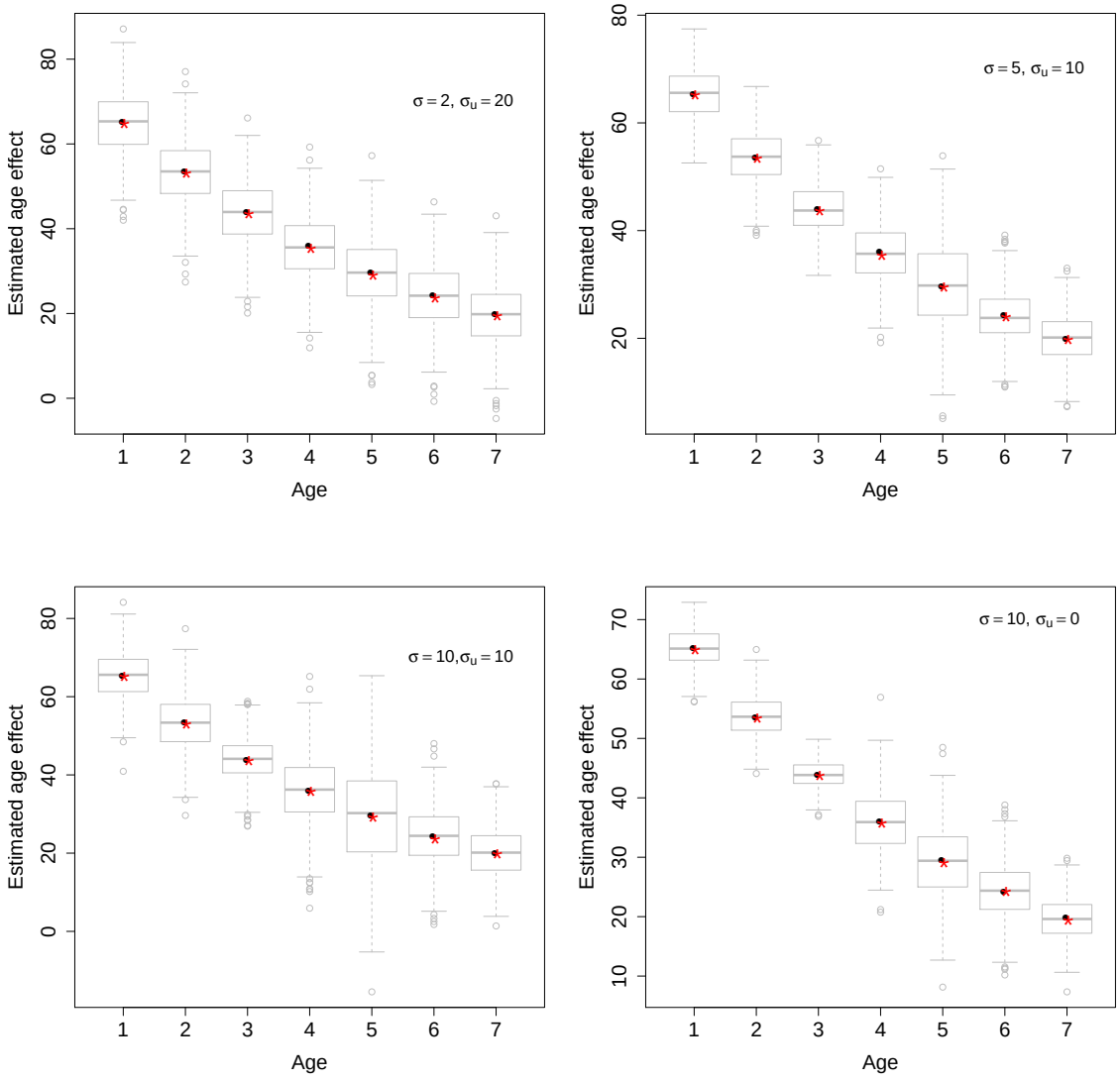


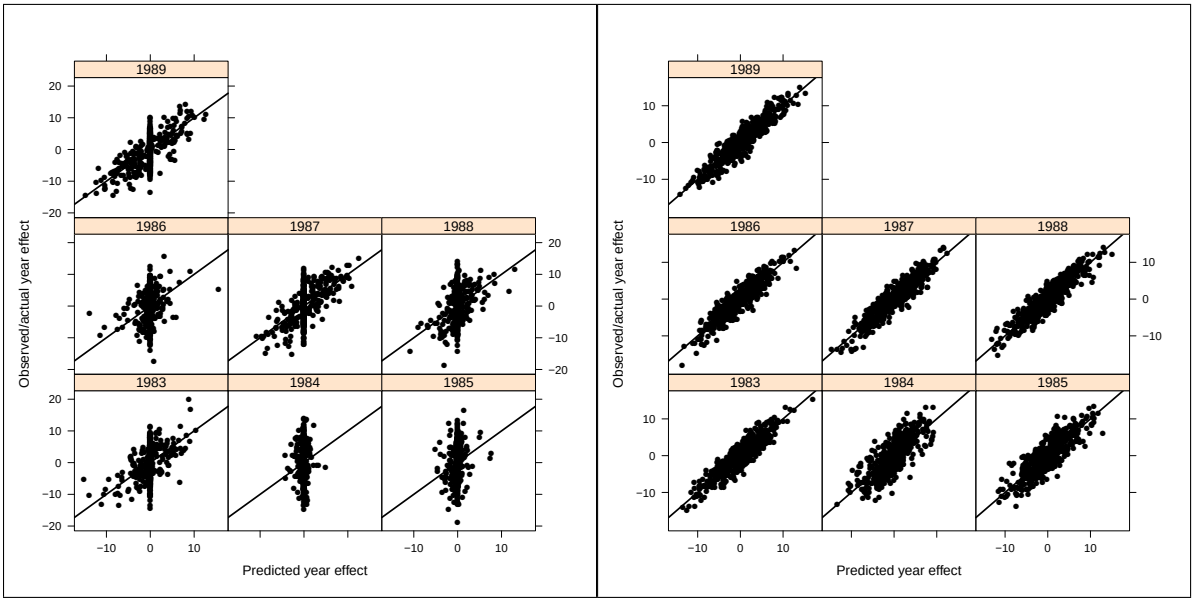
Figure 4.5: $\hat{\boldsymbol{\beta}}$ from 500 simulations. The black dots indicate the true parameter values. The red stars indicate the sample means of the estimated age effects from the 500 simulations.

σ	σ_u	A	B
1	5	0.98	0.89
2	20	1.00	0.88
5	10	1.05	0.87
5	20	0.98	0.90
10	0	0.97	0.99
10	5	1.05	0.83
10	10	1.08	0.82
10	20	1.03	0.86
20	5	1.02	0.90

Table 4.4: Ratios of the simulated standard error to the theoretical standard error of the age effects (A) and the sample mean of the standard errors from the simulations to the theoretical standard error of the age effects(B). The ratios are averaged over all seven of the estimated age effects, so only one ratio is displayed for each σ, σ_u combination.

and the sample mean of the standard errors of the estimated age effects from the simulations to the theoretical standard error of the age effects. The ratios of the simulated standard error to the theoretical standard error and the sample mean of the standard errors from the simulation to the theoretical standard error were calculated. To more concisely display the results, these ratios were averaged over all age effects, so there is only one ratio, even though there are seven age effects. The summarized results are shown in Table 4.4.

Ratios of the simulated standard error to the theoretical standard error are close to one for all combinations of σ and σ_u, suggesting the actual variability in the estimated age effects is similar to the theoretical variation. The ratios of the sample mean of the standard errors from the simulations to the theoretical standard error are all less than one. This suggests that the standard error of the age effects is being underestimated.



(a) $\sigma = 10$, $\sigma_u = 5$ (b) $\sigma = 1$, $\sigma_u = 5$

Figure 4.6: Observed versus predicted year effects from the 500 simulations.

4.1.3 u

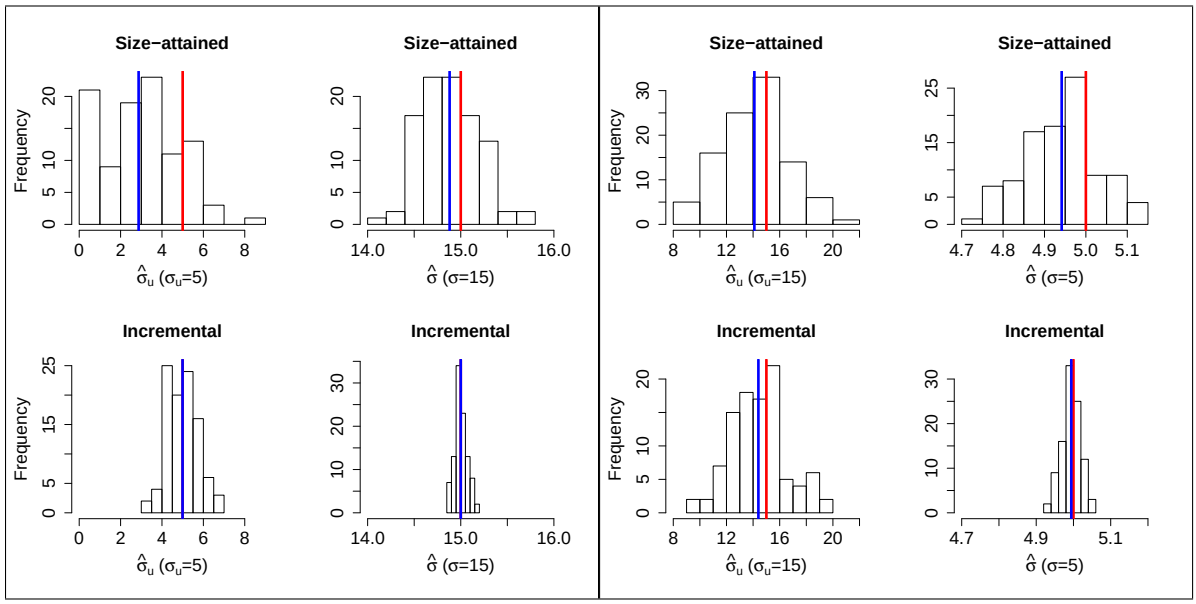
The predicted year effects, $\hat{\mathbf{u}}$, are highly affected by $\hat{\sigma}_u$ (see (2.31)). For example, it was seen in Figure 4.2 that σ_u was often underestimated when $\sigma = 10$ and $\sigma_u = 5$. Figure 4.6 shows scatterplots of the observed year effects, that is the random draw from $N(0, \sigma_u^2)$, versus the predicted year effects for each year from the 500 simulations. When $\sigma = 10$ and $\sigma_u = 5$, the predicted year effects are often relatively close to zero. On the other hand, when $\sigma = 1$ and $\sigma_u = 5$, the scatterplot shows that the observed and predicted year effects agree fairly well, indicated by the points being close to the $y = x$ line. This suggests that year effects may be usefully estimated only when σ_u is large relative to σ.

4.2 Comparing the size-attained and incremental models

The larger Drum data set was used to run simulations to compare estimates from the size-attained model to those from the incremental model. First, the incremental length measurements were simulated, then the corresponding lengths at capture were obtained by summing all increments for each fish. The actual ages and years of capture were used to form the $\mathbf{X}$ and $\mathbf{Z}$ matrices for the incremental model. For simplicity, no random fish effect was included. The $\boldsymbol{\beta}$ are shown as the black dots in Figure 4.8. In one set of simulations, $\sigma = 15$ and $\sigma_u = 5$ and in the other, $\sigma = 5$ and $\sigma_u = 15$. The random year effects were random draws from $N(0, \sigma_u^2)$. A new set of random draws was taken for each simulated data set. The random errors were random draws from $N(0, \sigma^2)$. One hundred simulated data sets were created for the two combinations of σ and σ_u , and both the incremental model (1.10) and size-attained model (2.6) were fit to each data set.

The $\hat{\sigma}$ and $\hat{\sigma_u}$ from the runs of the incremental and size-attained model using simulated data are shown in Figure 4.7. In both cases, there is more variability in the estimates from the size-attained model than in the estimates from the incremental model, but the difference is less extreme in the case when $\sigma = 5$ and $\sigma_u = 15$. The distribution of $\hat{\sigma_u}$ from the size-attained model appears more Normal when $\sigma = 5$ and $\sigma_u = 15$ than when $\sigma = 15$ and $\sigma_u = 5$, and the sample mean of the estimates is relatively close to the true parameter value.

The comparisons of the estimated age effects, $\hat{\boldsymbol{\beta}}$, from the size-attained model to the estimated age effects from the incremental model for the simulated Drum data sets are shown in Figure 4.8. For both combinations of σ and σ_u, the incremental model and size-attained model appear to produce unbiased estimates because the boxplots are centered near the true parameter value.



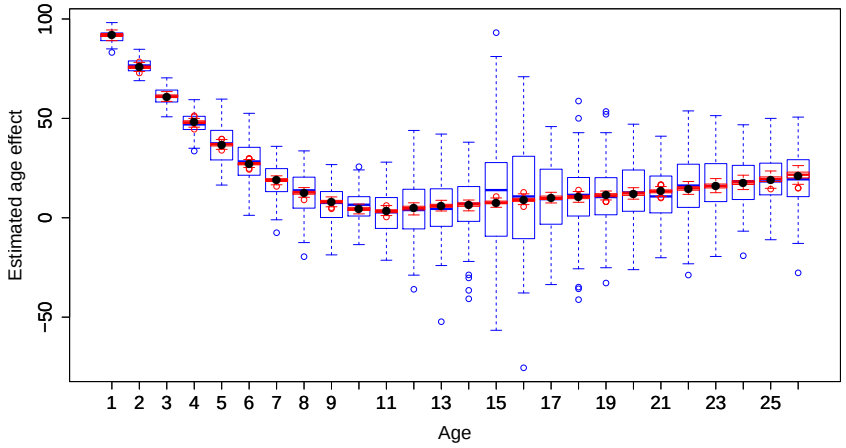
(a) $\sigma = 15$, $\sigma_u = 5$ (b) $\sigma = 5$, $\sigma_u = 15$

Figure 4.7: Estimated σ and σ_u. The vertical red line indicates the true parameter value. The vertical blue line is the sample mean of the estimates from the 100 runs of both the incremental and size-attained models using simulated data.

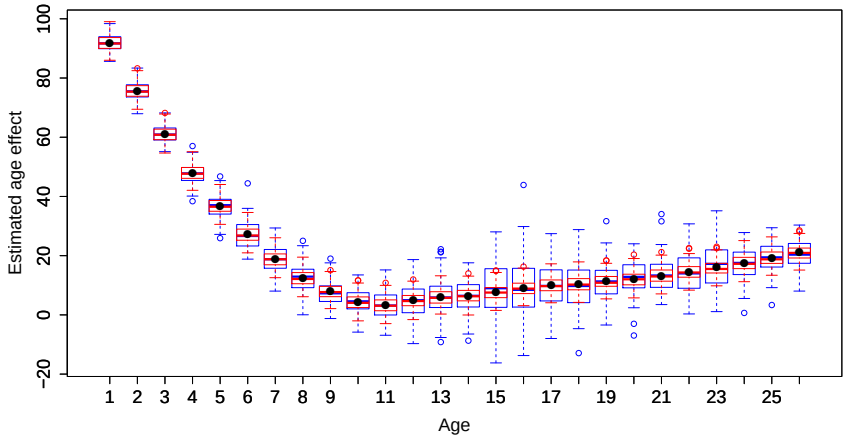
Figure 4.8a shows the estimated age effects from the simulations when $\sigma = 15$ and $\sigma_u = 5$. The boxplots showing the estimated age effects from the incremental model are much narrower than those from the size-attained model, indicating as expected that variability of the estimated age effects is smaller in the incremental model. For the estimates from the size-attained model, the variability is greater at older ages and ages where there were very few fish captured at that age. For example, there were only six drum captured at age fifteen, and the sample standard deviation of the estimated age fifteen effects is almost 28 mm. The sample standard deviations of the estimated age effects from the size-attained model range from 2.7 mm at age one to 28.0 mm at age fifteen. The sample standard deviations of the estimate age effects from the incremental model range from .98 mm at age three to 2.27 mm at age twenty-six.

