

ST 762, HOMEWORK 1 EXTRA PROBLEM SOLUTIONS, FALL 2009

1. (a) The loglikelihood is

$$\log L = \sum_{j=1}^n \{Y_j \log f(\mathbf{x}_j, \boldsymbol{\beta}) - f(\mathbf{x}_j, \boldsymbol{\beta}) - \log Y_j!\}.$$

Taking derivatives with respect to  $\boldsymbol{\beta}$  and setting equal to zero gives the estimating equation

$$\begin{aligned} \partial/\partial\boldsymbol{\beta} \log L &= \sum_{j=1}^n \{Y_j f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta})/f(\mathbf{x}_j, \boldsymbol{\beta}) - f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta})\} \\ &= \sum_{j=1}^n f^{-1}(\mathbf{x}_j, \boldsymbol{\beta}) \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta})\} f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) = \mathbf{0}. \end{aligned}$$

Note that, under the Poisson distribution,  $\text{var}(Y_j|\mathbf{x}_j) = f(\mathbf{x}_j, \boldsymbol{\beta})$ .

- (b) The loglikelihood is

$$\log L = \sum_{j=1}^n [Y_j \log f(\mathbf{x}_j, \boldsymbol{\beta}) + (k_j - Y_j) \log \{1 - f(\mathbf{x}_j, \boldsymbol{\beta})\}].$$

Thus, taking derivatives with respect to  $\boldsymbol{\beta}$  gives

$$\begin{aligned} \partial/\partial\boldsymbol{\beta} \log L &= \sum_{j=1}^n [Y_j f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta})/f(\mathbf{x}_j, \boldsymbol{\beta}) - (k_j - Y_j) f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta})/\{1 - f(\mathbf{x}_j, \boldsymbol{\beta})\}] \\ &= \sum_{j=1}^n [k_j f(\mathbf{x}_j, \boldsymbol{\beta}) \{1 - f(\mathbf{x}_j, \boldsymbol{\beta})\}]^{-1} \{Y_j - k_j f(\mathbf{x}_j, \boldsymbol{\beta})\} k_j f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) = \mathbf{0}. \end{aligned}$$

Note that, under the binomial distribution,  $\text{var}(Y_j|\mathbf{x}_j) = k_j f(\mathbf{x}_j, \boldsymbol{\beta}) \{1 - f(\mathbf{x}_j, \boldsymbol{\beta})\}$ .

(c) Both of these estimating equations are *linear* in the data  $Y_j$ . In addition, they both have a specific form, that of the GLS-type equation in (3.2) of the notes. That is, they have the form of a deviation (response–mean) times a gradient and a “weight” equal to the inverse of the variance of the response. This is no accident, as we will see: Both of these distributions are members of a special class with this property.

2. (a) The likelihood is

$$L = \prod_{j=1}^n \frac{Y_j^{1/\sigma^2 - 1} \exp[-Y_j/\{\sigma^2 f(\mathbf{x}_j, \boldsymbol{\beta})\}]}{\Gamma(1/\sigma^2) \{\sigma^2 f(\mathbf{x}_j, \boldsymbol{\beta})\}^{1/\sigma^2}},$$

so that

$$\log L = \sum_{j=1}^n [(1/\sigma^2 - 1) \log Y_j - Y_j/\{\sigma^2 f(\mathbf{x}_j, \boldsymbol{\beta})\} - (1/\sigma^2) \log \{\sigma^2 f(\mathbf{x}_j, \boldsymbol{\beta})\} - \log \Gamma(1/\sigma^2)].$$

Taking derivatives with respect to  $\boldsymbol{\beta}$  yields the estimating equation

$$\begin{aligned} \partial/\partial\boldsymbol{\beta} \log L &= \sum_{j=1}^n [Y_j f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta})/\{\sigma f(\mathbf{x}_j, \boldsymbol{\beta})\}^2 - (1/\sigma^2) f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta})/f(\mathbf{x}_j, \boldsymbol{\beta})] \\ &= (1/\sigma^2) \sum_{j=1}^n f^{-2}(\mathbf{x}_j, \boldsymbol{\beta}) \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta})\} f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) = \mathbf{0}. \end{aligned}$$

(b) Now we have

$$\log L = -n \log(2\pi)^{1/2} - n \log \sigma - \sum_{j=1}^n \log f(\mathbf{x}_j, \boldsymbol{\beta}) - \frac{1}{2\sigma^2} \sum_{j=1}^n \frac{[Y_j - f(\mathbf{x}_j, \boldsymbol{\beta})]^2}{f^2(\mathbf{x}_j, \boldsymbol{\beta})}.$$

