ST 762, HOMEWORK 3 EXTRA PROBLEMS SOLUTIONS, FALL 2009

1. By a conditioning argument, the cdf of X is $F(x) = \Phi_1(x)(1-\alpha) + \Phi_b(x)\alpha$, where

$$\Phi_b(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi b^2}} e^{-u^2/(2b^2)}\, du,$$

where the integrand is the normal density $\phi_b(x)$ with mean zero and variance b^2. Thus, the density is $f(x) = \phi_1(x)(1-\alpha) + \phi_b(x)\alpha$. It is straightforward to see that $E(X) = 0$, $E(X^2) = (1-\alpha) + \alpha b^2$, $E(X^3) = 0$, and

$$E(X^4) = (1-\alpha)\int_{-\infty}^{\infty} x^4\phi_1(x)\, dx + \alpha \int_{-\infty}^{\infty} x^4 \phi_b(x)\, dx,$$

which, by a change of variables in the second integral and using Problem 2(a), reduces to $E(X^4) = 3(1-\alpha) + 3b^4\alpha = 3\{(1-\alpha) + \alpha b^4\}$. Thus, for $\epsilon = X\{(1-\alpha)+\alpha b^2\}^{-1/2}$, $E(\epsilon) = 0$, $E(\epsilon^2) = 1$, $E(\epsilon^3) = 0$, and

$$\text{var}(\epsilon^2) = E(\epsilon^4) - 1 = \frac{3\{(1-\alpha)+\alpha b^4\}}{\{(1-\alpha)+\alpha b^2\}} - 1 = 2 + \kappa,$$

so that

$$\kappa = 3\left\{\frac{1-\alpha+\alpha b^4}{(1-\alpha+\alpha b^2)^2} - 1\right\}.$$

(b) Substitution with $\alpha = 0.01$ and $\alpha = 0.05$ gives the following values of κ

b	$\alpha = 0.01$	$\alpha = 0.05$
2	0.252	0.970
3	1.630	4.653
4	5.053	10.469

Thus, for $b = 3$, $\alpha = 0.01$, for instance, $\text{var}(\epsilon^2) = 3.630$ vs. 2 for the normal. The excess kurtosis increases fairly quickly. Even for a very small proportion of outliers like 0.01, the excess kurtosis can depart considerably from that of the normal. This suggests that deviation from normality caused by outliers may be important when estimating $\boldsymbol{\beta}$ by quadratic estimating equations, as we will see later.

2. (a) One way to do this is to use the moment generating function of the standard normal, given by $m(t) = e^{t^2/2}$. Taking derivatives gives $m'(t) = te^{t^2/2}$, $m''(t) = e^{t^2/2} + t^2e^{t^2/2}$, $m'''(t) = 3te^{t^2/2} + t^3e^{t^2/2}$, and $m^{iv} = 3e^{t^2/2} + 6t^2e^{t^2/2} + t^4e^{t^2/2}$. Thus, $E(\epsilon^3) = m'''(0) = 0$ and $E(\epsilon^4) = m^{iv}(0 = 3$, so that $\text{var}(\epsilon^2) = E(\epsilon^4) - \{E(\epsilon^2)\}^2 = 3 - 1 = 2$, and $\kappa = 0$.

(b) The gamma is a scaled exponential family random variable, so we can use the results in Homework 2 Extra Problems, Problem 1, as $E(\epsilon^3) = E\{(Y-\mu)^3/(\sigma^2\mu^2)^{3/2}\} = m_3/m^{3/2} = \zeta$ and $\text{var}(\epsilon^2) = E(\epsilon^4) - 1 = E\{(Y-\mu)^4/(\sigma^2\mu^2)^2\} - 1 = m_4/m_2^2 - 1$, so that $\kappa = m_4/m_2^2 - 3$. We wish to find these values for the gamma. Now from Homework 2 Extra Problems, Problem 1, recall that

$$\zeta = \sigma b_{\xi\xi\xi}(\xi)/\{b_{\xi\xi}(\xi)\}^{3/2},$$
$$\kappa = \sigma^2 b_{\xi\xi\xi\xi}(\xi)/\{b_{\xi\xi}(\xi)\}^2.$$

For the gamma, $b_(\xi) = -\log(-\xi)$, $b_\xi(\xi) = -1/\xi$, $b_{\xi\xi}(\xi) = 1/xi^2$, $b_{\xi\xi\xi}(\xi) = -2/\xi^3$, and $b_{\xi\xi\xi\xi}(\xi) = 6/\xi^4$. As $\xi = -1/\mu$, we obtain $\zeta = 2\sigma$ and $\kappa = 6\sigma^2$. Note that the skewness and excess kurtosis are of orders σ and σ^2, a fact that will be of interest to us later.