Figure 4.8b shows the estimated age effects from the simulations when $\sigma = 5$ and $\sigma_u = 15$. In this case, the boxplots showing the estimated age effects from the incremental model are also narrower than those from the size-attained model, but the difference is less than the case when $\sigma = 15$ and $\sigma_u = 5$. For example, for the first four years, the width of the boxplots for the incremental and size-attained estimates is similar and usually the first and third quartile of the estimates from the size-attained model stay within the minimum and maximum of the estimates from the incremental model.

Figure 4.9 shows the estimated age effects for a random sample of four of the one hundred simulations for each of the two σ, σ_u combinations. This further shows that the size-attained model performs well at the younger ages where the estimated age effects from the size-attained model are similar to those from the incremental model. In the case when $\sigma = 5$ and $\sigma_u = 15$ (Figure 4.9b) the estimated age effects are more similar than in the case when $\sigma = 15$ and $\sigma_u = 5$ (Figure 4.9a); the median absolute difference between the estimates from the size-attained and incremental models is

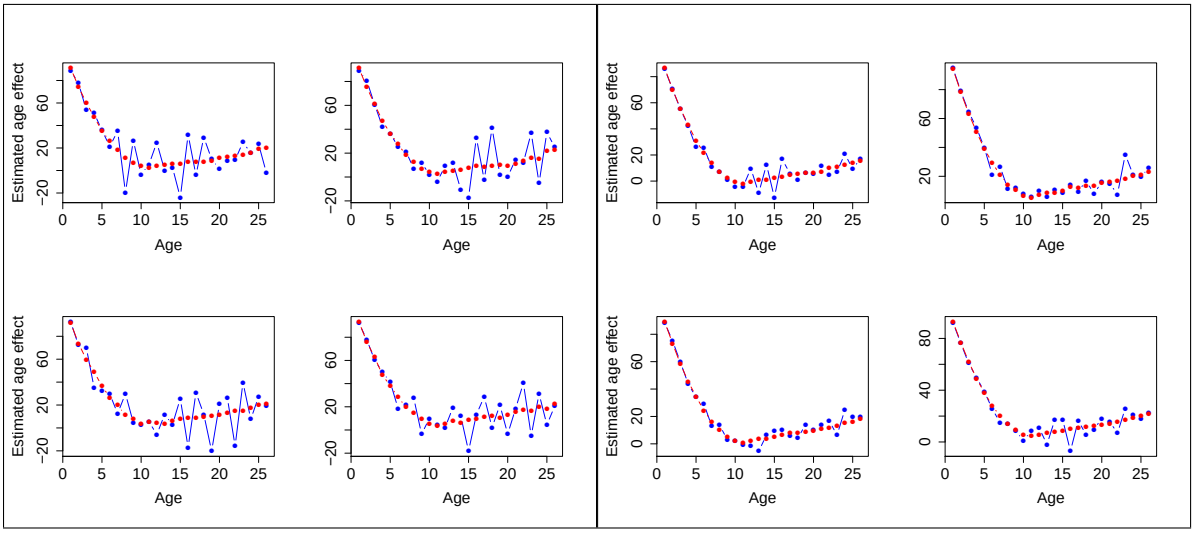


(a) $\sigma = 15$, $\sigma_u = 5$



(b) $\sigma = 5$, $\sigma_u = 15$

Figure 4.8: Estimated age effects from the size-attained and incremental models using simulated Drum data. The black dots indicate the true parameter values. The blue boxplots are the estimated age effects from the size-attained model. The red boxplots are the estimated age effects from the incremental model.



(a) $\sigma = 15$, $\sigma_u = 5$ (b) $\sigma = 5$, $\sigma_u = 15$

Figure 4.9: Estimated age effects from a random sample of four of the one hundred runs of the size-attained and incremental models using simulated Drum data. Blue dots are estimates from the size-attained model; red dots are from the incremental model.

about 2.7 and 7.5, respectively.

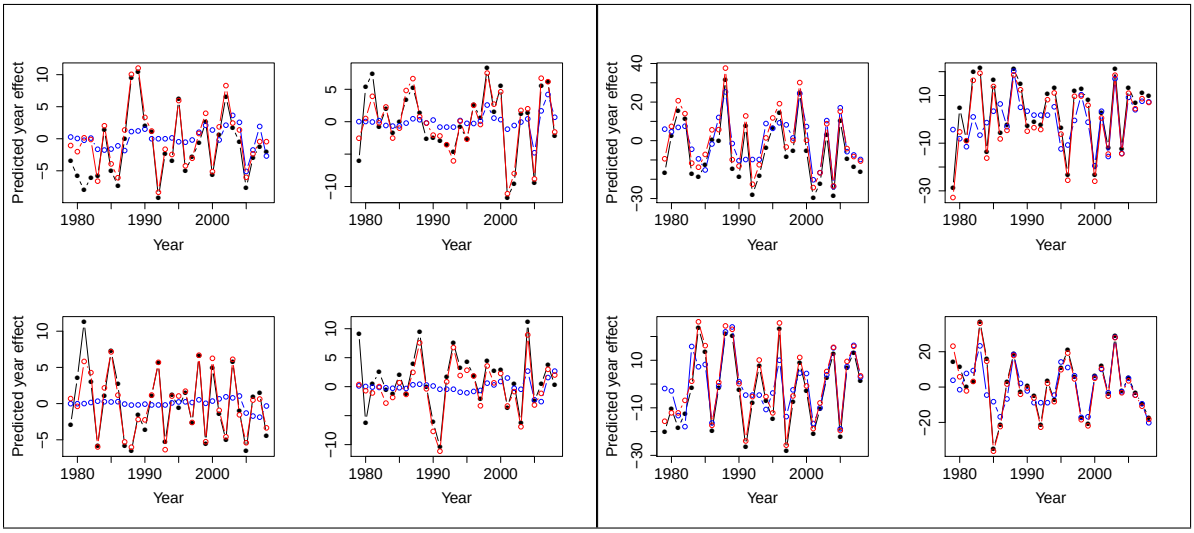
The squared ratio of the standard errors of the estimated age effects from the size-attained model to the standard errors of the estimated age effects from the incremental model gives an idea of how much larger the sample size needs to be when using the size-attained model to yield the same precision as the incremental model. This ratio was computed for each of the estimated age effects, with the standard errors being estimated by the sample standard deviation of the estimates from the one-hundred runs of the model with simulated data. Table 4.5 shows the minimum, median, and maximum ratio for each of the simulation scenarios: $\sigma = 15$, $\sigma_u = 5$ and $\sigma = 5$, $\sigma_u = 15$. When $\sigma = 5$ and $\sigma_u = 15$, a sample size about twelve times larger in the size-attained model would almost always lead to the same precision as in the incremental model, whereas a sample about four times larger would lead to equivalent precision

Parameters	Minimum	Median	Maximum
$\sigma = 15, \sigma_u = 5$	6.46	102.60	734.40
$\sigma = 5, \sigma_u = 15$	1.10	3.44	12.42

Table 4.5: Squared ratio of standard errors of estimated age effects from the size-attained model to standard errors of estimated age effects from the incremental model. The standard errors are estimated by the sample standard deviation of the estimates from the one-hundred runs of the model with simulated data

in about half the estimates. When $\sigma = 15$ and $\sigma_u = 5$, much larger sample sizes are needed in the size-attained model to have equivalent precision to the incremental model. Table 4.5 shows that even when the sample size is about one-hundred times larger, only half the estimates from the size-attained model would have equivalent precision to the incremental model.

Figure 4.10a shows the predicted year effects for a random sample of four of the one hundred simulations with $\sigma = 15$ and $\sigma_u = 5$. The size-attained model does not predict the year effects well. The predicted effects are relatively close to zero compared to the observed values of the random draws from independent $N(0, \sigma_u)$. The few year effects that appear to be predicted well occur in the most recent few years. When $\sigma = 5$ and $\sigma_u = 15$, shown in Figure 4.10b, the size-attained model does much better predicting the year effects. Predictions are especially close to the observed values for the most recent ten years.



(a) $\sigma = 15$, $\sigma_u = 5$ (b) $\sigma = 5$, $\sigma_u = 15$

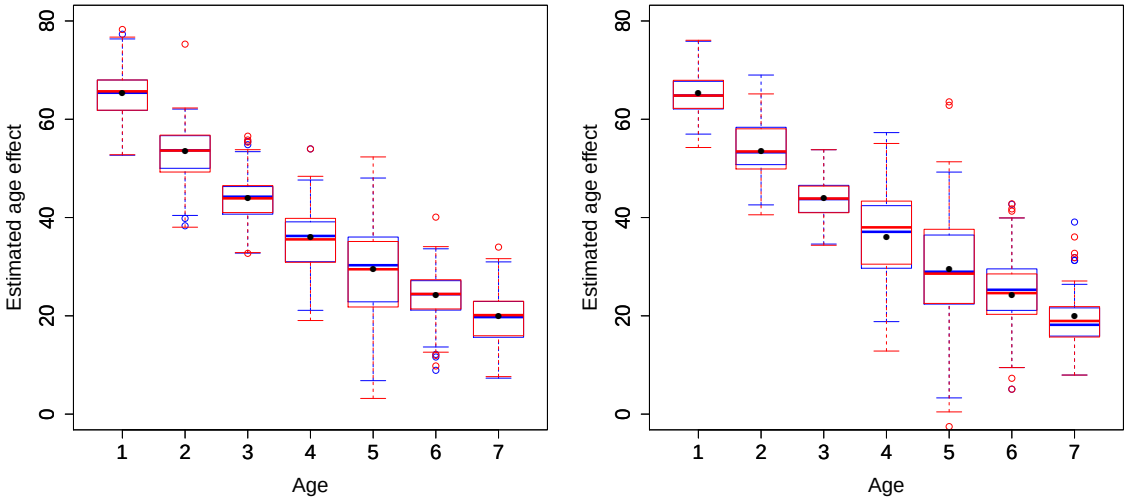
Figure 4.10: Predicted year effects from a random sample of four of the one hundred runs of the size-attained and incremental models using simulated Drum data. Blue dots are predictions from the size-attained model, red dots are from the incremental model, and black dots are the observed values.

4.3 Estimation using model for means

The model for means was proposed in Section 3.3 and is shown in (3.8). It models the mean size of fish in unique (c, p) class, rather than modeling the size of a single fish in yearclass c. We used simulation to explore how the estimates using this method compare to the estimates using the size-attained model. Two sets of 100 simulated data sets based on the West Bearskin Lake data, one with $\sigma = 5$ and $\sigma_u = 10$ and the other with $\sigma = 10$ and $\sigma_u = 5$, were used to make the comparisons. The same $\boldsymbol{\beta}$ parameter that was used in the previous section was used in these simulations and is indicated by the black dots in Figure 4.11. For each data set, the size-attained model was fit to the data, and the mean length of fish in each (c, p) class was calculated and the model for means fit to that data. The estimated age effects from the size-attained model and model for means are shown in Figure 4.11. The pooled estimator of σ, σ_p, is not used as an estimator in these simulations.

The blue boxplots show the distribution of the estimated age effects from the size-attained model. The red boxplots show the distribution of the estimated age effects from the model for means. For both sets of σ and σ_u parameters, the red and blue boxplots at each age are similar, suggesting that the estimated age effect and the variability of the estimated age effect are similar using the size-attained model and model for means. Interestingly, all but two of the sample means of the standard errors of the age effects derived from the size-attained model are larger than for those derived from the model for means (see Table 4.6).

Figure 4.12 shows scatterplots of $\hat{\sigma}$ and $\hat{\sigma}_u$ estimated from the size-attained model versus from the model for means for the one-hundred runs of each of the models with the simulated data. The $y = x$ line is drawn to show where all points would lie if the estimates were the same for the two models for all simulations. The estimates of σ_u from the size-attained model and model for means are similar, especially from



(a) $\sigma = 5$, $\sigma_u = 10$ (b) $\sigma = 10$, $\sigma_u = 5$

Figure 4.11: $\hat{\boldsymbol{\beta}}$ from 100 runs of the size-attained model and model for means using simulated data. The black dots indicate the true parameter values. The blue boxplots are the estimated age effects from the size-attained model. The red boxplots are the estimated age effects from the model for means.

Parameters	Model	A1	A2	A3	A4	A5	A6	A7
$\sigma = 10, \sigma_u = 5$	Size-attained	3.49	4.00	2.72	5.97	7.57	5.09	4.10
	Means	3.12	3.49	2.65	4.81	6.37	4.14	3.50
$\sigma = 5, \sigma_u = 10$	Size-attained	3.79	4.13	3.63	4.93	7.39	4.28	3.93
	Means	3.80	3.95	3.69	4.39	5.72	4.08	3.89

Table 4.6: Sample means of standard errors of the estimated age effects from the 100 runs of the size-attained model and model for means fit to simulated data.

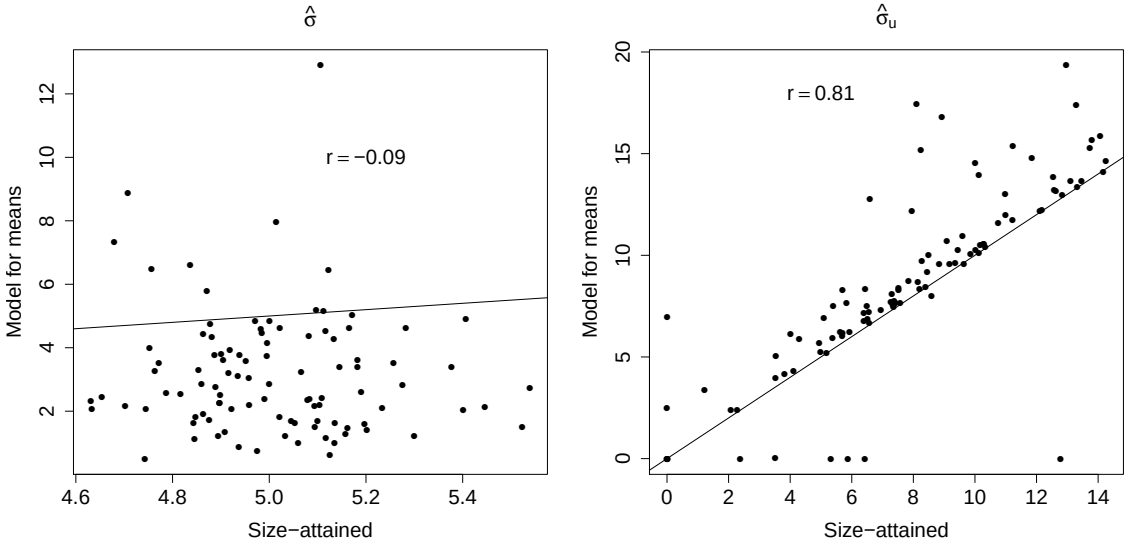
the simulations when $\sigma = 5$ and $\sigma_u = 10$ ($r = 0.81$). Many points lie above the $y = x$ line, indicating that estimates from the model for means are often greater than estimates from the size-attained model. For both σ and σ_u combinations, the range of $\hat{\sigma_u}$ is similar for the size-attained model and model for means.

The estimates of σ from the size-attained model and model for means are not similar, with correlations of $r = -0.09$ and $r = 0.02$ for the simulations when $\sigma = 5$ and $\sigma_u = 10$ and when $\sigma = 10$ and $\sigma_u = 5$, respectively. Many points lie below the $y = x$ line, meaning estimates from the model for means are often less than estimates from the size-attained model. The model for means also produces a wider range of values than the size-attained model. For example, in the simulations where $\sigma = 10$ and $\sigma_u = 5$, $\hat{\sigma}$ from the model for means ranges from close to zero to about thirteen, whereas the estimates from the size-attained model range from about nine to eleven.