Thus,

$$\begin{aligned} \partial/\partial\boldsymbol{\beta} \log L &= \sum_{j=1}^n \lambda_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) + (1/\sigma^2) \sum_{j=1}^n \frac{[Y_j - f(\mathbf{x}_j, \boldsymbol{\beta})]^2}{f^2(\mathbf{x}_j, \boldsymbol{\beta})} \lambda_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) \\ &\quad + (1/\sigma^2) \sum_{j=1}^n f^{-2}(\mathbf{x}_j, \boldsymbol{\beta}) \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta})\} f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) \\ &= (1/\sigma^2) \sum_{j=1}^n f^{-2}(\mathbf{x}_j, \boldsymbol{\beta}) \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta})\} f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) \\ &\quad + \sum_{j=1}^n \left( \frac{[Y_j - f(\mathbf{x}_j, \boldsymbol{\beta})]^2}{\sigma^2 f^2(\mathbf{x}_j, \boldsymbol{\beta})} - 1 \right) \lambda_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) = \mathbf{0}. \end{aligned}$$

(c) It is straightforward to derive the form of the lognormal density given only the information in the problem, which we do here. If  $Z = \log Y$ , then the Jacobian of the transformation is  $1/Y$ , and the density of  $Y$  is thus  $n(\log Y; m, \gamma^2)Y^{-1}$ , where  $n(\cdot; m, \gamma^2)$  is the normal density with mean  $m$  and variance  $\gamma^2$ . Thus, the desired density is

$$(2\pi)^{-1/2}(\gamma Y)^{-1} \exp \left\{ -\frac{(\log Y - m)^2}{2\gamma^2} \right\}.$$

We would like this in terms of the mean and variance of  $Y$ . If  $E(Y) = f$ , then using the moment generating function of a normal, we have

$$E(Y) = E(e^Z) = e^{m+\gamma^2/2} = f.$$

We also have  $E(Y^2) = E(e^{2Z}) = e^{2m+2\gamma^2}$ , so that

$$\text{var}(Y) = (e^{\gamma^2} - 1)\{E(Y)\}^2.$$

Thus,  $\sigma^2 = e^{\gamma^2} - 1$ , and we may deduce that

$$\gamma^2 = \log(\sigma^2 + 1), \quad m = \log f - \log\{(\sigma^2 + 1)/2\}.$$

Applying this to our problem and ignoring constants, we have

$$\log L = - \sum_{j=1}^n \log Y_j - (n/2) \sum_{j=1}^n \log\{\log(\sigma^2 + 1)\} - \sum_{j=1}^n \frac{[\log Y_j - \log f(\mathbf{x}_j, \boldsymbol{\beta}) + \log\{(\sigma^2 + 1)/2\}]^2}{\log(\sigma^2 + 1)} \frac{f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta})}{f(\mathbf{x}_j, \boldsymbol{\beta})}.$$

Taking derivatives with respect to  $\boldsymbol{\beta}$  and simplifying yields the estimating equation

$$\sum_{j=1}^n [\log Y_j - \log f(\mathbf{x}_j, \boldsymbol{\beta}) + \log\{(\sigma^2 + 1)/2\}] \frac{f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta})}{f(\mathbf{x}_j, \boldsymbol{\beta})} = \mathbf{0}.$$

(d) The equation in (a) is of exactly the same ‘‘GLS’’ form as those in Problem 1(a) and (b) – linear in the data with ‘‘weighting’’ by the inverse of the variance under the distributional

assumption. The equation in (b) is the sum of two terms, one of which is identical to that in 2(a). The other term is a *quadratic* function of the data. Thus, the estimating equation one solves for maximum likelihood estimation under the normal distribution is *not* linear, but rather is quadratic in the data. However, it is interesting that the first term is linear and of the same GLS form. The equation in (c) is also not *linear* in the data. Rather, it is linear in the *logarithm* of the data. Note that both (b) and (c) also depend on  $\sigma^2$  (the equation in (a) does not, as  $\sigma^2$  only enters in a multiplicative fashion).

(e) We have already noted the similarities and differences for (a) and (b) in (d). A possible reason for the difference, namely, the presence of the second, quadratic term in (b) is the following. The gamma distribution has the intrinsic property that the specific relationship between mean and variance is of the form variance  $\propto$  mean<sup>2</sup>. Thus, once the mean is known, so is the variance – there is nothing additional to be gained by being told the variance once the form of the mean is given. On the other hand, the mean and variance of a normal distribution need not obey any such relationship; the mean and variance determine the normal distribution, and they may be anything, and be related or unrelated. Thus, an interpretation is that the “extra” quadratic term in the estimating equation for  $\beta$  in (b) is that for the normal distribution the variance has additional information about  $\beta$  beyond that in the mean. In contrast, for the gamma distribution, all the information on  $\beta$  is in the mean, as it completely determines the form of the variance. We will discuss this phenomenon in gory detail later.

