ST 762, HOMEWORK 4 EXTRA PROBLEMS SOLUTIONS, FALL 2009

1. Here, there are no covariates.

(a) Using the notation in the notes, we have $f_{\beta j} = mh_{\beta}(\beta) = me^{\beta}/(1+e^{\beta})^2$, and $g_j^2 = me^{\beta}/(1+e^{\beta})^2$. Thus, these terms cancel in the GLS estimating equation, and the GLS estimator $\hat{\beta}$ satisfies at $C = \infty$

$$\sum_{j=1}^{n} \{Y_j - me^{\hat{\beta}}/(1+e^{\hat{\beta}})\} = 0.$$

Letting $\bar{Y} = n^{-1} = n^{-1}\sum_{j=1}^{n} Y_j$, we have

$$\hat{\beta} = \log\left(\frac{\bar{Y}}{m - \bar{Y}}\right).$$

By the weak law of large numbers, $\bar{Y} \xrightarrow{p} me^{\beta_0}/(1+e^{\beta_0})$. Because $\log\{\eta/(m-\eta)\}$ is continuous in η, we thus have

$$\hat{\beta} \xrightarrow{p} \log\left\{\frac{me^{\beta_0}}{m(1+e^{\beta_0}) - me^{\beta_0}}\right\} = \beta_0.$$

(b) We have $n^{1/2}(\hat{\beta} - \beta_0) = n^{1/2}[\log\{\bar{Y}/(m-\bar{Y})\} - \beta_0]$. Because $\bar{Y} \xrightarrow{p} me^{\beta_0}/(1+e^{\beta_0})$, we can approximate by a first-order Taylor series

$$\begin{aligned}
\log\{\bar{Y}/(m-\bar{Y})\} &\approx \log\left\{\frac{mh(\beta_0)}{m - mh(\beta_))}\right\} + \left[\{(\bar{Y}/(m-\bar{Y})\}^{-1}\left\{\frac{1}{m-\bar{Y}} + \frac{\bar{Y}}{(m-\bar{Y})^2}\right\}\right]_{\bar{Y}=mh(\beta_0)} \\
&\quad \times \left(\bar{Y} - \frac{me^{\beta_0}}{1+e^{\beta_0}}\right) \\
&= \beta_0 + \left\{\frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right\}^{-1}\left(\bar{Y} - \frac{me^{\beta_0}}{1+e^{\beta_0}}\right).
\end{aligned}$$

Thus,

$$n^{1/2}(\hat{\beta} - \beta_0) \approx n^{1/2}\{n^{-1}\sum_{j=1}^{n} Y_j - me^{\beta_0}/(1+e^{\beta_0})\}\left\{\frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right\}.$$

Now

$$n^{1/2}\{n^{-1}\sum_{j=1}^{n} Y_j - me^{\beta_0}/(1+e^{\beta_0})\} \xrightarrow{L} \mathcal{N}\left\{0, \frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right\}.$$

Thus, the full expression satisfies

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{L} \mathcal{N}\left\{0, \left(\frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right)^{-1}\right\}.$$

From the folklore theorem, the limiting variance should be, using shorthand notation,

$$\sigma_0^2\left(n^{-1}\sum_{j=1}^{n} g_{0j}^{-2} f_{\beta 0j} f_{\beta 0j}^T\right)^{-1} = \left\{n^{-1}\sum_{j=1}^{n}\left(\frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right)^{-1}\left(\frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right)^2\right\}^{-1},$$

which reduces to the above ($\sigma_0 = 1$ here). So the above result coincides with the general folklore result.

(c) The GLS estimator is still the same as above. However, now the variance is different. We now have

$$n^{-1}\sum_{j=1}^{n} \text{var} Y_j = n^{-1}\sum_{j=1}^{n}\left\{\frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right\}\{1+\theta(m-1)\},$$

which is a constant, so that $n^{-2}\sum_{j=1}^{n}\text{var}(Y_j) \to 0$. Thus, the weak law of large numbers still applies as in (a), and we have the result that, despite the misspecification of variance, $\hat{\beta} \xrightarrow{p} \beta_0$.