(c) We are interested in finding the coefficient of skewness ζ and the excess kurtosis κ for a binomial random variable with mean kp and variance $kp(1-p)$. We can use the same approach as in (b) here, as this is a scaled exponential family; in this case, it is easy to show that $b(\xi) = k\log(1+e^\xi)$, so that $b_\xi(\xi) = ke^\xi/(1+e^\xi) = kp$, $b_{\xi\xi}(\xi) = ke^\xi/(1+e^\xi)\{1-e^\xi/(1+e^\xi)\} = kp(1-p)$,

$$b_{\xi\xi\xi}(\xi) = \frac{ke^\xi}{1+e^\xi} - \frac{3ke^{2\xi}}{(1+e^\xi)^2} + \frac{2ke^{3\xi}}{(1+e^\xi)^3} = kp(1-p)(1-2p).$$

$$b_{\xi\xi\xi\xi}(\xi) = \frac{ke^\xi}{1+e^\xi} - \frac{7ke^{2\xi}}{(1+e^\xi)^2} + \frac{12ke^{3\xi}}{(1+e^\xi)^3} - \frac{6ke^{4\xi}}{(1+e^\xi)^4} = kp(1-p)(1-6p+6p^2).$$

Thus, substituting into the expressions for ζ and κ, we obtain

$$\zeta = \frac{1-2p}{\{kp(1-p)\}^{1/2}}, \quad \kappa = \frac{1-6p+6p^2}{kp(1-p)}.$$

Note here that the skewness and excess kurtosis in fact *depend* on the mean kp. This shows that, for regression models and this distribution, the third and fourth moments of $\epsilon_j|\boldsymbol{x}_j$ must depend on the mean; hence, the assumption that the ϵ_j are independent of the $\boldsymbol{x}_j$ would be inappropriate.

3. (a) If we define as in the notes

$$\boldsymbol{s}_j = \begin{pmatrix} Y_j \\ (Y_j - f_j)^2 \end{pmatrix}, \quad \boldsymbol{m}_j = \begin{pmatrix} f_j \\ \sigma^2 g_j^2 \end{pmatrix},$$

using obvious shorthand notation, then we have

$$\boldsymbol{D}_j = \begin{pmatrix} f_{\boldsymbol{\beta}j} & 2\sigma^2 g_j^2 \nu_{\beta j} \\ 0 & 2\sigma g_j^2 \end{pmatrix}$$

as in the notes. However, now we have

$$\boldsymbol{V}_j = \begin{pmatrix} \sigma^2 g_j^2 & \zeta_j \sigma^3 g_j^3 \\ \zeta_j \sigma^3 g_j^3 & (2+\kappa_j)\sigma^4 g_j^4 \end{pmatrix}.$$

Now

$$\boldsymbol{V}_j^{-1} = \frac{1}{|\boldsymbol{V}_j|} \begin{pmatrix} (2+\kappa_j)\sigma^4 g_j^4 & -\zeta_j \sigma^3 g_j^3 \\ -\zeta_j \sigma^3 g_j^3 & \sigma^2 g_j^2 \end{pmatrix},$$
$$|\boldsymbol{V}_j| = \sigma^6 g_j^6 (2+\kappa_j - \zeta_j^2).$$

Thus, the equation may be written as

$$\sum_{j=1}^n \begin{pmatrix} f_{\beta j} & 2\sigma^2 g_j^2 \nu_{\beta j} \\ 0 & 2\sigma g_j^2 \end{pmatrix} \frac{1}{\sigma^6 g_j^6 (2+\kappa_j - \zeta_j^2)} \begin{pmatrix} (2+\kappa_j)\sigma^4 g_j^4 & -\zeta_j \sigma^3 g_j^3 \\ -\zeta_j \sigma^3 g_j^3 & \sigma^2 g_j^2 \end{pmatrix} \begin{pmatrix} Y_j - f_j \\ (Y_j - f_j)^2 - \sigma^2 g_j^2 \end{pmatrix} = \mathbf{0}.$$