3. (a) Following the same argument as on page 37 of the notes, we have

$$\begin{aligned} Y_j &= h^{-1}[h\{f(\mathbf{x}_j, \beta), \lambda\} + q(\mathbf{x}_j, \theta)e_j, \lambda] \\ &\approx h^{-1}[h\{f(\mathbf{x}_j, \beta), \lambda\}, \lambda] + \{d/du h^{-1}(u, \lambda)\}_{u=h\{f(\mathbf{x}_j, \beta), \lambda\}} q(\mathbf{x}_j, \theta)(e_j - 0) \\ &\approx f(\mathbf{x}_j, \beta) + \{d/du h^{-1}(u, \lambda)\}_{u=h\{f(\mathbf{x}_j, \beta), \lambda\}} q(\mathbf{x}_j, \theta)e_j \\ &\approx f(\mathbf{x}_j, \beta) + q(\mathbf{x}_j, \theta)f^{1-\lambda}(\mathbf{x}_j, \beta)e_j. \end{aligned}$$

Thus, we have approximately that

$$E(Y_j|\mathbf{x}_j) = f(\mathbf{x}_j, \beta), \quad \text{var}(Y_j|\mathbf{x}_j) = \sigma^2 q^2(\mathbf{x}_j, \theta) f^{2(1-\lambda)}(\mathbf{x}_j, \beta).$$

(b) The issue is whether we can obtain  $\text{var}(Y_j|\mathbf{x}_j) = \sigma^2[\theta_1 + f(\mathbf{x}_j, \beta)^{2\theta_2}]$ . It does not look promising; if we identify  $\theta_2 = 1 - \lambda$ , we still cannot achieve an additive form, as  $f(\mathbf{x}_j, \beta)^{2(1-\lambda)}$  multiplies  $q(\mathbf{x}_j, \theta)$  and does not depend on  $f$ . Thus, although this generalization of the TBS model accommodates a “fancier” implied (approximate) variance on the original scale, it is still not flexible enough to represent the variance model in question. The moral: it is not necessarily the case the all variance heterogeneity may be handled by transformation, even a “flexible” transformation.

(c) The usual model with  $q \equiv 1$  is

$$h\{Y, \lambda\} = h\{f(\mathbf{x}, \beta), \lambda\} + \sigma e.$$

Under the Box-Cox transformation

$$h(y, \lambda) = \frac{y^\lambda - 1}{\lambda} \implies y = \{\lambda h(y, \lambda) + 1\}^{1/\lambda}.$$

If  $y > 0$ , this implies that the transformed version must satisfy  $h(y, \lambda) > -1/\lambda$ . If  $h(Y, \lambda)$  is supposed to be normal, however, it must be able to take on values on the entire real line.

Only when  $\lambda = 0$  (the log transformation) will this be the case; otherwise, this is not possible. The implication is that, technically, for positive response, it is impossible for the TBS model to hold! However, if  $P\{h(Y, \lambda) > -1/\lambda\}$  is close to 1, for all practical purposes we may ignore this technical detail in applications. This explains why this model has been successfully and widely used even with positive response on the original scale.

4. (a) For (i), note immediately that

$$\frac{1}{(\beta_0 + \beta_1/x)^{-1}} = \beta_0 + \beta_1/x.$$

Thus, we may write this model alternatively as

$$\frac{Y^{-1} - 1}{1} = \frac{f^{-1}(x, \boldsymbol{\beta}) - 1}{1} + e,$$

with  $f(x, \boldsymbol{\beta}) = (\beta_0 + \beta_1/x)^{-1}$ . It follows that this is of the form in 3(a) with  $\lambda = -1$ ,  $\theta = 0$ , so that  $E(Y_j|x_j) = f(x_j, \boldsymbol{\beta})$  and  $\text{var}(Y_j|x_j) = \sigma^2 f^4(x_j, \boldsymbol{\beta})$ . Thus, this model makes an approximate assumption that the variance increases drastically with the mean.