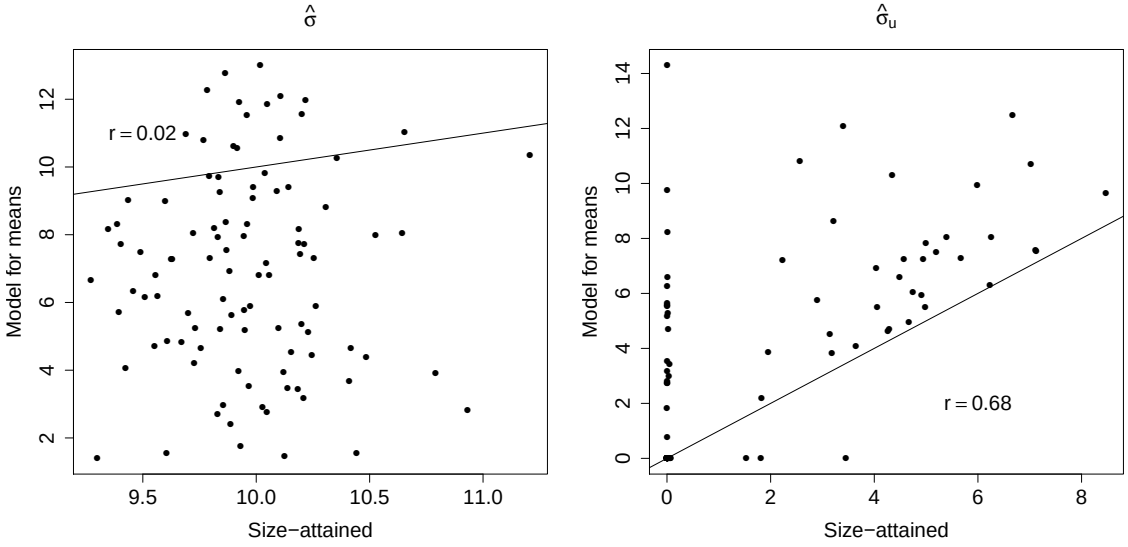
Figure 4.13 shows that the predicted year effects are similar from the size-attained model and model for means as the points lie close to the $y = x$ line. There is a scatterplot for each year effect, with the predictions from the size-attained model on the x-axis and the predictions from the model for means on the y-axis. For the simulations where $\sigma = 5$ and $\sigma_u = 10$ (Figure 4.13a), the predictions are very similar. This is probably due to the high correlation of the estimates of σ_u between the two model methods (see Figure 4.12a).

In Section 3.3 a new estimator was proposed for σ, $\hat{\sigma}_p$. Details of its derivation are shown in (3.11). In the two sets of 100 simulations run using the West Bearskin Lake data, $\hat{\sigma}_p$ was also computed. Figure 4.14 shows scatterplots of $\hat{\sigma}$ from the size-attained model versus the pooled estimator, $\hat{\sigma}_p$. With both correlations very close to one, it is obvious that the pooled estimator is similar to the $\hat{\sigma}$ obtained from the size-attained model. Thus, the pooled estimator, rather than $\hat{\sigma}$ from the model for means, should be used to estimate σ.

Once the pooled estimator is calculated, it can be put into the likelihood equation

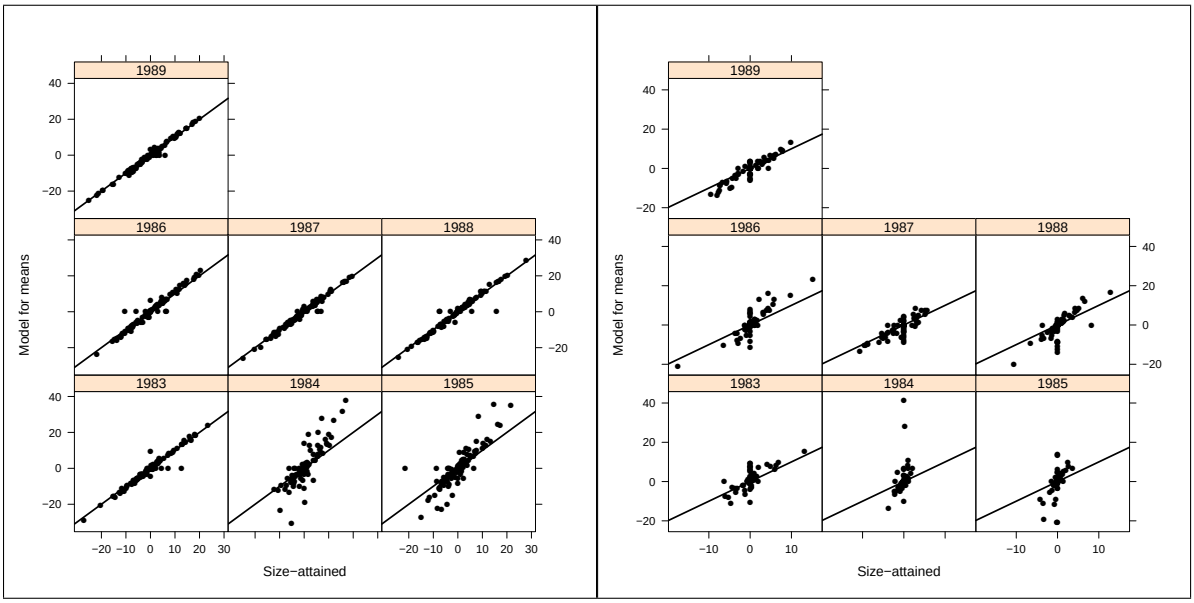


(a) $\sigma = 5$, $\sigma_u = 10$



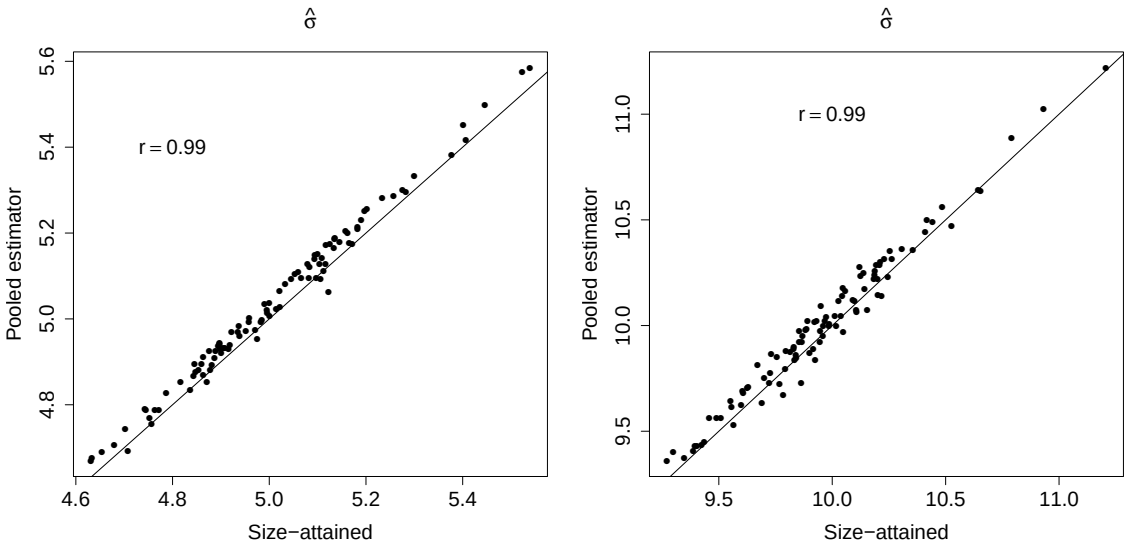
(b) $\sigma = 10$, $\sigma_u = 5$

Figure 4.12: Scatterplots of $\hat{\sigma}$ and $\hat{\sigma}_u$ estimated from the size-attained model versus from the model for means for the 100 simulations.



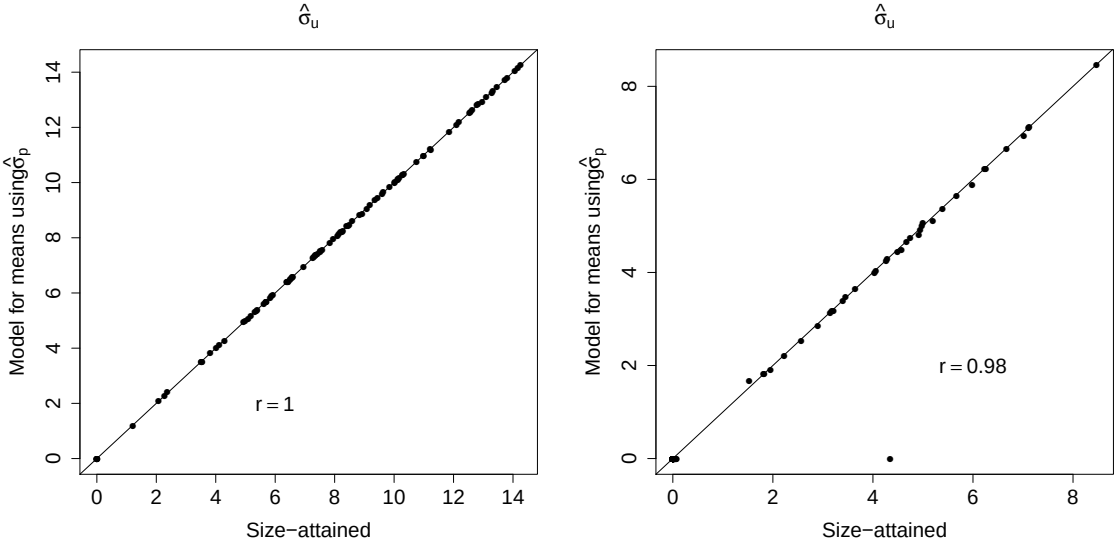
(a) $\sigma = 5$, $\sigma_u = 10$ (b) $\sigma = 10$, $\sigma_u = 5$

Figure 4.13: Scatterplots of predicted year effects from the size-attained model versus predicted year effects from the model for means from the 100 simulations.



(a) $\sigma = 5$, $\sigma_u = 10$ (b) $\sigma = 10$, $\sigma_u = 5$

Figure 4.14: Scatterplots of the estimated σ from the size-attained model versus pooled estimator, $\hat{\sigma}_p$, from 100 simulations.



(a) $\sigma = 5$, $\sigma_u = 10$ (b) $\sigma = 10$, $\sigma_u = 5$

Figure 4.15: Scatterplots of $\hat{\sigma}_u$ estimated from the model for means using $\hat{\sigma}_p$ for σ versus $\hat{\sigma}_u$ from the size-attained model.

for σ in (2.16), and σ_u is estimated by finding the value of σ_u that maximizes the log-likelihood. The R code used to do this is shown in Figure 3.9. This was done for each of the two sets of one hundred simulations.

Figure 4.15 shows scatterplots of $\hat{\sigma}_u$ estimated from the model for means but using $\hat{\sigma}_p$ to estimate σ versus $\hat{\sigma}_u$ estimated from the size-attained model. For both combinations of σ and σ_u, the estimates are almost perfectly correlated. This is a great improvement when compared to the original correlations between the $\hat{\sigma}_u$ from the model for means and size-attained model of $r = 0.81$ when $\sigma = 5$ and $\sigma_u = 10$ and $r = 0.68$ when $\sigma = 10$ and $\sigma_u = 5$.

The sample means of the estimated age effects for the two sets of one hundred simulations were calculated and are displayed in Table 4.7. The sample means of the estimates obtained from the model for means using $\hat{\sigma}_p$ to estimate σ are similar to the

Parameters	Model	A1	A2	A3	A4	A5	A6	A7
$\sigma = 10, \sigma_u = 5$	Size-attained	65.14	54.11	43.64	36.61	29.03	25.18	19.03
	Means	65.18	54.16	43.67	36.57	28.92	25.22	19.03
$\sigma = 5, \sigma_u = 10$	Size-attained	64.93	52.97	43.51	35.30	29.26	23.80	19.54
	Means	64.93	52.97	43.51	35.30	29.27	23.80	19.54

Table 4.7: Sample means of the estimated age effects from the 100 runs of the size-attained model and model for means fit to simulated data.

Parameters	Model	A1	A2	A3	A4	A5	A6	A7
$\sigma = 10, \sigma_u = 5$	Size-attained	3.49	4.00	2.72	5.97	7.57	5.09	4.10
	Means	3.49	4.00	2.72	5.98	7.57	5.10	4.10
$\sigma = 5, \sigma_u = 10$	Size-attained	3.79	4.13	3.63	4.93	7.39	4.28	3.93
	Means	3.79	4.13	3.63	4.94	7.41	4.29	3.94

Table 4.8: Sample means of standard errors of the estimated age effects from the 100 runs of the size-attained model and model for means fit to simulated data.

those obtained from the size-attained model. This was also seen in Figure 4.11, where the medians of the estimated age effects from the model for means were similar to the medians of the estimated age effects from the size-attained model. But, Table 4.8 indicates that using $\hat{\sigma}_p$ in the model for means also leads to estimates of the standard errors of the estimated age effects being similar to those from the size-attained model. This is in contrast to what was seen in Table 4.6, where the standard errors of the estimated age effects from the model for means were often less than those from the size-attained model.

Using $\hat{\sigma}_p$ to estimate σ and then finding the σ_u that maximizes the likelihood equation of the model for means seems to provide similar results to using the size-attained model. This solution should be used when the sample size, N, is too large to obtain a solution using the size-attained model in a reasonable amount of time.

4.4 Sensitivity to assumptions

4.4.1 Correlated year effects

In this section, we investigate how deviations from the model assumptions affect the estimation of parameters. We start by modifying the assumption that $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_m)$. Rather than the year effects being independent, neighboring year effects will be more highly correlated than years that are farther apart. Specifically, let

$$\mathbf{V} = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{m-1} \\ \rho & 1 & \rho & \cdots & \rho^{m-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \rho^{m-3} & \cdots & 1 \end{pmatrix},$$

where $0 \leq \rho < 1$ and assume $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{V})$.

The West Bearskin Lake data was used to create simulated lengths at time of capture. The parameter values used for $\boldsymbol{\beta}$ are shown in Table 4.11, and we let $\sigma = 5$ and $\sigma_u = 10$ or $\sigma = 10$ and $\sigma_u = 5$. One hundred simulated lengths at time of capture were created for each of $\rho = 0, .2, .4, .6,$ and .8 for both σ and σ_u combinations.

The value of ρ does not appear to affect the estimation of σ. As seen in Table 4.9a and Table 4.9b, the sample mean of the estimates (A) is similar to the true parameter value for all values of ρ. Also, the simulated standard error (B), which is the standard deviation of the estimates, is similar to the mean of the standard errors of $\hat{\sigma}$ from the simulations (C) for all values of ρ. These values do not change much for different values of ρ.

Table 4.10a and Table 4.10b show that for all values of ρ, $\hat{\sigma}_u$ underestimates the true value of σ_u, and the bias tends to be larger for larger values of ρ. The simulated standard errors (B) when $\rho \neq 0$ appear similar to the simulated standard error when

ρ	A	B	C
0	4.96	0.17	0.17
0.2	4.98	0.16	0.17
0.4	4.99	0.17	0.17
0.6	4.96	0.17	0.17
0.8	4.98	0.16	0.17

(a) $\sigma = 5$, $\sigma_u = 10$

ρ	A	B	C
0	9.98	0.36	0.35
0.2	9.93	0.32	0.34
0.4	9.96	0.36	0.34
0.6	9.93	0.36	0.34
0.8	9.93	0.36	0.34

(b) $\sigma = 10$, $\sigma_u = 5$

Table 4.9: Sample mean of $\hat{\sigma}$ (A), standard deviation of $\hat{\sigma}$ (B), and sample mean of standard error of $\hat{\sigma}$ (C) from the 100 runs of the size-attained model fit to simulated data for each of the five values of ρ.

ρ	A	B	C
0	7.94	3.43	3.86
0.20	7.56	3.50	4.26
0.40	6.94	3.41	4.90
0.60	6.27	3.09	4.94
0.80	4.59	3.27	8.23

(a) $\sigma = 5$, $\sigma_u = 10$

ρ	A	B	C
0	1.61	2.33	44.88
0.2	1.89	2.59	40.25
0.4	0.94	1.89	49.54
0.6	1.22	1.96	41.14
0.8	0.61	1.37	50.50

(b) $\sigma = 10$, $\sigma_u = 5$

Table 4.10: Sample mean of $\hat{\sigma}_u$ (A), standard deviation of $\hat{\sigma}_u$ (B), and sample mean of standard error of $\hat{\sigma}_u$ (C) from the 100 runs of the size-attained model fit to simulated data for each of the five values of ρ.