Multiplying out and simplifying, we eventually arrive at the joint estimating equations

$$\sum_{j=1}^n \frac{(2+\kappa_j) f_{\beta j}/(\sigma g_j) - 2\nu_{\beta j}\zeta_j}{\sigma g_j (2+\kappa_j - \zeta_j^2)}(Y_j - f_j) + \frac{2\nu_{\beta j} - \zeta_j f_{\beta j}/(\sigma g_j)}{\sigma^2 g_j^2 (2+\kappa_j - \zeta_j^2)}\{(Y_j - f_j)^2 - \sigma^2 g_j^2\} = 0,$$

$$\sum_{j=1}^{n} \frac{-2\zeta_j}{\sigma^2 g_j^2(2+\kappa_j-\zeta_j^2)}(Y_j - f_j) + \frac{2}{\sigma^3 g_j^3(2+\kappa_j-\zeta_j^2)}\{(Y_j - f_j)^2 - \sigma^2 g_j^2\} = 0.$$

Thus, we may identify

$$\boldsymbol{a}_j = \frac{1}{\sigma g_j(2+\kappa_j-\zeta_j^2)} \begin{pmatrix} (2+\kappa_j) f_{\beta j}/(\sigma g_j) - 2\nu_{\beta j}\zeta_j \\ -2\zeta_j/\sigma \end{pmatrix},$$

$$\boldsymbol{c}_j = \frac{1}{\sigma^2 g_j^2(2+\kappa_j-\zeta_j^2)} \begin{pmatrix} -\zeta_j f_{\beta j}/(\sigma g_j) + 2\nu_{\beta j} \\ 2/\sigma \end{pmatrix}.$$

(b) Here, $\sigma = 1$. Write $p_j = p(\boldsymbol{x}_j, \boldsymbol{\beta})$ for short. From Problem 2(c), we know that $\zeta_j = (1-2p_j)/\{k_j p_j(1-p_j)\}^{1/2}$ and $\kappa_j = (1-6p_j+6p_j^2)/\{k_j p_j(1-p_j)\}$. We also have $f_j = k_j p_j$ and $g_j = \{k_j p_j(1-p_j)\}^{1/2}$. Using the fact that p_j has the form of a logistic regression model, it is straightforward to show that $f_{\beta j} = k_j p_j(1-p_j)\boldsymbol{x}_j$ and $\nu_{\beta j} = (1-2p_j)\boldsymbol{x}_j/2$. Substituting yields $(2+\kappa_j-\zeta_j) = 2(k_j-1)/k_j$. The first p rows of $\boldsymbol{c}_j$ depend on

$$2\nu_{\beta j} - \zeta_j f_{\beta j}/g_j = (1-2p_j)\boldsymbol{x}_j - \frac{(1-2p_j)}{\{k_j p_j(1-p_j)\}^{1/2}} \frac{k_j p_j(1-p_j)\boldsymbol{x}_j}{\{k_j p_j(1-p_j)\}^{1/2}} = \boldsymbol{0};$$

thus, the quadratic term vanishes. Similarly, for the linear term, the first p rows depend on

$$(2+\kappa_j) f_{\beta j}/g_j - 2\nu_{\beta j}\zeta_j = \left(\frac{2(k_j-1}{k_j}\right) \frac{k_j p_j(1-p_j)\boldsymbol{x}_j}{\{k_j p_j(1-p_j)\}^{1/2}}$$

after some algebra, so that the first p rows of $\boldsymbol{a}_j$ become

$$\left(\frac{2(k_j-1}{k_j}\right)\left(\frac{2(k_j-1}{k_j}\right)^{-1} \frac{k_j p_j(1-p_j)\boldsymbol{x}_j}{k_j p_j(1-p_j)} = \frac{f_{\beta j}}{g_j^2} = \boldsymbol{x}_j.$$

Thus, the first p rows of the quadratic estimating equation for estimation of $\boldsymbol{\beta}$ in this model reduce to

$$\sum_{j=1}^{n}(Y_j - f_j)\boldsymbol{x}_j = \sum_{j=1}^{n} g_j^{-2}(Y_j - f_j) f_{\beta j};$$

the first expression shows these reduce to just the usual maximum likelihood estimating equations for logistic regression; the second expression shows furthermore that these equations are linear in Y_j and have the form of the GLS equations.