For (ii), using the above, we may immediately identify  $\lambda = -1$  and  $\theta = -1$ , so that  $E(Y_j|x_j) = f(x_j, \boldsymbol{\beta})$  and  $\text{var}(Y_j|x_j) = \sigma^2 x_j^{-2} f^4(x_j, \boldsymbol{\beta})$ . Thus, this model perhaps makes a less severe approximate assumption on variance, as it “tempers” the power of the mean with the inverse of the square of  $x_j$ .

(b) By some algebra, we can write this model as

$$Y = (\beta_0 + \beta_1/x)^{-1}(1 + \beta_1 e).$$

Thus, we see that this model makes the assumption that  $E(Y_j|x_j) = (\beta_0 + \beta_1/x)^{-1}$  and  $\text{var}(Y_j|x_j) = \sigma^2 \beta_1^2 \{E(Y_j|x_j)\}^2$ , so that the model does assume constant coefficient of variation, where the coefficient of variation is  $\sigma \beta_1$ .

Note further that

$$\log Y = \log(\beta_0 + \beta_1/x)^{-1} + \log(1 + \beta_1 e) \approx \log(\beta_0 + \beta_1/x)^{-1} + \beta_1 e$$

by a Taylor series in the second term about  $e = 0$ . Thus, this model is of the form in part (a) with  $\lambda = \theta = 0$  and “errors” with variance  $\sigma \beta_1$ .

5. We have

$$\begin{aligned} f(\mathbf{x}_j, \boldsymbol{\beta}) &\approx f(\mathbf{x}_j, \boldsymbol{\beta}^*) + f_{\boldsymbol{\beta}}^T(\mathbf{x}_j, \boldsymbol{\beta}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}^*), \\ f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) &\approx f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}^*) + \mathbf{f}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}^*), \\ g^{-2}(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{x}_j) &\approx g^{-2}(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j) + G^T(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j)(\boldsymbol{\beta} - \boldsymbol{\beta}^*), \end{aligned}$$

where  $G(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{x}_j) = \partial/\partial \boldsymbol{\beta} g^{-2}(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{x}_j)$ , say (a  $(q \times 1)$  vector).

Substituting these approximations into the GLS equation (3.12) on page 59 gives

$$\begin{aligned} \mathbf{0} &\approx \sum_{j=1}^n \{g^{-2}(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j) + G^T(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j)(\boldsymbol{\beta} - \boldsymbol{\beta}^*)\} \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta}^*) - f_{\boldsymbol{\beta}}^T(\mathbf{x}_j, \boldsymbol{\beta}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}^*)\} \\ &\quad \times \{f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}^*) + \mathbf{f}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}^*)\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^n g^{-2}(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j) \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta}^*)\} f_{\beta}(\mathbf{x}_j, \boldsymbol{\beta}^*) \\
&\quad - \sum_{j=1}^n g^{-2}(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j) f_{\beta}(\mathbf{x}_j, \boldsymbol{\beta}^*) f_{\beta}^T(\mathbf{x}_j, \boldsymbol{\beta}^*) (\boldsymbol{\beta} - \boldsymbol{\beta}^*) \\
&\quad + \sum_{j=1}^n g^{-2}(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j) \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta}^*)\} \mathbf{f}_{\beta\beta}(\mathbf{x}_j, \boldsymbol{\beta}^*) (\boldsymbol{\beta} - \boldsymbol{\beta}^*) \\
&\quad + \sum_{j=1}^n g^{-2}(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j) f_{\beta}^T(\mathbf{x}_j, \boldsymbol{\beta}^*) (\boldsymbol{\beta} - \boldsymbol{\beta}^*) \mathbf{f}_{\beta\beta}(\mathbf{x}_j, \boldsymbol{\beta}^*) (\mathbf{x}_j, \boldsymbol{\beta}^*) \\
&\quad + \sum_{j=1}^n G^T(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j) (\boldsymbol{\beta} - \boldsymbol{\beta}^*) \times \text{terms involving } \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta}^*)\}, (\boldsymbol{\beta} - \boldsymbol{\beta}^*).
\end{aligned}$$

The first two terms involve the data and are linear in  $(\boldsymbol{\beta} - \boldsymbol{\beta}^*)$ . The third depends on the product of  $\{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta}^*)\}$  and  $(\boldsymbol{\beta} - \boldsymbol{\beta}^*)$ , which is expected to be “smaller” for  $\boldsymbol{\beta}^*$  close to  $\boldsymbol{\beta}$  than the first two terms. The fourth term is quadratic in  $(\boldsymbol{\beta} - \boldsymbol{\beta}^*)$ , so should also be “smaller” than the first two. The remaining terms involve *at least* products of  $\{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta}^*)\}$  and  $(\boldsymbol{\beta} - \boldsymbol{\beta}^*)$ , so are also “smaller.”