$\rho = 0$ when $\sigma = 5$ and $\sigma_u = 10$. When $\sigma = 10$ and $\sigma_u = 5$, the simulated standard errors range from 1.37 when $\rho = 0.8$ to 2.59 when $\rho = 0.2$ For both combinations of σ and σ_u, the sample means of the standard errors from the simulations (C) are larger than the simulated standard errors. This is exaggerated in the case when $\sigma = 10$ and $\sigma_u = 5$.

The sample means of the estimated age effects, $\hat{\boldsymbol{\beta}}$, are shown in Table 4.11 and Table 4.13 and the simulated standard errors of the estimated age effects (A) and sample means of standard errors of the estimated age effects from the simulations (B) in Table 4.12 and Table 4.14. The value of ρ does not appear to affect the

		Age						
		1	2	3	4	5	6	7
Parameter		65.27	53.55	43.92	36.03	29.56	24.25	19.89
ρ	0	65.46	53.63	43.92	35.75	29.88	24.18	20.06
	0.2	65.72	54.48	44.57	36.10	29.68	25.44	20.37
	0.4	65.74	53.93	44.23	36.46	30.59	24.85	20.00
	0.6	65.01	53.50	43.81	35.57	28.53	23.99	19.77
	0.8	65.11	53.65	43.81	36.05	30.66	23.94	20.38

Table 4.11: Sample mean of estimated age effects ($\hat{\boldsymbol{\beta}}$) from the 100 runs of the size-attained model fit to simulated data with $\sigma = 5$ and $\sigma_u = 10$ for each of the five values of ρ.

actual estimates as the mean of the estimates is close to the true parameter value for all values of ρ. But, Table 4.12 and Table 4.14 indicate that, often, the simulated standard error of the estimated age effects increases for larger values of ρ, while the mean of the standard errors of the estimated age effects from the simulations decreases. This suggests that the standard errors of the estimated age effects derived from the model could be greatly underestimating the true standard error in cases where ρ is large. This is more exaggerated when $\sigma = 5$ and $\sigma_u = 10$. For example, when $\rho = 0.8$, the simulated standard error for the estimated age three effect is about 8.2, but the mean of the standard errors from the simulations is only about 2.5. In contrast, when $\rho = 0$, the simulated standard error for the estimated age three effect is about 4.6, while the mean of the standard errors from the simulations is about 3.6.

The ρ parameter could be incorporated into the model and estimated along with $\boldsymbol{\beta}$, σ, and σ_u. But, the goal of this section was just to investigate the robustness of this method under the assumption that $\rho = 0$. The larger values of ρ, in general, do not appear to affect the actual estimators of σ or $\boldsymbol{\beta}$. Larger values of ρ did appear to lead to further underestimation of σ_u. The larger values of ρ also appeared to increase the standard error of $\hat{\sigma}_u$ and decrease the standard error of the estimated age effects.

		Age													
		1		2		3		4		5		6		7	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B
ρ	0	4.7	3.8	5.0	4.1	4.6	3.6	5.6	4.9	8.5	7.4	5.1	4.3	4.7	3.9
	0.2	4.7	3.7	5.2	4.0	4.5	3.5	6.0	4.8	9.2	7.2	5.3	4.2	4.8	3.8
	0.4	7.0	3.5	7.4	3.8	6.9	3.3	8.0	4.6	10.0	7.0	7.3	4.0	6.9	3.6
	0.6	7.4	3.3	7.4	3.6	7.1	3.1	8.6	4.4	10.8	6.7	7.2	3.8	7.3	3.4
	0.8	8.2	2.8	8.9	3.1	8.2	2.5	8.9	4.0	12.8	5.9	9.1	3.4	8.5	3.0

Table 4.12: Standard deviation of estimated age effects (A) and sample mean of standard error of estimated age effects (B) from the 100 runs of the size-attained model fit to simulated data with $\sigma = 5$ and $\sigma_u = 10$ for each of the five values of ρ.

		Age						
		1	2	3	4	5	6	7
Parameter		65.27	53.55	43.92	36.03	29.56	24.25	19.89
ρ	0	65.06	53.57	43.93	36.09	30.24	23.84	19.75
	0.2	65.21	53.57	43.83	36.95	30.23	24.06	19.79
	0.4	65.24	54.08	44.13	36.50	27.84	25.02	19.87
	0.6	64.66	53.18	43.38	35.80	28.46	24.27	19.32
	0.8	65.07	53.46	43.74	36.05	31.50	22.92	19.98

Table 4.13: Sample mean of estimated age effects ($\hat{\boldsymbol{\beta}}$) from the 100 runs of the size-attained model fit to simulated data with $\sigma = 10$ and $\sigma_u = 5$ for each of the five values of ρ.

		Age													
		1		2		3		4		5		6		7	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B
ρ	0	4.4	3.5	5.3	4.0	3.8	2.7	7.3	6.0	10.5	7.5	5.6	5.1	5.0	4.1
	0.2	4.1	3.5	4.9	4.1	3.8	2.8	6.1	6.0	11.7	7.7	5.9	5.1	4.7	4.1
	0.4	5.2	3.3	5.2	3.8	4.0	2.5	6.6	5.8	11.4	7.2	6.2	5.0	5.6	4.0
	0.6	5.3	3.4	5.7	3.8	5.0	2.6	7.0	5.8	12.2	7.3	6.1	5.0	5.6	4.0
	0.8	5.4	3.2	6.8	3.7	5.2	2.4	7.4	5.7	11.6	6.9	6.0	4.9	6.2	3.9

Table 4.14: Standard deviation of estimated age effects (A) and sample mean of standard error of estimated age effects (B) from the 100 runs of the size-attained model fit to simulated data with $\sigma = 10$ and $\sigma_u = 5$ for each of the five values of ρ.

	1	2	3	4	5	6	7	$\hat{\sigma}$	$\hat{\sigma}_u$	k
Parameter	65.27	53.55	43.92	36.03	29.56	24.25	19.89	5	$\sqrt{2}$	1
A	65.20	53.71	43.83	36.63	28.87	24.41	19.81	4.97	0.19	–
B	1.62	2.42	1.38	3.18	4.43	2.96	2.18	0.17	0.60	–
C	1.61	1.81	1.19	2.82	3.42	2.43	1.93	0.17	28.68	–
Parameter	65.27	53.55	43.92	36.03	29.56	24.25	19.89	5	$\sqrt{20}$	10
A	65.05	53.35	43.80	35.82	29.94	24.06	19.66	4.99	2.28	–
B	3.06	2.97	2.93	4.20	5.79	3.41	2.93	0.17	2.13	–
C	2.11	2.44	1.80	3.39	4.66	2.85	2.37	0.17	19.53	–

Table 4.15: Sample mean of the estimates (A), sample standard deviation of the estimates (B), and sample mean of standard error of the estimates (C) from the 100 simulations with the parameter values shown. The numbers 1-7 represent the estimated age effects for those ages.

4.4.2 Non-Normal year effects

In this section, we explore taking the year effects as random draws from a $\sigma_u^2/(2k)\left[\chi_k^2 - k\right]$ distribution, rather than from a Normal distribution. This is a multiple of the χ_k^2 distribution, scaled to have standard deviation σ_u and centered to have mean zero. Thus this distribution is skewed to the right, with skewness controlled by the degrees of freedom, k. Smaller values of k correspond to a more skewed distribution. For one simulation, $k = 1$ and $\sigma_u = \sqrt{2}$, and for another simulation, $k = 10$ and $\sigma_u = \sqrt{20}$. In both cases, $\sigma = 5$ and the parameter values used for $\boldsymbol{\beta}$ are shown in Table 4.15. One hundred simulated lengths at time of capture were created for the two cases and the size-attained model was fit to the data. Table 4.15 shows the results.

In both cases, the estimated age effects are close to the true parameter values, but the sample mean of the standard error of the estimated age effects are less than the sample standard deviation of the estimated age effects. This same phenomenon was seen in the cases when the year effects were a random draw from $N(\mathbf{0}, \sigma_u^2\mathbf{I}_m)$ (see Table 4.4), so likely is not due to taking the year effects as random draws from χ^2 distribution.

The estimate of σ is also near its true parameter value in both cases, and the sample standard deviation of $\hat{\sigma}$ is about the same as the sample mean of the standard errors of $\hat{\sigma}$. The σ_u parameter is underestimated in both cases. When $\sigma_u = \sqrt{2} \approx 1.41$ and $k = 1$, the sample mean of the $\hat{\sigma}_u$s from the one hundred simulations is 0.19. The sample mean of the $\hat{\sigma}_u$ from the one hundred simulations when $\sigma_u = \sqrt{20} \approx 4.47$ and $k = 10$ is 2.28. Underestimation of $\hat{\sigma}_u$ was also seen in the original model when σ_u was less than σ (see Figure 4.2). The sample mean of the standard errors of $\hat{\sigma}_u$ is much larger than the sample standard deviation of $\hat{\sigma}_u$, which also agrees with what was observed in the original model with year effects drawn from $N(\mathbf{0}, \sigma_u^2 \mathbf{I}_m)$.

Chapter 5

Example: Lake Winnebago Drum

5.1 The data

Freshwater drum (*Aplodinotus grunniens*) are common fish in North America, inhabiting lakes and rivers in the eastern two-thirds of the continent, from the Hudson Bay drainage to the Gulf States and from Montana and eastern Mexico to Pennsylvania (NatureServe 2010 as cited in Davis-Foust 2011). We are interested in learning about the growth of drum in Lake Winnebago, Wisconsin, where this fish has the highest biomass of any fish species. Davis-Foust (2011, p. 3) states that "Lake Winnebago is an economically important sport fishery ... [with] ... few consumption advisories, but it is also one of the few remaining waterbodies that is relatively free of the aquatic invasive species that plaque most of the rest of the Great Lakes region." Due to the drum's abundance, an accurate picture of its growth cycle is needed in order to assess its impact on the ecological system of Lake Winnebago.

The drum were collected from a variety of samples, including fishing tournaments, boom electroshocking, and assessment trawling. Data that will be used in modeling growth include date of capture, age at capture, length at capture (mm), otolith radius at capture (pixels), yearly growth of the otolith radius (pixels), and sex, which is only available for a subset of the data.

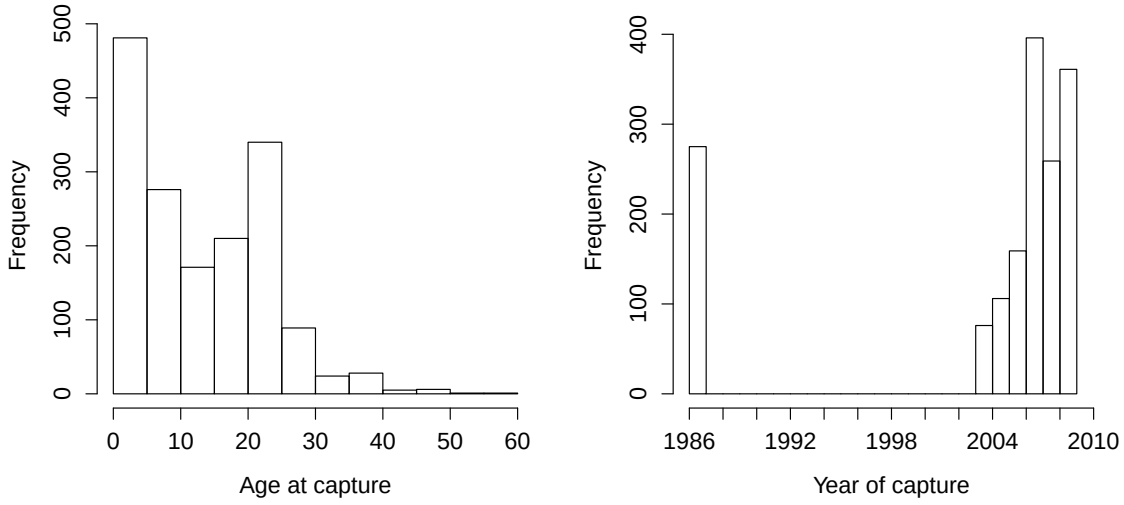


Figure 5.1: Distributions of age at capture and year of capture for drum in the original data set.

There are records for 1632 drum. The distributions of age at capture and year of capture are shown in Figure 5.1. The median age at capture is twelve years with just under 95% being less than twenty-seven years. Two hundred seventy-five drum were captured in 1986 with the rest being captured between 2003 and 2009. Since there are no data on drum captured between 1986 and 2003, we will eliminate all drum captured in 1986 from the data set. Keeping them in the data set would result in early predicted year effects being based only on drum captured in year 1986.

Eliminating the drum caught in 1986 leaves 1357 drum. We also eliminate drum that were captured after the age of 26, as there are only sixty of them and no more than seven in each age category. We call this the *reduced data set*. All proceeding analyses are performed on the reduced data set.

Distributions for the age at capture and yearclass (year of birth) for the reduced data set are shown in Figure 5.2. The median age at capture is eleven years.

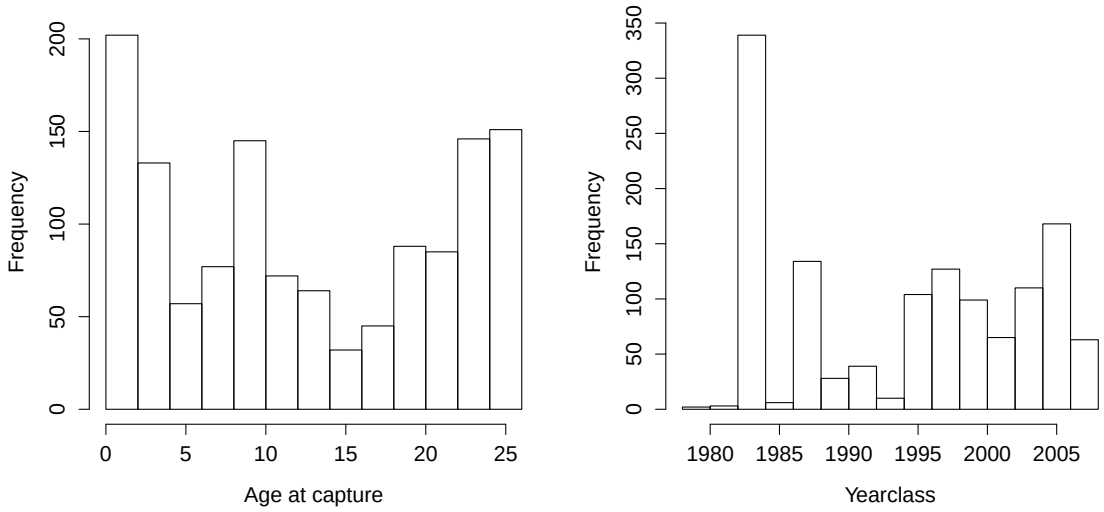
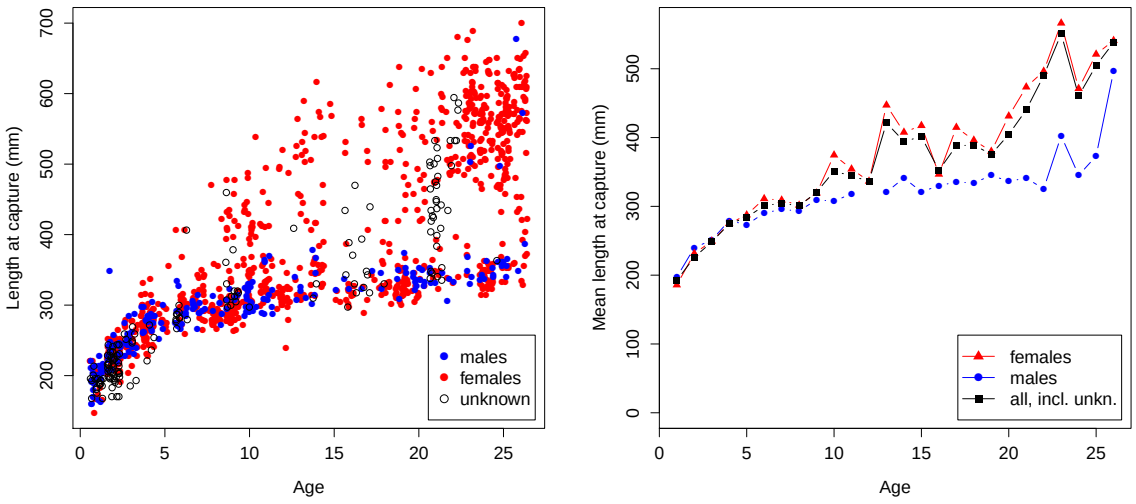


Figure 5.2: Distribution of age at capture and yearclass for drum in the reduced data set.