(c) We have $\sigma = 1$, $\zeta_j = f_j^{-1/2}$, and $\kappa_j = f_j^{-1}$. Thus, $\zeta_j = g_j^{-1}$, $(2+\kappa_j-\zeta_j^2) = 2+f_j^{-1}-f_j^{-1} = 2$, $2\nu_{\beta j} = f_{\beta j}/f_j$, and then $\boldsymbol{a}_j$ and $\boldsymbol{c}_j$ simplify to

$$\boldsymbol{a}_j = \begin{pmatrix} f_{\beta j}/f_j \\ -1/f_j \end{pmatrix}, \quad \boldsymbol{c}_j = \begin{pmatrix} 0 \\ 1/f_j \end{pmatrix}.$$

The first equation, which corresponds to $\boldsymbol{\beta}$, thus reduces to

$$\sum_{j=1}^{n} f_j^{-1}(Y_j - f_j) f_{\beta j} = \boldsymbol{0}.$$

(d) Under these conditions, we have $\sigma = 0.5$, $g_j = f_j$, $\zeta_j = 1$, and $\kappa_j = 1.5$. Then $\nu_{\beta j} = f_{\beta j}/f_j$, so that $\boldsymbol{a}_j$ and $\boldsymbol{c}_j$ simplify to

$$\boldsymbol{a}_j = \begin{pmatrix} 4f_{\beta j}/f_j^2 \\ -16/5 \end{pmatrix}, \quad \boldsymbol{c}_j = \begin{pmatrix} 0 \\ 32/5j \end{pmatrix}.$$

The first equation, which corresponds to $\boldsymbol{\beta}$, thus reduces to

$$4\sum_{j=1}^{n} f_j^{-1}(Y_j - f_j)f_{\beta j} = \mathbf{0},$$

where we may ignore the constant "4," where this equation is linear in Y_j.

(e) We see that, in all these cases, the quadratic estimation equation for $\boldsymbol{\beta}$ reduces to the usual linear maximum likelihood equations, which have the GLS form. The binomial (b), gamma (c), and Poisson (d) distributions are members of the scaled exponential family class, so we know that the maximum likelihood estimating equations for $\boldsymbol{\beta}$ must have this property. For distributions in this class, the variance is completely determined by the mean. Thus, an obvious explanation for the way that this turned out is that trying to gain information on $\boldsymbol{\beta}$ from the variance via a quadratic equation is pointless when the first four moments of the data, which are used in forming the quadratic equation, are the same as those of a distribution in the scaled exponential family class (the binomial here). We can't improve on the maximum likelihood estimator even if we try to do this. We will discuss this more later in the course.

4. Expanding (1) about $\boldsymbol{\alpha}^*$ "close to" $\boldsymbol{\alpha}$ gives

$$\begin{aligned}\mathbf{0} \;\approx\; & \sum_{j=1}^{n} \boldsymbol{D}_j^T(\boldsymbol{\alpha}^*)\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha}^*)\{\boldsymbol{s}_j(\boldsymbol{\alpha}^*) - \boldsymbol{m}_j(\boldsymbol{\alpha}^*)\} \\ & + \left[\sum_{j=1}^{n} \partial/\partial\boldsymbol{\alpha}\,\{\boldsymbol{D}_j^T(\boldsymbol{\alpha}^*)\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha}^*)\}\{\boldsymbol{s}_j(\boldsymbol{\alpha}^*) - \boldsymbol{m}_j(\boldsymbol{\alpha}^*)\}\right. \\ & \left. + \sum_{j=1}^{n} \boldsymbol{D}_j^T(\boldsymbol{\alpha}^*)\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha}^*)\partial/\partial\boldsymbol{\alpha}\,\{\boldsymbol{s}_j(\boldsymbol{\alpha}^*) - \boldsymbol{m}_j(\boldsymbol{\alpha}^*)\}\right](\boldsymbol{\alpha} - \boldsymbol{\alpha}^*).\end{aligned}$$

Now $\partial/\partial\boldsymbol{\alpha}\,\boldsymbol{s}_j(\boldsymbol{\alpha}) = \partial/\partial\boldsymbol{\alpha}\boldsymbol{s}_j[\{Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta})\}^2]$, where this notation is meant to indicate that $\boldsymbol{s}_j(\boldsymbol{\alpha})$ is a function of $\boldsymbol{\alpha}$ only through this argument. By the chain rule, letting "$'$" denote differentiation with respect to the argument, we have

$$\begin{aligned}& \boldsymbol{s}_j'[\{Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta})\}^2]\,\partial/\partial\boldsymbol{\alpha},[\{Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta})\}^2] \\ &= -2\{Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta}^*)\}\partial/\partial\boldsymbol{z}\,\boldsymbol{s}_j[\{Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta}^*)\}^2]\partial/\partial\boldsymbol{\alpha}\,\{f(\boldsymbol{x}_j,\boldsymbol{\beta}^*)\}.\end{aligned}$$