(d) We may make the same approximation as in (b) to arrive at

$$n^{1/2}(\hat{\beta}-\beta_0) \approx n^{1/2}\{n^{-1}\sum_{j=1}^{n} Y_j - me^{\beta_0}/(1+e^{\beta_0})\}\left\{\frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right\}.$$

Now, however, by the central limit theorem, we have

$$n^{1/2}\{n^{-1}\sum_{j=1}^{n} Y_j - me^{\beta_0}/(1+e^{\beta_0})\} \xrightarrow{L} \mathcal{N}\left[0, \frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\{1+\theta(m-1)\}\right],$$

so that, after simplification, we have

$$n^{1/2}(\hat{\beta}-\beta_0) \xrightarrow{L} \mathcal{N}\left[0, \left(\frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right)^{-1}\{1+\theta(m-1)\}\right].$$

From page 224, we expect that the covariance matrix should be of the form

$$\sigma_0^2\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1},$$

where

$$\mathbf{\Sigma}_1^{-1} \approx n^{-1}\sum_{j=1}^{n} k_{0j}^{-2} f_{\beta 0j} f_{\beta 0j}^T$$

and k_{0j} is the misspecified variance function. It is straightforward to obtain

$$\Sigma_1 \approx \frac{me^{\beta_0}}{(1+e^{\beta_0})^2},$$

and

$$\Sigma_2 \approx \frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\{1+\theta(m-1)\}.$$

Thus,

$$\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} = \left\{\frac{me^{\beta_0}}{(1+e^{\beta_0})^2}\right\}^{-1}\{1+\theta(m-1)\},$$

as required.

2. (a) This is the same argument as in the notes, only messier. You should have ended up with

$$\boldsymbol{C}_n^* = n^{-1/2}\sum_{j=1}^{n}(2+\kappa_j-\zeta_j^2)^{-1}\begin{pmatrix} \{(2+\kappa_j)g_{0j}^{-1}f_{\beta 0j} - 2\zeta_j\sigma_0\nu_{\beta 0j}\}\epsilon_j + \{-\zeta_j g_{0j}^{-1}f_{\beta 0j} + 2\sigma_0\nu_{\beta 0j}\}(\epsilon_j^2-1) \\ \{-2\zeta_j\epsilon_j + 2(\epsilon_j^2-1)\tau_{\theta 0j}\end{pmatrix}$$

and

$$\boldsymbol{A}_n^* = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + (\boldsymbol{X}_*^T\boldsymbol{W}^{1/2} - 2\sigma_0\boldsymbol{R}_*^T)(\boldsymbol{W}^{1/2}\boldsymbol{X}_* - 2\sigma_0\boldsymbol{R}_*) & -2(\boldsymbol{X}_*^T - 2\sigma_0\boldsymbol{R}_*^T)\boldsymbol{Q}_* \\ -2\boldsymbol{Q}_*^T(\boldsymbol{X}_* - 2\sigma_0\boldsymbol{R}_*) & 4\boldsymbol{Q}_*^T\boldsymbol{Q}_* \end{pmatrix}.$$

(b) We could proceed to evaluate the RHS of

$$n^{1/2}\begin{pmatrix} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)/\sigma_0 \\ (\hat{\sigma} - \sigma_0)/\sigma_0 \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} \approx \boldsymbol{A}_n^{*\,-1}\boldsymbol{C}_n^*.$$

However, we know from the usual M-estimator argument that when the "covariance matrix" $\boldsymbol{V}_j$ in the original estimating equation is correctly specified, we have immediately that the LHS converges in distribution to a normal with mean $\mathbf{0}$ and covariance matrix $\{\lim n^{-1}\boldsymbol{A}_n^{*\,-1}\}^{-1}$. So all we really need to do is find the upper left-hand $(p \times p)$ submatrix of $\boldsymbol{A}_n^{*\,-1}$. From the results for inversion of partitioned matrices on page 252, we can find that this matrix is

$$\{\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + (\boldsymbol{X}_*^T\boldsymbol{W}^{1/2} - 2\sigma_0\boldsymbol{R}_*^T)(\boldsymbol{W}^{1/2}\boldsymbol{X}_* - 2\sigma_0\boldsymbol{R}_*)$$
$$-(\boldsymbol{X}_*^T\boldsymbol{W}^{1/2} - 2\sigma_0\boldsymbol{R}_*^T)\boldsymbol{Q}_*(\boldsymbol{Q}_*^T\boldsymbol{Q}_*)^{-1}\boldsymbol{Q}_*^T(\boldsymbol{W}^{1/2}\boldsymbol{X}_* - 2\sigma_0\boldsymbol{R}_*)\}^{-1}$$

which is equal to

$$\{\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + (\boldsymbol{X}_*^T\boldsymbol{W}^{1/2} - 2\sigma_0\boldsymbol{R}_*^T)\boldsymbol{P}_*(\boldsymbol{W}^{1/2}\boldsymbol{X}_* - 2\sigma_0\boldsymbol{R}_*)\}^{-1}.$$