Yearclasses range from 1979 to 2008. Yearclass 1983 is noticeably larger than any other yearclass, containing 329 drum. This is probably due to a much larger than usual hatch in that year (Davis-Foust, 2011).

Figure 5.3a shows the lengths at capture by age at capture for males, females, and unknown sex. At the younger ages, there appears to be less variability in the lengths and length does not appear to be significantly different for males and females. At the older ages, there appears to be more variability in length, especially for female drum. Figure 5.3b summarizes the raw length at capture data, showing the mean length at time of capture by age at capture. Females are shown as red triangles, males as blue dots, and all drum, including those with unknown gender, as black squares. Drum grow quickly in the first year where both males and females average nearly 200 millimeters in length. Growth of male and female drum appear fairly similar before age ten. After this point males grow very little while females continue to grow,



(a) Length at time of capture for males, females, and unknown sex.

(b) Mean length at time of capture.

Figure 5.3: Individual lengths and mean length by age at capture.

averaging around 350 and 450 millimeters in length by the mid-twenties, respectively.

5.2 Length-attained model

First, the data are analyzed using the size-attained model, (2.4), with length at capture as the response variable. We call this the *length-attained* model. The initial analysis pools all the data for males, females, and unknowns, but we will test adding a sex effect later. The cumulative estimated age effects and predicted year effects from this model are shown in Figure 5.4. As expected from what was observed in Figure 5.3a, the estimated growth from birth to age one is large, about 189.0 millimeters. After the first year, most estimated age effects are less than 50 millimeters.

The cumulative estimated age effects are shown, rather than the estimated age effects, to better illustrate the growth pattern of a fish. Let $\mathbf{C}$ be a square matrix with

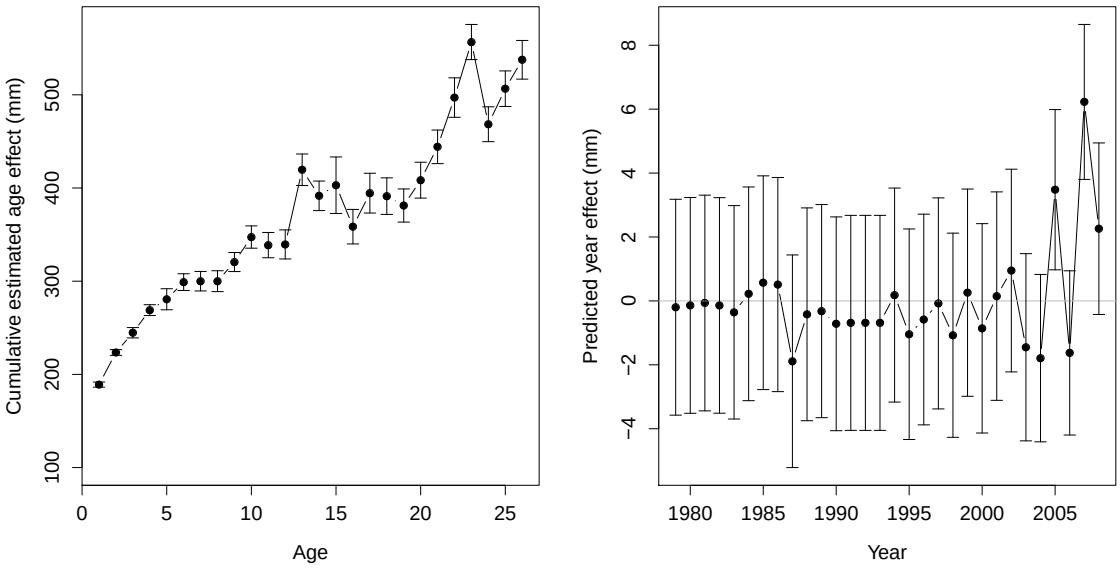


Figure 5.4: Cumulative estimated age effects ($\mathbf{C}\hat{\boldsymbol{\beta}}$) and predicted year effects ($\hat{\mathbf{u}}$) with standard error bars using length at capture as the response in the size-attained model.

the number of rows and columns equal to the number of estimated age effects, where there are zeros in the upper triangle and ones in the lower triangle and diagonal. Then the cumulative estimated age effects can be written as $\mathbf{C}\hat{\boldsymbol{\beta}}$. If $\mathbf{V} = \text{var}(\hat{\boldsymbol{\beta}})$, then $\text{var}(\mathbf{C}\hat{\boldsymbol{\beta}}) = \mathbf{C}\mathbf{V}\mathbf{C}^{\text{T}}$. The cumulative estimated age effects are not constantly increasing, as might be expected. Instead, cumulative estimated age effects sometimes decrease as age increases. For example, the cumulative estimated age effect at age twenty-three is over 550 mm and at age twenty-four is under 500 mm. The standard errors for the cumulative estimated age effects tend to be larger at the older ages.

Predicted year effects are also shown in Figure 5.4. The predicted year effects are all relatively close to zero, with the largest being 6.2 millimeters in year 2007. The standard errors are large compared to the size of the predicted effects.

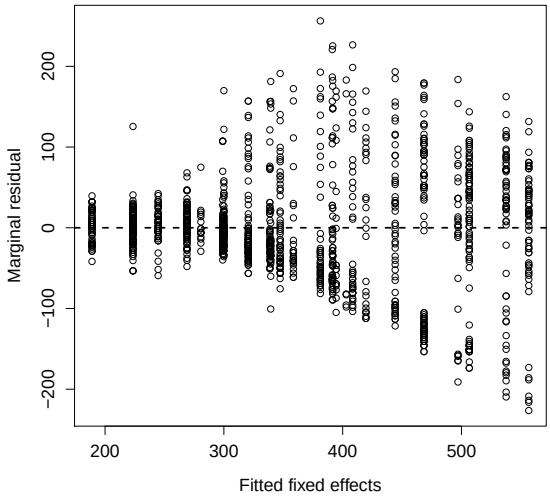
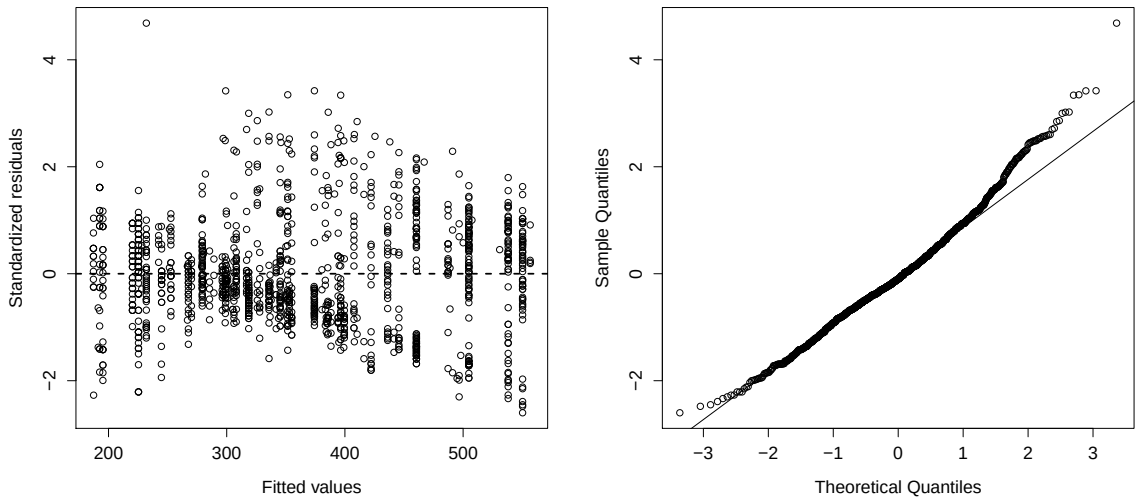


Figure 5.5: Plot of marginal residuals, $\hat{\boldsymbol{\xi}}$, versus the fitted fixed effects, $\mathbf{X}\hat{\boldsymbol{\beta}}$.

We can check some model assumptions through residual analysis, which was outlined in Section 3.4. To verify that the model follows the linearity assumption, we examine a plot of the marginal residuals, $\hat{\boldsymbol{\xi}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, versus the fitted fixed effects, $\mathbf{X}\hat{\boldsymbol{\beta}}$. There appears to be some evidence of non-linearity as seen by the curved pattern in Figure 5.5. There are more residuals less than zero when the fitted fixed effects are between 300 and 400 and more residuals greater than zero when the fitted fixed effects are greater than 500.

Standardized conditional residuals, $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}})/(\hat{\sigma}\sqrt{p})$, were computed and plotted versus the fitted values, $\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$ to assess heteroscedasticity and a Q-Q plot was created to assess normality of $\boldsymbol{\delta}$. They are shown in Figure 5.6. Figure 5.6a shows random spread around zero but non-linearity is still evidenced here, as it was in Figure 5.5. Figure 5.6b shows some evidence of non-normality; the distribution of the residuals is slightly right-skewed. The same plots were created with the studentized conditional residuals, instead of the standardized conditional residuals. They were



(a) Standardized conditional residuals versus fitted values.

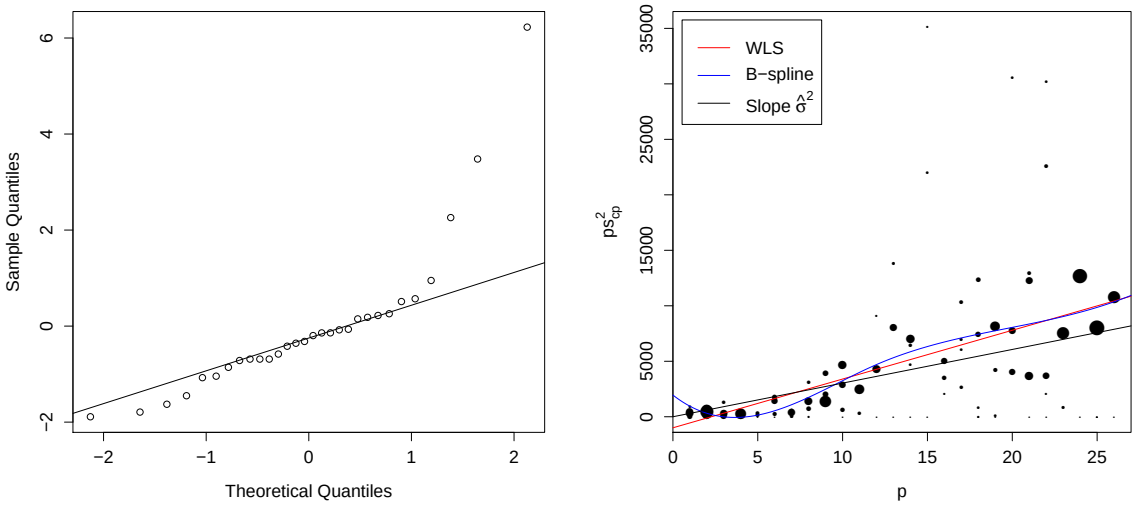
(b) Q-Q plot of standardized conditional residuals.

Figure 5.6: Diagnostic plots of standardized conditional residuals.

similar to those seen in Figure 5.6.

A Q-Q plot of the $\hat{\mathbf{u}}$ in Figure 5.7a shows evidence that they are non-normal with some potential outliers, although with only thirty predicted random effects this is difficult to assess. There were not any outlying points in the plot of $\mathbf{Z}\hat{\mathbf{u}}$ versus observation number.

In the size-attained model (2.3), it is assumed that the ε_{cka} are independent random draws from $N(0, \sigma^2)$ and thus the δ_{ck} are independent $N(0, p\sigma^2)$. One way to check this model assumption is by examining the within (c, p) class variances in size, s^2_{cp} (see (3.9) for a formal definition). Figure 5.7b shows a scatterplot of ps^2_{cp} on the y-axis and age at capture, p, on the x-axis. The area of the dots is proportional to the sample size for the (c, p) class. If $\text{var}(\delta_{ck}) = p\sigma^2$, then this plot should show a linear relationship between ps^2_{cp} and p. A weighted least squares line and b-spline curve



(a) Q-Q plot of predicted random effects.

(b) Scatterplot of ps_{cp}^2 versus p.

Figure 5.7: More diagnostic plots used to check model assumptions.

were fit to these data and are indicated on the scatterplot. The line with intercept zero and slope $\hat{\sigma}^2$ is also shown as a point of reference. The b-spline curve does not deviate much from the weighted least squares line which provides evidence that the assumption of $\text{var}(\delta_{ck}) = p\sigma^2$ holds to a reasonable approximation.

Male and female drum appear to grow differently, as evidenced in Figure 5.3. We first explore adding a sex term to the model, thus fitting the model shown in (3.14). This model fits parallel growth curves for males and females. The estimated coefficient for the sex term is 13.0 with a standard error of 3.37, indicating that, on average, female fish are about 13.0 mm longer than male fish. A likelihood ratio test ($\chi_1^2 = 14.288$, p-value ≈ 0.0002) shows this model is preferred over the model with no sex term (the model was refit with unknowns taken out so that the same number of observations were used in each model). To further improve this model, we allow the growth patterns to be different for males and females, fitting the model shown in

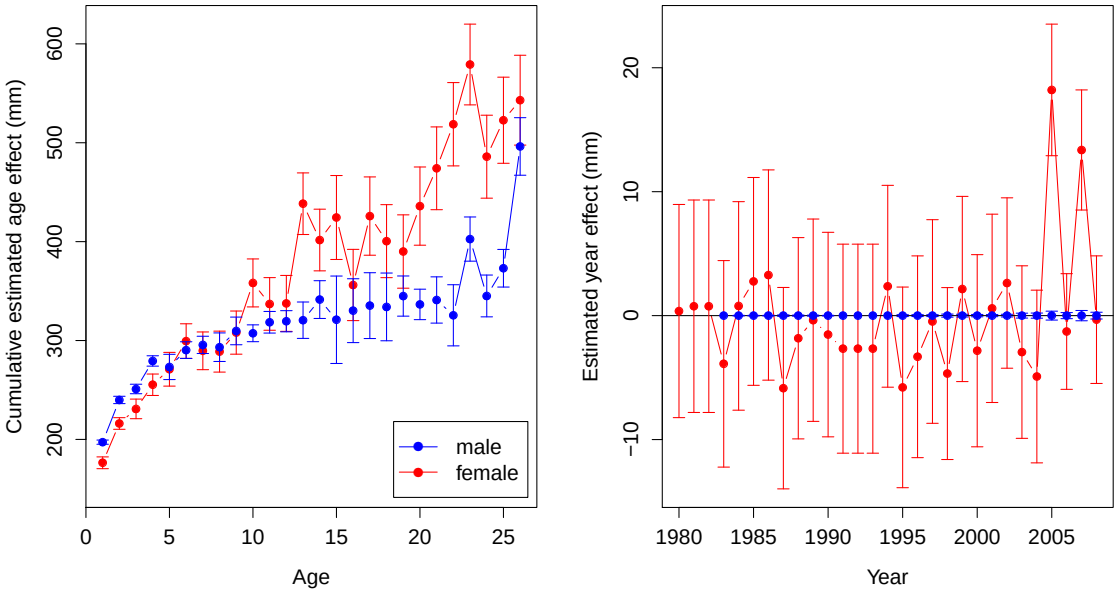


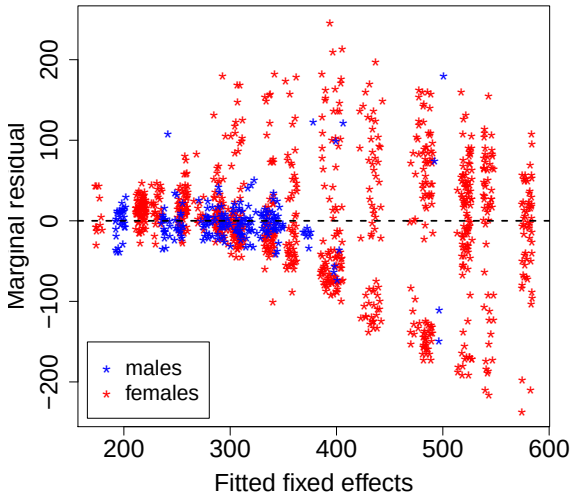
Figure 5.8: Cumulative estimated age effects ($\mathbf{C}\hat{\boldsymbol{\beta}}$) and predicted year effects ($\hat{\mathbf{u}}$) with standard error bars for male and female drum using length at capture as the response in the size-attained model.