Thus, as in the argument in Section 3.2, we disregard these terms. Note that the presence of  $\boldsymbol{\beta}$  in the “weights” really doesn’t affect the form of the linear approximation at  $\boldsymbol{\beta}^*$ . We are left with

$$\sum_{j=1}^n g^{-2}(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j) \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta}^*)\} f_{\beta}(\mathbf{x}_j, \boldsymbol{\beta}^*) \approx \sum_{j=1}^n g^{-2}(\boldsymbol{\beta}^*, \boldsymbol{\theta}, \mathbf{x}_j) f_{\beta}(\mathbf{x}_j, \boldsymbol{\beta}^*) f_{\beta}^T(\mathbf{x}_j, \boldsymbol{\beta}^*) (\boldsymbol{\beta} - \boldsymbol{\beta}^*),$$

which may be rewritten in obvious matrix notation as

$$\{\mathbf{X}^T(\boldsymbol{\beta}^*) \mathbf{W}(\boldsymbol{\beta}^*) \mathbf{X}(\boldsymbol{\beta}^*)\} (\boldsymbol{\beta} - \boldsymbol{\beta}^*) \approx \mathbf{X}^T(\boldsymbol{\beta}^*) \mathbf{W}(\boldsymbol{\beta}^*) \{\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta}^*)\},$$

yielding the required updating scheme.

6. (a) If you plot the data you will notice that there are two distinct phases of decay. It turns out that my algorithm uses the final 6 observations to fit the “second phase.” The remaining 5 observations are used to fit the “first phase.”

The method I used is based on the hint. I start with the last 3 observations (the farthest out in time) and fit a straight line to them by simple linear regression, using  $\log Y$  as the response. This is based on the fact that, in this region,  $Y \approx \beta_3 e^{-\beta_4 x}$ , or  $\log Y \approx \log \beta_3 - \beta_4 x = \beta_3^* + \beta_4^* x$ , say. Thus, I obtain estimates for  $\beta_3^*$  and  $\beta_4^*$ . I then construct a 90% prediction interval for  $\log Y$  at  $x_0$  corresponding to the time of the next observation (backward in time). For this, I use the standard prediction interval formula for simple linear regression based on the fit with the  $n = 3$  final values. That is, the estimated standard deviation of the prediction error at  $x_0$  is

$$\hat{\sigma} \left\{ 1 + n^{-1} \frac{(x_0 - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \right\} = SD,$$

where  $\bar{x}$  is the mean of the  $n$   $x$  values involved in the fit and  $\hat{\sigma}^2$  is the usual estimator for constant variance in simple linear regression based on the  $n$  observations. I obtain the 90% prediction interval as

$$(\hat{\beta}_3^* + \hat{\beta}_4^* x_0) \pm t_{0.95} SD,$$

where  $t_{0.95}$  is the  $t$  critical value with  $(n - 2)$  degrees of freedom with 0.95 area to its right. I then check whether  $\log Y$  corresponding to  $x_0$  is contained in the interval. If yes, repeat this process with  $n = 4$ , including this observation in the simple linear regression fit. If not, stop, and declare the current estimates  $\exp(\hat{\beta}_3^*)$  and  $-\hat{\beta}_4^*$  to be starting values for  $\beta_3$  and  $\beta_4$ . Now, I note that the model implies that  $Y - \beta_3 e^{-\beta_4 x} \approx \beta_1 e^{-\beta_2 x}$ . This suggests that

$$\log(Y - \beta_3 e^{-\beta_4 x}) \approx \log \beta_1 - \beta_2 x = \beta_1^* + \beta_2^* x.$$

I thus form “residuals”  $Y_j - e^{\hat{\beta}_3^* + \hat{\beta}_4^* x} = r_j$  for all observations not included in the final second phase fit (thus thought to be in the first phase) and regress  $\log(r_j)$  on  $x_j$  in the first phase to obtain estimates for  $\beta_1^*$  and  $\beta_2^*$  and hence for  $\beta_1$  and  $\beta_2$ .

The starting values are thus  $(e^{\hat{\beta}_1^*}, -\hat{\beta}_2^*, e^{\hat{\beta}_3^*}, -\hat{\beta}_4^*)$ . This is implemented in the program.

(b) For these data, I obtained start values (2.608, 2.743, 0.309, 0.310). Using these, I obtained the IRWLS estimate for  $\beta$  within 7 iterations. Comparing these values to the final estimates shows that this *ad hoc* method did a pretty good job of producing starting values that are “in the ballpark.”