Thus

$$\boldsymbol{\Sigma} \approx \{n^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + n^{-1}(\boldsymbol{X}_*^T\boldsymbol{W}^{1/2} - 2\sigma_0\boldsymbol{R}_*^T)\boldsymbol{P}_*(\boldsymbol{W}^{1/2}\boldsymbol{X}_* - 2\sigma_0\boldsymbol{R}_*)\}^{-1}.$$

(c) We know that the GLS estimator has approximate large sample covariance matrix $\{n^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\}^{-1}$. So we want to show that $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} - \{\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + (\boldsymbol{X}_*^T\boldsymbol{W}^{1/2} - 2\sigma_0\boldsymbol{R}_*^T)\boldsymbol{P}_*(\boldsymbol{W}^{1/2}\boldsymbol{X}_* - 2\sigma_0\boldsymbol{R}_*)\}^{-1}$ is nonnegative definite Because $\boldsymbol{P}_*$ is symmetric and idempotent, we can write this second term, letting $\boldsymbol{G}_* = (\boldsymbol{W}^{1/2}\boldsymbol{X}_* - 2\sigma_0\boldsymbol{R}_*)$, as

$$\{\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \boldsymbol{G}_*^T\boldsymbol{P}_*^T\boldsymbol{P}_*\boldsymbol{G}_*\}^{-1} = \{\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \boldsymbol{K}_*^T\boldsymbol{K}_*\}^{-1},$$

say. Now from the partitioned inverse results, we have that this may be written equivalently as

$$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} - (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\{\boldsymbol{K}_*^T\boldsymbol{K}_* + (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\}^{-1}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1},$$

which is clearly $\leq (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$, assuming all matrices in question are positive definite. Thus, the quadratic estimator is at least as efficient as GLS.

(d) When $\zeta_j = 0$ and $\kappa_j = \kappa$, then it is straightforward to see that $\boldsymbol{Z} = \mathbf{0}$, $\boldsymbol{H} = (2+\kappa)\boldsymbol{I}$, and thus $\boldsymbol{X}_* = \mathbf{0}$, $\boldsymbol{R}_* = (2+\kappa)^{-1/2}\boldsymbol{R}$, $\boldsymbol{Q}_* = (2+\kappa)^{-1/2}\boldsymbol{Q}$, and $\boldsymbol{P}_* = \boldsymbol{P}$. Under these conditions, the general result reduces to the one for (c) on page 256 of the notes.

(e) When $g_j = f_j^\theta$, $\log g_j = \theta \log f_j$, and thus $\nu_{\beta 0j} = \theta_0 f_{\beta 0j}/f_j$. Now when $\zeta_j = \zeta = a\sigma_0$ and $\kappa_j = \kappa$, $\boldsymbol{X}_* = \zeta \boldsymbol{X}/(2+\kappa-\zeta^2)^{1/2}$, $\boldsymbol{Q}_* = \boldsymbol{Q}/(2+\kappa-\zeta^2)^{1/2}$, and $\boldsymbol{R}_* = \boldsymbol{R}/(2+\kappa-\zeta^2)^{1/2}$, so that $\boldsymbol{P}_* = \boldsymbol{P}$. Moreover, defining $\boldsymbol{F} = \text{diag}(f_{0j}^{-1})$, $\boldsymbol{R} = \theta_0 \boldsymbol{F}\boldsymbol{X}$. Now for $\zeta = a\sigma_0$ and $\kappa = b\sigma_0^2$, $2+\kappa-\zeta^2 = 2+(b-a^2)\sigma_0^2$. Substituting all of this into the approximation to $\boldsymbol{\Sigma}$ we found in (b), we obtain

$$\boldsymbol{\Sigma} \approx \{n^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + n^{-1}(2+(b-a^2)\sigma_0^2)^{-1}\sigma_0^2\boldsymbol{X}^T(a\boldsymbol{W}^{1/2}-2\boldsymbol{F})\boldsymbol{P}(a\boldsymbol{W}^{1/2}-2\boldsymbol{F})\boldsymbol{X}\}^{-1}.$$