(3.15). The likelihood ratio test shows this to be the preferred model ($\chi^2_{25} = 140.732$, p-value ≈ 0).

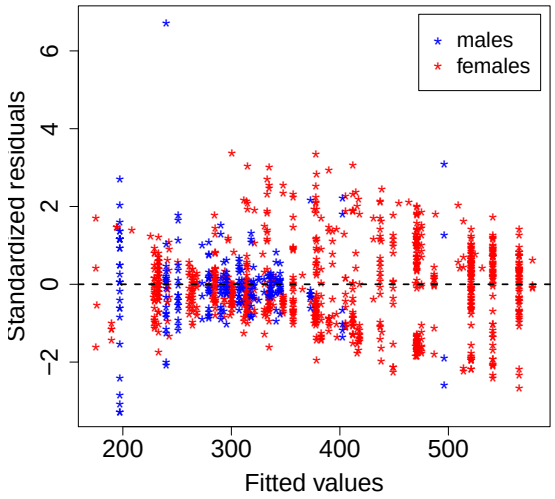
We could also allow the predicted year effects to be different for males and females. This can be accomplished by fitting the size-attained model separately for males and females. The likelihood ratio test shows this to be the best of all the models we examined ($\chi^2_{29} = 59.912$, p-value ≈ 0.0006). The results of modeling males and females separately are shown in Figure 5.8. The cumulative estimated age effects for males stay fairly flat after age ten, except at the last four years which are largely due to the five male fish that appear abnormally large (see Figure 5.3a). The cumulative estimated age effects for females continue to increase with no clear evidence as to where growth might level off. The predicted year effects for males and the standard

errors are all very close to zero. This is largely because $\hat{\sigma}_u \approx 0.0014$. Most of the predicted year effects for females are close to zero compared to their standard errors. The largest predicted effects are in years 2005 and 2007, which are both positive.

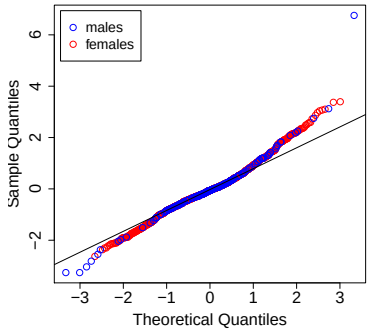
Residual plots were examined for this model. Figure 5.9a shows a plot of the marginal residuals, $\hat{\boldsymbol{\xi}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, versus the fitted fixed effects, $\mathbf{X}\hat{\boldsymbol{\beta}}$. There appears to be some evidence of non-linearity as seen by the curved pattern. The plot of the standardized conditional residuals, $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}})/(\hat{\sigma}\sqrt{p})$, versus the fitted values, $\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$, in Figure 5.9b shows similar spread around zero throughout the plot. This gives evidence that the variance assumption on $\boldsymbol{\delta}$ is correct. Figure 5.9c shows a Q-Q plot of the standardized conditional residuals. The long tails indicate a potential violation of the normality assumption on $\boldsymbol{\delta}$. In the model with all fish, the standardized conditional residuals were right skewed (see Figure 5.6b). Q-Q plots of the predicted random effects, $\hat{\mathbf{u}}$, were created for both the males and females and are shown in Figure 5.9d and Figure 5.9e, respectively. The plot for males indicates long tails in the distribution. The plot for females shows a couple potential outliers and otherwise does not appear to deviate from normality.



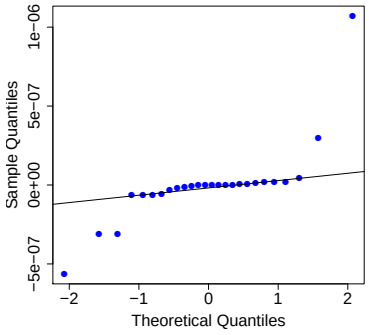
(a) Standardized conditional residuals versus fitted values.



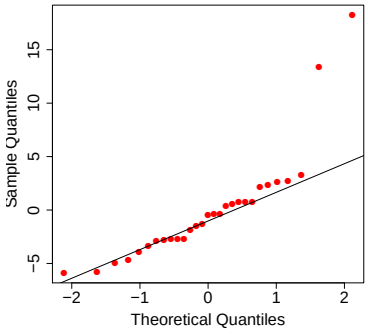
(b) Standardized conditional residuals versus fitted values.



(c) Q-Q plot of standardized conditional residuals.

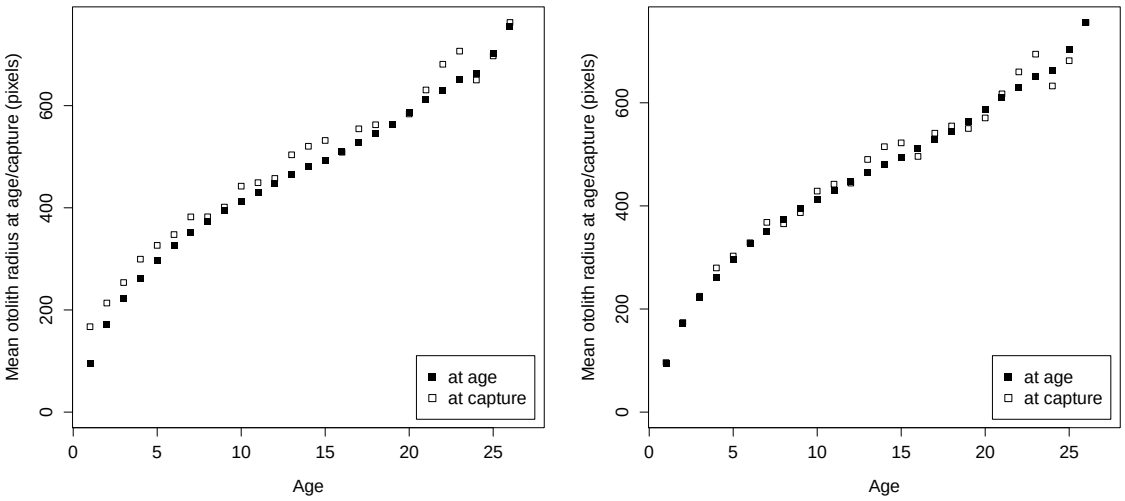


(d) Q-Q plot of predicted random effects for males.



(e) Q-Q plot of predicted random effects for females.

Figure 5.9: Diagnostic plots when females and males are modeled separately.



(a) Includes extra growth in the time of capture measurements.

(b) Excludes extra growth in the time of capture measurements.

Figure 5.10: Mean otolith radius, either at time of capture or at age.

5.3 Comparing the size-attained and incremental models

In order to compare the size-attained model, (2.4), to the incremental model, (1.9), we will use data collected on otoliths. We use otolith radius at time of capture as the response variable in the size-attained model and the yearly growth of the otolith as the response variable in the incremental model. For simplicity, these models are referred to as the *size-attained* and *incremental* models, respectively.

First, some exploratory plots are examined. Figure 5.10 shows the mean otolith radius at time of capture (unfilled square) or the mean otolith radius at age (filled square). The mean otolith radius at time of capture uses only the measurements of the otolith radius at the age at which the drum was captured. So, for example, the

otolith radius for a drum caught at age ten would only be used in computing the mean otolith radius at age ten. These are the data that are used in the size-attained model. The mean otolith radius at age uses all measurements of the otolith radius up until the age at which the drum was captured. So, the otolith radius from a drum caught at age ten would be used in computing the mean otolith radius at ages one through ten, where the radius at each age is computed from the yearly growth of the otolith radius. For the graph on the left, the otolith radius at time of capture is computed simply by summing all the increments, including the most recent partial growth. So, a ten year old fish would actually have some growth from its current year included. The graph on the right does not include the partial growth in the otolith radius at time of capture.

In the graph on the left, the mean otolith radius at time of capture is almost always larger than the mean otolith radius at age. This is especially noticeable at age one, where the mean otolith radius at time of capture is 167.2 pixels and the mean otolith radius at age is 94.8 pixels. In the graph on the right, the mean otolith radius at time of capture and mean otolith radius at age are very similar, suggesting that the differences in the graph on the left are mostly due to including the extra growth beyond the last ring formation. Although using the radius at time of capture that includes the current year's growth is more reflective of what would be used in the size-attained model with length data, we will use the radius at time of capture without extra growth so that we can more accurately compare the methods. Differences in estimates and predictions will mostly be due to differences in modeling methods rather than differences in data.

Cumulative estimated age effects and predicted year effects from the size-attained model and the incremental model fit to all the otolith data (males, females, and uknowns) shown in Figure 5.11. The cumulative estimated age effects from the size-attained model have larger standard errors than those from the incremental model.

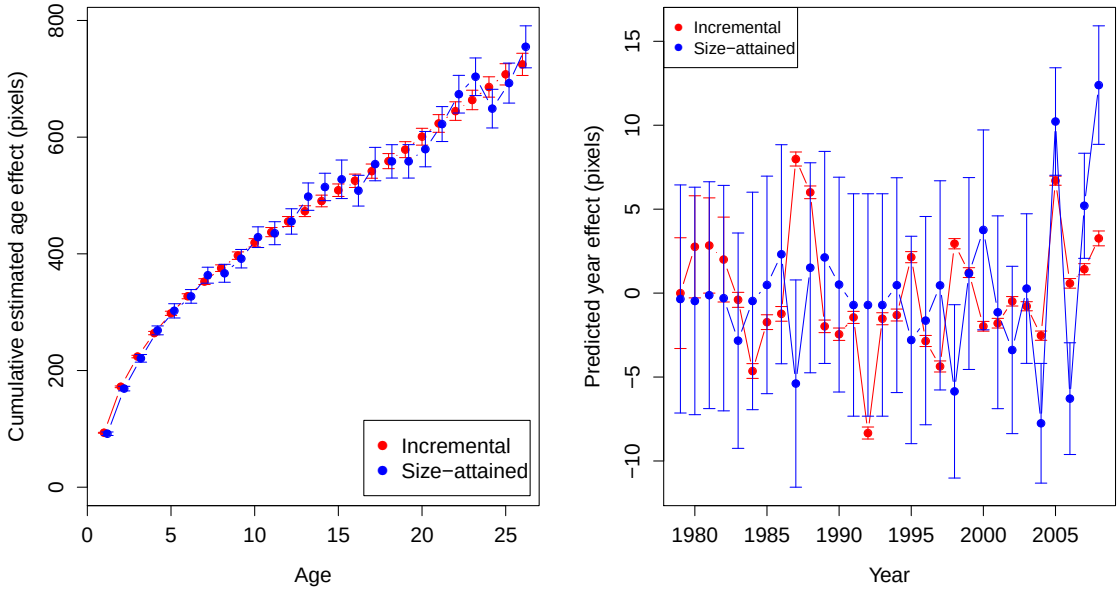


Figure 5.11: Cumulative estimated age effects ($\mathbf{C}\hat{\boldsymbol{\beta}}$) and predicted year effects ($\hat{\mathbf{u}}$), with one standard error bar, using otolith data. The blue points are the estimates/predictions from the size-attained model using otolith radius. The red points are the estimates/predictions from the incremental model using otolith increments.

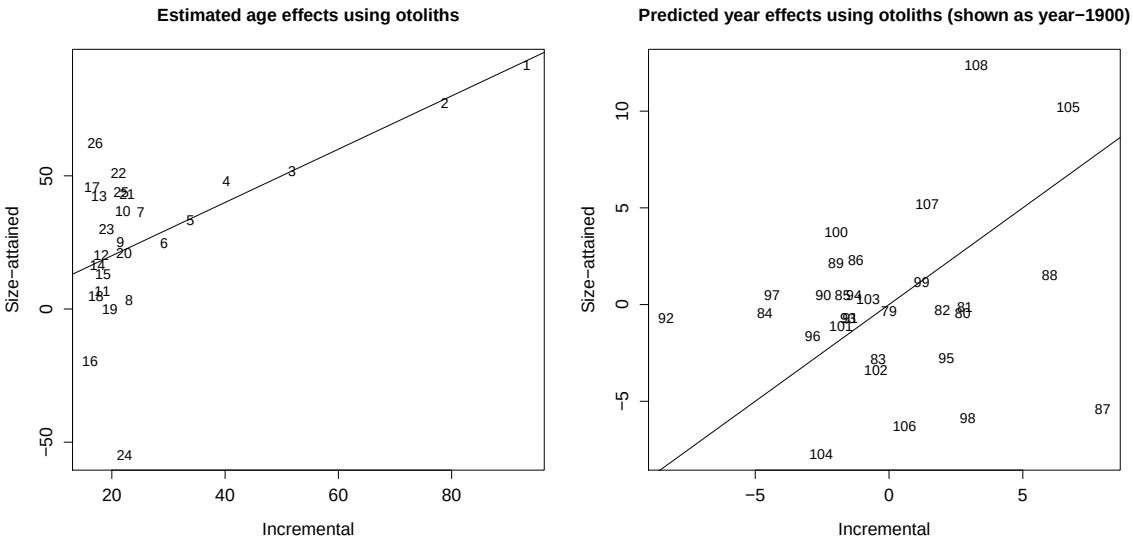


Figure 5.12: Scatterplots of estimated age effects ($\hat{\boldsymbol{\beta}}$) and predicted year effects ($\hat{\mathbf{u}}$) using otolith data. The vertical axis shows the estimates/predictions from the size-attained model. The horizontal axis shows the estimates/predictions from the incremental model.

The cumulative estimates from the size-attained model are similar to those from the incremental model, especially up until age thirteen. The predicted year effects from the size-attained model also have larger standard errors than those from the incremental model; the predicted year effects from the two models do not appear to follow similar patterns.

In order to better compare the estimates and predictions from the two models, scatterplots are created showing the estimates or predictions from the size-attained model on the y-axis and from the incremental model on the x-axis. The $y = x$ line is shown to indicate where the points would lie if the two methods produced exactly the same results.

As seen in both Figure 5.11 and Figure 5.12, the first six estimated age effects agree very well. After this point there is much more fluctuation in the estimates

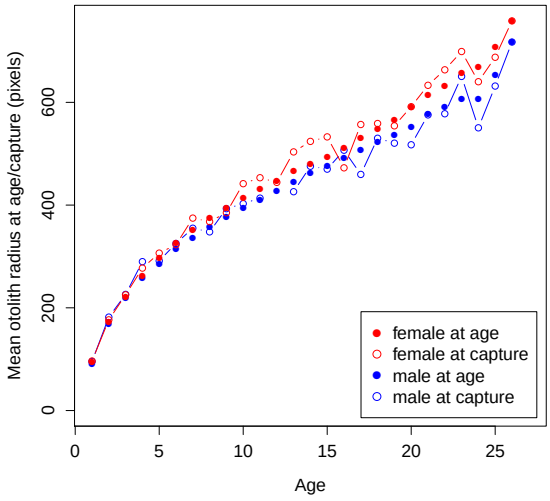


Figure 5.13: Mean otolith radius at capture/age for male and female drum.

from the size-attained model. While the estimated age effects for ages seven through twenty-six in the incremental model range from about 16 to 25 pixels, these same estimated effects in the size-attained model range from about -55 to 92 pixels. The correlation between the estimated age effects using the two different models is about $r = .61$. Regarding the predicted year effects, the points are widely spread around the $y = x$ line with a small correlation of only about .18.