Substituting this into the above and using the definition of $\boldsymbol{D}_j(\boldsymbol{\alpha})$, we obtain

$$\begin{aligned}\mathbf{0} \;\approx\; & \sum_{j=1}^{n} \boldsymbol{D}_j^T(\boldsymbol{\alpha}^*)\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha}^*)\{\boldsymbol{s}_j(\boldsymbol{\alpha}^*) - \boldsymbol{m}_j(\boldsymbol{\alpha}^*)\} \\ & + \sum_{j=1}^{n} \partial/\partial\boldsymbol{\alpha}\,\{\boldsymbol{D}_j^T(\boldsymbol{\alpha}^*)\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha}^*)\}\{\boldsymbol{s}_j(\boldsymbol{\alpha}^*) - \boldsymbol{m}_j(\boldsymbol{\alpha}^*)\}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\end{aligned}$$

$$+\sum_{j=1}^{n} \boldsymbol{D}_j^T(\boldsymbol{\alpha}^*)\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha}^*)\boldsymbol{D}_j(\boldsymbol{\alpha}^*)(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)$$

$$+-2\sum_{j=1}^{n}\{Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta}^*)\}\boldsymbol{D}_j^T(\boldsymbol{\alpha}^*)\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha}^*)\boldsymbol{s}_j'(\boldsymbol{\alpha}^*)\partial/\partial\boldsymbol{\alpha}\,\{f(\boldsymbol{x}_j,\boldsymbol{\beta}^*)\}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*).$$

The term on the second line is negligible by virtue of the fact that $E(\boldsymbol{s}_j(\boldsymbol{\alpha}^*)|\boldsymbol{x}_j) \approx \boldsymbol{m}_j(\boldsymbol{\alpha}^*)$, so that this term contains the product of two "small" terms. The term in the fourth line depends on the product $(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)\{Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta}^*)\}$, which is also "small." Disregarding these terms yields the approximation

$$\begin{aligned}\boldsymbol{0} \;\approx\; & \sum_{j=1}^{n} \boldsymbol{D}_j^T(\boldsymbol{\alpha}^*)\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha}^*)\{\boldsymbol{s}_j(\boldsymbol{\alpha}^*) - \boldsymbol{m}_j(\boldsymbol{\alpha}^*)\} \\ & +\sum_{j=1}^{n} \boldsymbol{D}_j^T(\boldsymbol{\alpha}^*)\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha}^*)\boldsymbol{D}_j(\boldsymbol{\alpha}^*)(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*),\end{aligned}$$

which may be rearranged to give (6.14). The rest of the problem follows by the same manipulations as in the notes.

5. (a) First, we need to determine γ. Note that γ must satisfy

$$\begin{aligned} E(\epsilon^2) &= \frac{1}{2\gamma}\int_{-\infty}^{\infty} \epsilon^2 e^{-|\epsilon|/\gamma} d\epsilon \\ &= \frac{1}{\gamma}\int_{0}^{\infty} \epsilon^2 e^{-\epsilon/\gamma} d\epsilon \\ &= 2\gamma^2 \end{aligned}$$

either by brute-force integration by parts or by noting that this is the second moment of an exponential distribution. To ensure $E(\epsilon^2) = 1$, then, we obtain that $\gamma = 1/\sqrt{2}$.

We may now write down the form of the density of $Y_j|\boldsymbol{x}_j$; the joint density will be the product. We have for this choice of γ

$$f(\epsilon) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}\epsilon},$$

and $\epsilon = (Y-f)/(\sigma g)$ in shorthand, so that the density of Y (given $\boldsymbol{x}$, suppressed here) has the form

$$f_Y(y) = f\{(Y-f)/(\sigma g)\}\frac{1}{\sigma g}.$$

Thus, the joint density of interest is

$$\prod_{j=1}^{n} \frac{1}{\sqrt{2}\sigma g(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)} \exp\left\{-\sqrt{2}\frac{|Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta})|}{\sigma g(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)}\right\}.$$