(f) Note that if in fact $\theta_0 = 1$, $\boldsymbol{W}^{1/2} = \boldsymbol{F}$, so if we further take $a = 2$, the second term in $\boldsymbol{\Sigma}$ is zero, as long as $2+(b-a^2)\sigma_0^2 \neq 0$, $\boldsymbol{\Sigma} \approx (n^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{x})^{-1}$, as required. Thus, with $a = 2$, b must satisfy $2+(b-4)\sigma_0^2 \neq 0$. The only way this can equal zero is if $\sigma_0^2 = 2/(4-b)$. Note that if $b = 6$, then these choices of a and b with θ_0 correspond second, third, and fourth moments of a gamma distribution (and certainly this value of b does not violate the above, as $\sigma_0^2 > 0$). This shows that if we set up the estimating equations using the third and fourth moments of the gamma distribution, and in fact the variance is identical to that of a gamma (power model with $\theta_0 = 1$), then there is no benefit to trying to incorporate information from the variance about $\boldsymbol{\beta}$ to gain efficiency; the ML estimator under these conditions solves the linear GLS equations, so this would seem to be the best we can do.

3. (a) The optimal estimating equation is

$$\sum_{j=1}^n (f_{\beta j}, 2\sigma^2 g_j^2 \nu_{\beta j}) \begin{pmatrix} \sigma^2 g_j^2 & 0 \\ 0 & (2+\kappa)\sigma^4 g_j^4 \end{pmatrix}^{-1} \begin{pmatrix} Y_j - f_j \\ (Y_j - f_j)^2 - \sigma^2 g_j^2 \end{pmatrix} = \boldsymbol{0}.$$

Here $f_j = \beta$, $g_j = f_j^{1/2} = \beta^{1/2}$, $f_{\beta j} = 1$, $\nu_{\beta j} = 1/(2\beta)$ for $j = 1, \ldots, m$, and $f_j = 1/\beta$, $g_j = f_j^{1/2} = (1/\beta)^{1/2}$, $f_{\beta j} = -1/\beta^2$, $\nu_{\beta j} = -1/(2\beta)$. Putting all this together, the estimating equation becomes, after some algebra,

$$\sum_{j=1}^m \left[\frac{(Y_j-\beta)}{\sigma^2\beta} + \frac{(Y_j-\beta)^2-\sigma^2\beta}{\sigma^2(2+\kappa)\beta^2}\right] - \sum_{j=m+1}^n \left[\frac{(Y_j-\beta)}{\sigma^2\beta} + \frac{(Y_j-1/\beta)^2-\sigma^2/\beta}{\sigma^2(2+\kappa)}\right] = 0.$$

(b) Defining T_1, T_2, Z_1, Z_2 as described, we may write the estimating equation in (a) as, after algebra, as

$$\frac{m}{\sigma^2\beta}(T_1-T_2)\{1-2/(2+\kappa)\} + \frac{m}{\sigma^2}(1/\beta^2-1)\{1-1/(2+\kappa)\} + \frac{m}{\sigma^2(2+\kappa)}(Z_1/\beta^2-Z_2) = 0.$$

Multiplying by $\beta^2\sigma^2(2+\kappa)/m$ yields an expression that does not depend on σ^2, namely

$$\beta^2(Z_2+1+\kappa) - \beta\kappa(T_1-T_2) - (Z_1+1+\kappa) = 0.$$

This is a quadratic equation in β, so the estimator may be found using the quadratic formula. If $\beta > 1$ then $T_1 > T_2$ is likely, so choose the positive root, and conclude

$$\widehat{\beta}_Q = \frac{\kappa(T_1-T_2) + \{\kappa^2(T_1-T_2)^2 + 4(Z_2+1+\kappa)(Z_1+1+\kappa)\}^{1/2}}{2(Z_2+1+\kappa)}.$$

(c) We know under these conditions as $m \to \infty$ by the WLLN that $T_1 \xrightarrow{p} \beta_0$, $T_2 \xrightarrow{p} 1/\beta_0$, $Z_1 \xrightarrow{p} \beta_0^2 + \sigma_0^2\beta_0$, and $Z_2 \xrightarrow{p} 1/\beta_0^2 + \sigma_0^2/\beta_0$. Because $\widehat{\beta}_Q$ is a continuous function of these quantities, we have

$$\widehat{\beta}_Q \xrightarrow{p} \frac{\kappa(\beta_0-1/\beta_0) + \{\kappa^2(\beta_0-1/\beta_0)^2 + 4(1/\beta_0^2+\sigma_0^2/\beta_0+1+\kappa)(\beta_0^2+\sigma_0^2\beta_0+1+\kappa)\}^{1/2}}{2(1/\beta_0^2+\sigma_0^2/\beta_0+1+\kappa)}.$$

Yuck! It may be shown after some algebra that the stuff in { } is in fact a perfect square and equal to

$$\{(2+\kappa)/\beta_0 + (2+\kappa)/\beta_0 + 2\sigma_0^2\}^2$$

so that it follows that $\widehat{\beta}_Q \xrightarrow{p} \beta_0$. If the algebra was too brutal, it would have still been possible to deduce this; as we know that the estimating equation is unbiased in general, we already are expecting $\widehat{\beta}_Q \xrightarrow{p} \beta_0$. Thus, we could work backwards to determine that the quantity in { } has to be a perfect square.