Unlike the length-attained data where there appeared to be a difference in the growth of male and female drum (see Figure 5.3), the difference in otolith growth for the two sexes is much less dramatic. Figure 5.13 shows the mean otolith radius at capture and at age for both male and female drum. The average otolith growth for female drum appears slightly greater than for males at the older ages but neither male nor female growth appears to be leveling off. This is in contrast to what was observed in the length data where the growth for male drum appeared to level off around age fifteen.

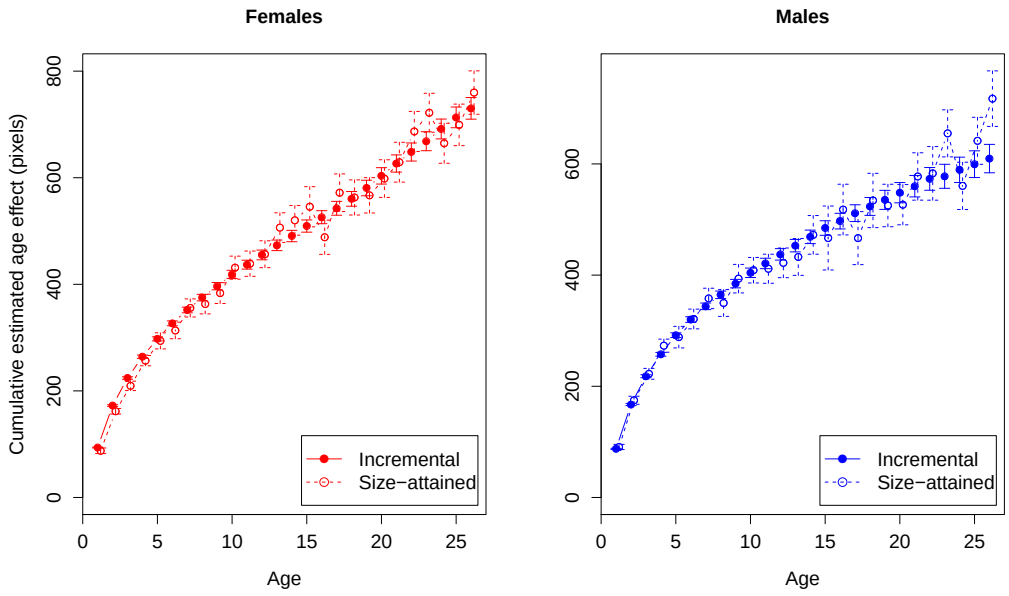


Figure 5.14: Cumulative estimated age effects with one standard error bar for male and female drum from the size-attained and incremental models using otoliths.

The cumulative estimated age effects and predicted year effects for male and female drum using the incremental and size-attained models are shown in Figure 5.14 and Figure 5.15. For both male and female drum, the cumulative estimated age effects from the size-attained model are similar to those from the incremental model in about the first ten years. Whereas the estimates from the incremental model show a constantly increasing pattern throughout, the estimates from the size-attained model at times decrease as age increases. The estimated year effects from the two models greatly differ. Year effects for the early years predicted from the size-attained model are relatively small compared to their large standard errors, but from the incremental model some of these year effects appear significantly different from zero. From the size-attained model the only potentially significant year effects are 2005 and 2008 for male drum and 2005, 2007, and 2008 for female drum.

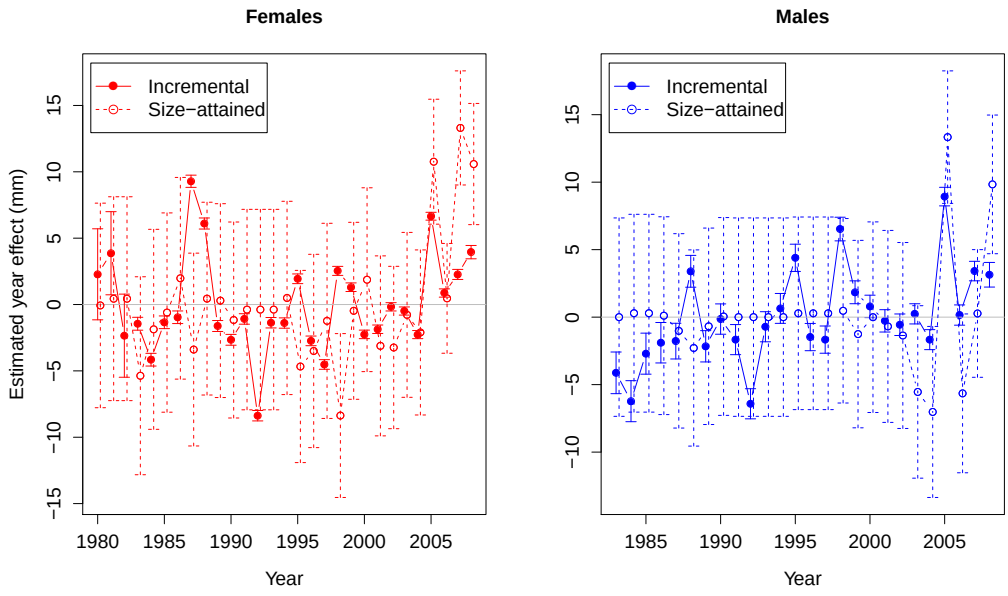


Figure 5.15: Predicted year effects with one standard error bar for male and female drum from the size-attained and incremental models using otoliths.

5.4 Comparing the otolith and length models

Also of interest is to compare the model using total otolith radius as the response in the size-attained model to the model using length at time of capture as the response in the size-attained model. We call these the *otolith-attained* and *length-attained* models, respectively. In both cases, we include the current year's growth. Since the units will be different, we will just compare the estimates and predictions by looking at the scatterplots of estimated and predicted effects from both models. These are shown in Figure 5.16.

Under the hypothesis that otolith growth and body growth occur at the same rate, we would expect the points to fall on a straight line. The line should have a positive slope, which would mean that large age and year effects from the length-attained model also have large age and year effects in the otolith-attained model.

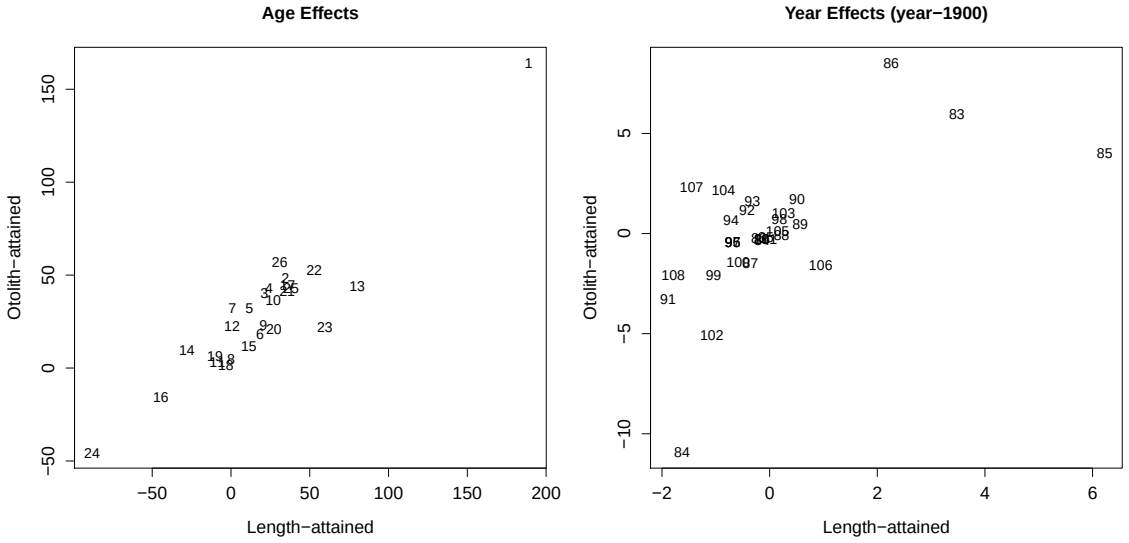


Figure 5.16: Scatterplots of estimated age effects ($\hat{\boldsymbol{\beta}}$) and predicted year effects ($\hat{\mathbf{u}}$) from the otolith-attained and length-attained models.

The correlation between the estimated age effects using length-attained and otolith-attained is about .93, indicating a strong linear relationship. The correlation between the predicted year effects using length-attained and otolith-attained is about .63, indicating a moderate linear relationship.

5.5 von Bertalanffy growth

The large fluctuations of the estimated age effects in the size-attained model (see Figure 5.4) do not make intuitive or biological sense. One way to force a smooth relationship is by replacing the mean function that uses fixed age effects with a growth family. The growth family we use is a modified version of the von Bertalanffy function, which is commonly used in modeling fish growth (Haddon and Haddon, 2010). The size-attained model is

	Estimate	St. Err.
L_∞	734.94	341.76
K_0	14.79	8.66
a_0	-5.52	0.93

Table 5.1: Estimates and standard errors of the parameters in the von Bertalanffy function, (5.1).

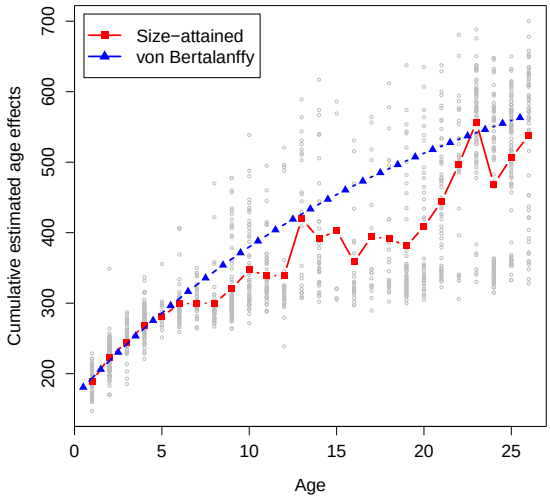


Figure 5.17: Fit of the von Bertalanffy growth function and cumulative estimated age effects from the size-attained model.

$$y_{ck} = L_\infty(1 - e^{-(\log(2)/K_0)(p-a_0)}) + \sum_{a=1}^{p} h_{c+a-1} + \delta_{ck}. \tag{5.1}$$

In the mean function, L_∞, K_0, and a_0 are the parameters to be estimated. L_∞ is the asymptote for the model of average length-at-age, K_0 is the age at which the fish reaches 50% of its asymptotic growth, and a_0 is the age at which the average size is 0, although this is usually an extrapolation thus not necessarily meaningful (Ogle, 2010).

	Female	Male
L_∞	734.92	362.19
K_0	14.33	2.12
a_0	-5.08	-1.53

Table 5.2: Estimates of the parameters in the von Bertalanffy function, (5.1), where male and female drum are modeled separately

This model was fit to the length at capture data. Estimates and standard errors of the parameters in the von Bertalanffy function are shown in Table 5.1. Figure 5.17 shows the fit of the von Bertalanffy function to these data and also shows the cumulative estimated age effects from the size-attained model. At the younger ages, the estimated von Bertalanffy curve closely matches the cumulative estimated age effects from the size-attained model, but at the older ages, the fitted values from the von Bertalanffy curve are usually greater than the cumulative estimated age effects from the size-attained model. Although the von Bertalanffy growth function yields a worse fit than the fixed age effects ($X^2 = 334.8$ on 23 degrees of freedom, p-value ≈ 0), one could argue that the fluctuations seen in the estimated age effects do not make any intuitive sense.

The von Bertalanffy model, (5.1) was also fit separately for males and females. Estimates are shown in Table 5.2. The fitted curves and cumulative estimated age effects from the size-attained model are shown in Figure 5.18; the predicted year effects are shown in Figure 5.19. For female drum, the curve and the cumulative estimated age effects match well at the younger ages. After that, the von Bertalanffy curve usually estimates a larger cumulative age effect compared to the size-attained model. For the male drum, the von Bertalanffy curve and cumulative estimated age effects from the size-attained model only match well up until age two or three. The curve's ascent to the asymptote seems too quick since in the middle ages it has much larger cumulative estimated age effects compared to the size-attained model. It is

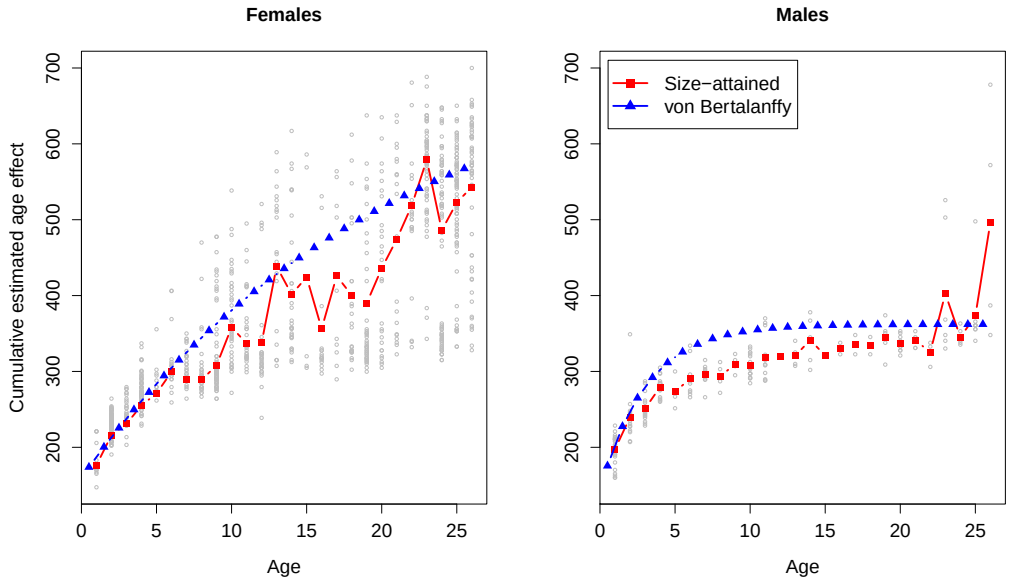


Figure 5.18: Fit of the von Bertalanffy growth function to the female and male drum data and respective cumulative estimated age effects from the size-attained models.

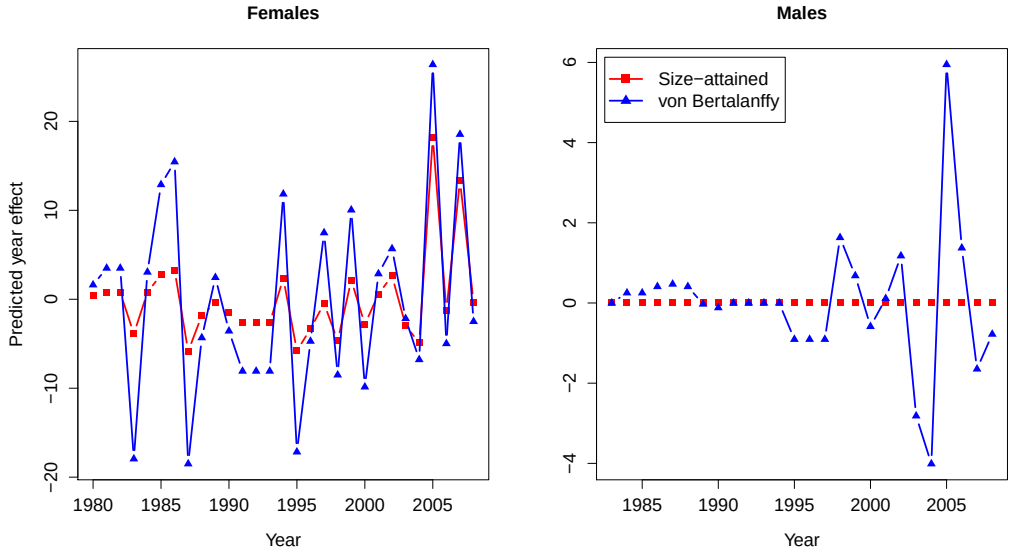


Figure 5.19: Predicted year effects from the size-attained model and the von Bertalanfyy model fit separately to the male and female drum data.

possible that another growth curve would be more appropriate for these data, but no other functions were explored.