(b) First, we need $E(|\epsilon|)$ for the double exponential. From (a), this is, with $\gamma = 1/\sqrt{2}$,

$$E(|\epsilon|) = \frac{1}{2\gamma}\int_{-\infty}^{\infty} |\epsilon| e^{-|\epsilon|/\gamma} d\epsilon = \frac{2}{\sqrt{2}}\int_{0}^{\infty} u e^{-\sqrt{2}u} du = 1/\sqrt{2}.$$

Then we have that $e^\eta = \sigma/\sqrt{2}$. Thus, substituting this in the joint density in (a), we may rewrite the joint density as

$$L = \prod_{j=1}^{n} \frac{1}{e^\eta g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)} \exp\left\{ -\frac{|Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})|}{e^\eta g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)} \right\}.$$

Treating this as a likelihood and taking logs and then differentiating with respect to η and $\boldsymbol{\theta}$, we obtain

$$\log L = -n \log e^\eta - \sum_{j=1}^{n} \log g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) - \sum_{j=1}^{n} \frac{|Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})|}{e^\eta g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)}, \tag{1}$$

$$\partial/\partial\eta \; \log L = -n + \sum_{j=1}^{n} \frac{|Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})|}{e^\eta g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)} = 0,$$

$$\partial/\partial\boldsymbol{\theta} \; \log L - \sum_{j=1}^{n} \nu_\theta(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) + \sum_{j=1}^{n} \frac{|Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})|}{e^\eta g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)} \nu_\theta(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) = \mathbf{0}.$$

It is straightforward to verify that these may be rewritten as (using shorthand),

$$\sum_{j=1}^{n} \left(\frac{|Y_j - f_j| - e^\eta g_j}{e^{2\eta} g_j^2} \right) e^\eta g_j \tau_{\theta j} = \mathbf{0},$$

which is (6.28) with $\lambda = 1$ up to a multiplicative constant.

(c) We need to come up with a "likelihood" for general $\lambda > 0$ such that differentiation leads to equation (6.28). To do this, it is reasonable to try to emulate the features of the normal loglikelihood ($\lambda = 2$, on page 131 in equation (6.17) and that in (1) for $\lambda = 1$ in (b). Define

$$L = \prod_{j=1}^{n} \{e^{\lambda\eta} g^\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)\}^{-1/\lambda} \exp\left\{ -\frac{|Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})|^\lambda}{\lambda e^{\lambda\eta} g^\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)} \right\}.$$

Note that this coincides with the "important part" of the normal or exponential joint density for $\lambda = 2, 1$. Thus,

$$\log L = -(n/\lambda) \log e^{\lambda\eta} - (1/\lambda) \sum_{j=1}^{n} \log g^\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) - (1/\lambda) \sum_{j=1}^{n} \frac{|Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})|^\lambda}{e^{\lambda\eta} g^\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)},$$

which may be simplified to

$$\log L = -n \log \eta - \sum_{j=1}^{n} \log g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) - (1/\lambda) \sum_{j=1}^{n} \frac{|Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})|^\lambda}{e^{\lambda\eta} g^\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)}.$$

Taking derivatives yields

$$\partial/\partial\eta \; \log L = -n + \sum_{j=1}^{n} \frac{|Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})|^\lambda}{e^{\lambda\eta} g^\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)} = 0,$$

$$\partial/\partial\boldsymbol{\theta} \; \log L - \sum_{j=1}^{n} \nu_\theta(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) + \sum_{j=1}^{n} \frac{|Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})|^\lambda}{e^{\lambda\eta} g^\lambda(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)} \nu_\theta(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) = \mathbf{0}.$$

It is straightforward to verify that these equations may be expressed in the form (6.28).

To obtain the "trick," note from above that η must satisfy

$$e^{\lambda\eta} = n^{-1}\sum_{j=1}^{n}\frac{|Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta})|^{\lambda}}{g^{\lambda}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)}.$$

Substituting this into $\log L$ in (1) gives the profile loglikelihood

$$\log L_{\max} = -(n/\lambda)\log\left\{n^{-1}\sum_{j=1}^{n}\frac{|Y_j - f(\boldsymbol{\beta},\boldsymbol{x}_j)|^{\lambda}}{g^{\lambda}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)}\right\} - (n/\lambda)\sum_{j=1}^{n}\log g^{\lambda/n}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j) - n/\lambda.$$