(d) Under these conditions, we have the much simpler expression

$$\widehat{\beta}_Q^* = \left(\frac{Z_1+1}{Z_2+1}\right)^{1/2}.$$

(Note that you could always derive this directly from the estimating equation if you got confounded by the algebra above.) Thus,

$$m^{1/2}(\widehat{\beta}_Q^* - \beta_0) = m^{1/2}\left\{\left(\frac{Z_1+1}{Z_2+1}\right)^{1/2} - \beta_0\right\}.$$

As $Z_1 \xrightarrow{p} \beta_0^2 + \sigma_0^2\beta_0$, and $Z_2 \xrightarrow{p} 1/\beta_0^2 + \sigma_0^2/\beta_0$, we can take a Taylor series about these values, which gives after much algebra

$$m^{1/2}(\widehat{\beta}_Q^* - \beta_0) \approx \frac{\beta_0}{2}(1/\beta_0 + \beta_0 + \sigma_0^2)^{-1}\left\{m^{-1/2}\sum_{j=1}^{m}(\epsilon_j^2 - 1)\sigma_0^2 - m^{-1/2}\sum_{m+1}^{n}(\epsilon_j^2-1)\sigma_0^2\right.$$

$$\left. + 2\sigma_0\beta_0^{1/2}m^{-1/2}\sum_{j=1}^{m}\epsilon_j - 2\sigma_0\beta_0^{-1/2}m^{-1/2}\sum_{j=m+1}^{n}\epsilon_j\right\}.$$

Applying the CLT, we obtain

$$m^{1/2}(\widehat{\beta}_Q^* - \beta_0) \xrightarrow{L} \mathcal{N}(0, \Delta),$$

where $\Delta = \sigma_0^2\beta_0^2(\beta_0^{-1} + \beta_0 + \sigma_0^2)^{-1}$. You can in fact check that this is correct by plugging into the generic results for normal ML, noting that $m^{1/2} = (n/2)^{1/2}$.

(e) Now we have $Z_1 \xrightarrow{p} \beta_0^2 + \sigma_0^2\beta_0^{\theta_0}$ and $Z_2 \xrightarrow{p} 1/\beta_0^2 + \sigma_0^2(1/\beta_0)^{\theta_0}$. Thus, by continuity, we arrive at, after some algebra,

$$\widehat{\beta}_Q^* \xrightarrow{p} \beta_0\left(\frac{\beta_0^2 + \sigma_0^2\beta_0^{\theta_0} + 1}{\beta_0^2 + \sigma_0^2\beta_0^{2-\theta_0} + 1}\right)^{1/2}.$$

Recall that $\beta_0 > 1$; so this case is excluded. Thus, we must have $\theta_0 = 2 - \theta_0$ to make the term in braces equal to 1, or $\theta_0 = 1$, in order that $\widehat{\beta}_Q^* \xrightarrow{p} \beta_0$. Thus, $\widehat{\beta}_Q^*$ is not consistent in general.

(f) The GLS estimator solves

$$m^{-1}\sum_{j=1}^{m}(Y_j - \beta)/\beta - m^{-1}\sum_{j=m+1}^{n}\beta(Y_j - 1/\beta)/\beta^2 = 0,$$

which yields $\beta^2 + \beta(T_1 - T_2) - 1 = 0$. This is quadratic in β, so we use the quadratic formula again to obtain the positive root

$$\widehat{\beta}_L = \frac{(T_1 - T_2) + \{(T_1 - T_2)^2 + 4\}^{1/2}}{2}.$$

(g) $T_1 \xrightarrow{p} \beta_0$ and $T_2 \xrightarrow{p} 1/\beta_0$ regardless of the form of the true variance of Y_j (as long as the conditions of the WLLN are satisfied, of course). Thus, regardless of whether the assumed variance model is correct, algebra shows that $\widehat{\beta}_L \xrightarrow{p} \beta_0$.