For the females, the predicted year effects from both models appear to follow a similar pattern in that when the size-attained model predicts effects greater than zero, the von Bertalanffy model also predicts effects greater than zero and likewise when effects are predicted less than zero. But, the predictions from the von Bertalanffy model are generally greater in absolute value than those from the size-attained model. The predicted effects for males do not appear similar, mostly due to the predicted effects from the size-attained model being very close to zero. These predictions are close to zero because the estimate of σ_u from the size-attained model is about .001. From the von Bertalanffy model, $\hat{\sigma}_u = 8.79$.

5.6 Summary

The size-attained model was fit to data on freshwater drum captured from Lake Winnebago, Wisconsin. Using the otolith data, we were able to compare estimates and predictions from the size-attained and incremental models. We found that the younger estimated age effects agreed fairly well while the older estimated age effects did not have as much agreement and could be greatly influenced by a few outlying points in the size-attained model. The predicted year effects from the size-attained and incremental models had low correlations. Standard errors for both the estimated age effects and predicted year effects were much larger from the size-attained model than from the incremental model.

Results from the size-attained model using otolith radius and the size-attained model using length were compared and found to be similar. This suggests that, for these fish, otolith growth may be a good proxy for body growth.

A von Bertalanffy growth function was used instead of the fixed age effects to

create a smooth curve. Our results showed that the model with fixed age effects fit the data better, but the smooth curve may still make more sense biologically. In some cases, other growth curves may have fit the data better than the von Bertalanffy function.

Chapter 6

Summary and Future Work

Many methods of modeling fish growth rely on having data on the yearly growth of the fish. Since longitudinal data like these are often not available on their body growth, scale and other hard part growth are often used to infer somatic growth. In this thesis, we proposed a method of modeling body length at time of capture that is based on a model where longitudinal data are available. We call it the size-attained model, and it models length at capture by the sum of the relevant age effects, plus the sum of the relevant year effects, plus an error term.

Due to the complicated covariance structure of the data, a specialized R script was written that finds estimates of the parameters and standard errors of those estimates. We also found a way of dealing with large sample sizes. Typically, when the sample is of size n, inverting an $n \times n$ matrix is part of the estimation process. Instead, we modeled the average length in a unique age and yearclass combination, thereby greatly reducing the sample size, and proposed a new estimator for the error variance. We found that estimates and standard errors using this method were comparable to those derived from using all the data in the size-attained model.

Simulation studies were used to verify properties of the size-attained model and to compare it to other models. It was found that the variability of the year effects (σ_u) was often underestimated. This occurred less often when the sample size was

increased. In most cases, standard errors of the parameter estimates were also smaller than the actual variability observed in the estimates. Estimates from the size-attained model matched estimates from a model that used longitudinal data more closely when the variability of the year effects (σ_u) was large compared to the random variability (σ).

6.1 Future work

There are a few ways in which this research could be extended; they are listed here.

- Simulation studies showed that standard errors were often underestimated. Kenward and Roger (1997) and Harville and Jeske (1992) proposed other ways to estimate standard errors for fixed and random effects, respectively. These could be explored to see if they lead to improvements.

- Currently, the size-attained model is not very flexible. Assumptions could be relaxed, for example, allowing for correlated year effects or error terms.

- In order for users to be able to easily implement this model, an R package could be developed.

References

Bartlett, J., Randerson, P., Williams, R., and Ellis, D. (1984). The use of analysis of covariance in the back-calculation of growth in fish. *Journal of Fish Biology*, 24:201–213.

Briffa, K. R. and Melvin, T. M. (2011). A closer look at regional curve standardization of tree-ring records: Justification of the need, a warning of some pitfalls, and suggested improvements in its application. In Hughes, M. K., Swetnam, T. W., and Diaz, H. F., editors, *Dendroclimatology: Progress and Prospects*. Springer.

Carlander, K. D. (1982). Standard intercepts for calculating lengths from scale measurements for some centrarchid and percid fishes. *Transactions of the American Fisheries Society*, 111:332–336.

Carlin, B. P. and Louis, T. A. (2009). *Bayesian Methods for Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL, third edition.

Casselman, J. M. (1990). Growth and relative size of calcified structures of fish. *Transactions of the American Fisheries Society*, 119:673–688.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, New York.

Davis-Foust, S. L. (2011). Bomb radiocarbon age validation and long-term popu-

lation dynamics of freshwater drum in the lake winnebago system. *Unpublished manuscript, University of Wisconsin - Milwaukee.*

Francis, R. (1990). Back-calculation of fish length: a critical review. *Journal of Fish Biology*, 36:883–902.

Francis, R. (1995). The analysis of otolith data – a mathematician's perspective (what, precisely, is your model?). In Secor, D. H., Dean, J. M., and Campana, S. E., editors, *Recent Developments in Fish Otolith Research*. University of South Carolina Press.

Haddon, M. and Haddon, M. (2010). *Modelling and quantitative methods in fisheries*. Chapman & Hall/CRC.

Harville, D. A. and Jeske, D. R. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, 87(419):724–731.

Kackar, R. N. and Harville, D. A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics - Theory and Methods*, 10:1249–1261.

Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effect in mixed linear models. *Journal of the American Statistical Association*, 79(388):853–862.

Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53(3):983–997.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS ®for Mixed Models*. SAS Institute Inc., Cary, NC, 2^{nd} edition.

Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs – a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337.

McCulloch, C. E., Searle, S., and Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. Wiley, Hoboken, NJ, 2^{nd} edition.

NatureServe (2010). *NatureServe Explorer: An Online Encyclopedia of Life*.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer, New York.

Ogle, D. H. (2010). Von bertalanffy growth model vignette. *Unpublished manuscript, Northland College, Ashland, WI*.

Pereira, D. L., Bingham, C., Spangler, G. R., Conner, D. J., and Cunningham, P. K. (1995). Construction of a 110-year biochronology from sagittae of freshwater drum (*Aplodinotus grunniens*). In Secor, D. H., Dean, J. M., and Campana, S. E., editors, *Recent Developments in Fish Otolith Research*. University of South Carolina Press.

Petersen, K. B. and Pedersen, M. S. (2008). The matrix cookbook.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects in S and S-PLUS*. Springer, New York.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Richmond, L. (2009). Unpublished photograph of sectioned otolith showing annual rings. png file.

Santos Nobre, J. and da Motta Singer, J. (2007). Residual analysis for linear mixed models. *Biometrical Journal*, 49(6):863–875.

Searle, S., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. Wiley, New York, NY.

Speer, J. H. (2010). *Fundamentals of Tree-Ring Research*. University of Arizona Press, Tucson.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

Vigliola, L. and Meekan, M. G. (2009). The back-calculation of fish growth from otoliths. In Green, B. S., Mapstone, B. D., Carlos, G., and Begg, G. A., editors, *Tropical Fish Otoliths: Information for Assessment, Management and Ecology*. Springer.

Weisberg, S. (1986). A linear model approach to backcalculation of fish length. *Journal of the American Statistical Association*, 81(396):922–929.

Weisberg, S. (1993). Using hard-part increment data to estimate age and environmental effects. *Canadian Journal of Fisheries and Aquatic Science*, 50:1229–1237.

Weisberg, S., Spangler, G., and Richmond, L. S. (2010). Mixed effects models for fish growth. *Canadian Journal of Fisheries & Aquatic Sciences*, 67(2):269 – 277.

Whitney, R. R. and Carlander, K. D. (1956). Interpretation of body-scale regression for computing body length of fish. *The Journal of Wildlife Management*, 20(1):21–27.

Appendix A

Annotated R code for fitting the size-attained model

```
#####################################################################
####   ANNOTATED R CODE FOR FITTING THE SIZE-ATTAINED MODEL ####
#####################################################################

#NEEDED FOR THE ginv() FUNCTION, WHICH CALCULATES THE
#MOORE-PENROSE GENERALIZED INVERSE
library(MASS)

#READ IN DATA FILE
dat <- read.table("", header=T)

#NAME OF LENGTH VARIABLE, ONE RECORD FOR EACH FISH
lengthvar <- dat$Length
#NAME OF AGE VARIABLE, ONE RECORD FOR EACH FISH
agevar <- dat$Age
#NAME OF YEAR OF CAPTURE VARIABLE, ONE RECORD FOR EACH FISH
yrvar <- dat$Yearcap

#FUNCTION USED TO CREATE THE Z MATRIX
createZ <- function(agecap, yearcap){
   maxage <- max(agecap)
   firstyear <- min(yearcap) - maxage
   lastyear <- max(yearcap) - 1
   nyears <- lastyear - firstyear + 1
   n <- length(yearcap)
   years <- matrix(0, nrow = n, ncol = nyears)
   for (i in 1:n) {
```

```
      a <- agecap[i]
      y <- yearcap[i]
      years[i, (y-a):(y-1) - (firstyear) + 1] <- 1
   }
   yobs <- apply(years, 2, function(x) max(x) - min(x) > 0)
   years <- years[, yobs]
   colnames(years) <- paste("GY", (firstyear:lastyear) - 1900, sep = "")[yobs]
   years
}

#FUNCTION USED TO CREATE THE X MATRIX
createX <- function(agecap){
   maxage <- max(agecap)
   n <- length(agecap)
   ages <- matrix(0, nrow = n, ncol= maxage)
   for (i in 1:n) {
      a <- agecap[i]
      ages[i, 1:a] <- 1
   }
   colnames(ages) <- paste("A", 1:maxage, sep = "")
   ages
}

#NEGATIVE LOG-LIKELIHOOD, USED IN THE optim() FUNCTION
#TO FIND MLES FOR eta AND rho
negloglkhd <- function(parms, Y, X, D, Z){
  eta <- parms[1]
  rho <- parms[2]
  N <- length(Y)
  Siginv <- (1/exp(2*eta)) *
    solve(diag(rep(1,N)) + exp(rho)* diag(1/D) %*% tcrossprod(Z), diag(1/D))
  B <- ginv(crossprod(X, Siginv %*% X))  %*% crossprod(X, Siginv %*% Y)
  nll <- (N/2)*log(2*pi) + (1/2)*(2*N*eta + sum(log(D)) +
  log(det(diag(rep(1,N)) + exp(rho) * diag(1/D) %*% tcrossprod(Z)))) +
    (1/2) * crossprod((Y - X %*% B), Siginv %*% (Y - X %*% B))
  nll
}

#THE sizeatt() FUNCTION OUTPUTS ESTIMATED AGE EFFECTS, PREDICTED YEAR EFFECTS,
#THEIR SEs, ESTIMATES OF sigma AND sigma_u, STANDARDIZED AND STUDENTIZED
#CONDITIONAL RESIDUALS, FITTED VALUES, AND THE VALUE OF THE LOG-LIKELIHOOD
#EVALUATED AT THE PARAMETER ESTIMATES.
sizeatt <- function(lengthvar, agevar, yrvar){
```

```
# CREATE Y VECTOR, X & Z MATRICES, AND DIAGONAL OF D MATRIX
  D2 <- as.vector(agevar)
  Y2 <- as.vector(lengthvar)
  X2 <- as.matrix(createX(agevar))
  Z2 <- as.matrix(createZ(agevar, yrvar))
# STARTING VALUES FOR eta & rho
  p <- c(log(summary(lm(lengthvar ~ agevar + yrvar))$sigma), log(.25))
# OPTIMIZATION TO ESTIMATE eta & rho.
  opt <- optim(p, negloglkhd, method = "L-BFGS-B", lower=c(-Inf,-Inf),
     upper=c(Inf,10), Y=Y2, X=X2, D=D2, Z=Z2)
# FINDS THE VALUE OF THE LOG-LIKELIHOOD AT THE ESTIMATED PARAMTER VALUES.
# THIS IS OUTPUTED AND CAN BE USED TO COMPARE MODELS.
  llike <- -1*opt$value
# COMPUTE ESTIMATES OF sigma, sigma_u, beta, & u.
  sig <- exp(opt$par[1])
  lambda <- exp(opt$par[2])
  sigu <- sig * sqrt(lambda)
  N <- length(Y2)
  Siginv <- (1/sig^2) * solve(diag(rep(1,N)) + lambda * diag(1/D2) %*%
     tcrossprod(Z2), diag(1/D2))
  Beta <- ginv(crossprod(X2, Siginv %*% X2)) %*% crossprod(X2, Siginv %*% Y2)
  Q <- Siginv - Siginv %*% X2 %*% ginv(crossprod(X2, Siginv %*% X2)) %*%
     crossprod(X2, Siginv)
  predran <- diag(rep(sigu^2, ncol(Z2))) %*% crossprod(Z2, Q) %*% Y2

#### COMPUTING SEs FOR ESTIMATED FIXED EFFECTS ###
  dSigdsigu <- 2 * sigu * tcrossprod(Z2)
  dSigdsig <- 2*sig*diag(D2)
  dSig2dsigudsigu <-  2* tcrossprod(Z2)
  dSig2dsigdsig <- 2 * diag(D2)
  I11 <- crossprod(X2, Siginv %*% X2)
  I21 <- matrix(rep(0,ncol(X2)), nrow=1, ncol=ncol(X2))
  I31 <- matrix(rep(0,ncol(X2)), nrow=1, ncol=ncol(X2))
  I22 <- 1/2 * (sum(diag(Siginv %*% dSigdsigu %*% Siginv %*% dSigdsigu)))
  I32 <- 1/2 * (sum(diag(Siginv %*% dSigdsig %*% Siginv %*% dSigdsigu)))
  I33 <- 1/2 * (sum(diag(Siginv %*% dSigdsig %*% Siginv %*% dSigdsig)))
#Imat IS THE EXPECTED INFORMATION MATRIX
  Imat <- cbind(rbind(I11,I21,I31),rbind(t(I21),I22,I32),
     rbind(t(I31),I32,I33))
  cov.mat <- ginv(Imat)
#SE OF THE ESTIMATED FIXED EFFECTS (age effects, sigma, and sigma_u)
  fix.se <- sqrt(diag(cov.mat))
```

```
### SE OF PREDICTED RANDOM EFFECTS ###
  keep <- c(dim(Imat)[1] - 1, dim(Imat)[1])
#Btheta CONTAINS THE sigma and sigma_u PORTIONS OF THE INVERSE OF THE
#EXPECTED INFORMATION MATRIX
  Btheta <- cov.mat[keep,keep]
  prederr_infl <- vector()
  for (i in 1:length(predran)){
    colm <- i
    c1t <- 2 * sigu * t(Z2[,colm]) - sigu^2 * crossprod(Z2[,colm], Q) %*%
       dSigdsigu
    c2t <- -sigu^2 * crossprod(Z2[,colm], Q) %*% dSigdsig
    c <- rbind(c1t, c2t)
    Atheta <- c %*% Q %*% t(c)
    m1 <- sigu^2 - sigu^4 * crossprod(Z2[,colm], Q) %*% Z2[,colm]
    infl <- sum(diag(Atheta %*% Btheta))
    prederr_infl[i] <- sqrt(m1 + infl)
  }

  fix_sizeatt <- data.frame(est=rbind(Beta, sigu, sig), se=fix.se,
     row.names=c(colnames(X2), "sigma_u", "sigma"))
  rand_sizeatt <- data.frame(pred=predran, sepred=prederr_infl,
     row.names=colnames(Z2))

#FITTED FIXED EFFECTS
  XB <- X2 %*% Beta
#FITTED RANDOM EFFECTS
  Zu <- Z2 %*% predran
#STANDARDIZED CONDITIONAL RESIDUALS
  rs <- (Y2 - XB - Zu)/(sig*sqrt(D2))
  sqrtmkk <- sig^2*sqrt(diag(diag(D2)%*%Q%*%diag(D2)))
#STUDENTIZED CONDITIONAL RESIDUALS
  rstar <- (Y2 - XB - Zu)/(sig*sqrtmkk)
#OUTPUT EVERYTHING IN A LIST
  list(fix_sizeatt, rand_sizeatt, cov.mat, XB, Zu, rs, rstar, llike)
}
```