Defining $\dot{g}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)$ as in the notes, we may write

$$\log L_{\max} = -(n/\lambda)\log\left\{n^{-1}\sum_{j=1}^{n}\frac{|Y_j - f(\boldsymbol{\beta},\boldsymbol{x}_j)|^{\lambda}\dot{g}^{\lambda}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)}{g^{\lambda}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)}\right\} - n/\lambda + (n/\lambda)\log n,$$

so that maximizing $\log L$ in $\boldsymbol{\theta}$ is equivalent to minimizing

$$\sum_{j=1}^{n}\left\{\frac{|Y_j - f(\boldsymbol{\beta},\boldsymbol{x}_j)|^{\lambda/2}\dot{g}^{\lambda/2}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)}{g^{\lambda/2}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)}\right\}^2.$$

Thus, we may use nonlinear regression procedures to maximize $\log L$, regressing "dummy" data equal to zero for all j on

$$F_j(\boldsymbol{\theta}) = \frac{|Y_j - f(\boldsymbol{\beta},\boldsymbol{x}_j)|^{\lambda/2}\dot{g}^{\lambda/2}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)}{g^{\lambda/2}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)}.$$

For the power variance model $g(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j) = f^{\theta}(\boldsymbol{x}_j,\boldsymbol{\beta})$, this becomes

$$F_j(\theta) = |Y_j - f(\boldsymbol{\beta},\boldsymbol{x}_j)|^{\lambda/2}\left\{\frac{\dot{f}(\boldsymbol{\beta})}{f(\boldsymbol{x}_j,\boldsymbol{\beta})}\right\}^{\theta\lambda/2}.$$

6. It is straightforward to see that differentiation of (7.11) with respect to σ yields

$$\sigma^2 = (n-p)^{-1}\sum_{j=1}^{n}\frac{\{Y_j - f(\boldsymbol{x}_j,\hat{\boldsymbol{\beta}})\}^2}{g^2(\hat{\boldsymbol{\beta}},\boldsymbol{\theta},\boldsymbol{x}_j)}.$$

Substituting this in (7.11) gives the profile objective function

$$-\frac{(n-p)}{2} - \frac{(n-p)}{2}\log\left[(n-p)^{-1}\sum_{j=1}^{n}\frac{\{Y_j - f(\boldsymbol{x}_j,\hat{\boldsymbol{\beta}})\}^2}{g^2(\hat{\boldsymbol{\beta}},\boldsymbol{\theta},\boldsymbol{x}_j)}\right] - \frac{(n-p)}{2}\log\{\dot{g}(\boldsymbol{\theta})\}^{(2n)/(n-p)}\}$$

$$-\frac{(n-p)}{2}\log|\boldsymbol{N}(\hat{\boldsymbol{\beta}},\boldsymbol{\theta})|^{1/(n-p)},$$

where $\dot{g}(\boldsymbol{\theta}) = \prod_{j=1}^{n} g^{1/2}(\hat{\boldsymbol{\beta}},\boldsymbol{\theta},\boldsymbol{x}_j)$. Note then that $\boldsymbol{\theta}$ maximizing this profile objective function must maximize

$$-\frac{(n-p)}{2}\log\left[\sum_{j=1}^{n}\frac{\{Y_j - f(\boldsymbol{x}_j,\hat{\boldsymbol{\beta}})\}^2}{g^2(\hat{\boldsymbol{\beta}},\boldsymbol{\theta},\boldsymbol{x}_j)}\dot{g}(\boldsymbol{\theta})\}^{(2n)/(n-p)}|\boldsymbol{N}(\hat{\boldsymbol{\beta}},\boldsymbol{\theta})|^{2/\{2(n-p)\}}\right].$$

Ignoring constants, then, the REML estimator for $\boldsymbol{\theta}$ must minimize

$$\sum_{j=1}^{n}\left[\frac{\{Y_j - f(\boldsymbol{x}_j, \hat{\boldsymbol{\beta}})\}}{g(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}, \boldsymbol{x}_j)}\dot{g}(\boldsymbol{\theta})^{n/(n-p)}|\boldsymbol{N}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta})|^{1/2(n-p)}\right]^2.$$

Thus

$$F_j(\boldsymbol{\theta}) = \frac{\{Y_j - f(\boldsymbol{x}_j, \hat{\boldsymbol{\beta}})\}}{g(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}, \boldsymbol{x}_j)}\dot{g}(\boldsymbol{\theta})^{n/(n-p)}|\boldsymbol{N}(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta})|^{1/2(n-p)}.$$