Consider the class of models

$$\boldsymbol{y} = \boldsymbol{x}(\boldsymbol{\beta}) + \boldsymbol{u}, \quad \boldsymbol{u} \sim N\big(\boldsymbol{0}, \boldsymbol{\Omega}(\boldsymbol{\alpha})\big). \tag{9.31}$$

By modifying the loglikelihood function (9.03) slightly, we find that the loglikelihood function corresponding to (9.31) is

$$\begin{aligned} \ell^n(\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = &-\tfrac{n}{2}\log(2\pi) - \tfrac{1}{2}\log|\boldsymbol{\Omega}(\boldsymbol{\alpha})| \\ &- \tfrac{1}{2}\big(\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta})\big)^{\top}\boldsymbol{\Omega}^{-1}(\boldsymbol{\alpha})\big(\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta})\big). \end{aligned} \tag{9.32}$$

There will be two sets of first-order conditions, one for $\boldsymbol{\alpha}$ and one for $\boldsymbol{\beta}$. The latter will be similar to the first-order conditions (9.05) for GNLS:

$$\boldsymbol{X}^{\top}(\hat{\boldsymbol{\beta}})\,\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\alpha}})\big(\boldsymbol{y} - \boldsymbol{x}(\hat{\boldsymbol{\beta}})\big) = \boldsymbol{0}.$$

The former will be rather complicated and will depend on precisely how $\boldsymbol{\Omega}$ is related to $\boldsymbol{\alpha}$. For a more detailed treatment, see Magnus (1978).

In Section 8.10, we saw that the information matrix for $\boldsymbol{\beta}$ and σ in a nonlinear regression model with covariance matrix $\sigma^2\mathbf{I}$ is block-diagonal between $\boldsymbol{\beta}$ and σ. An analogous result turns out to be true for the model (9.31) as well: The information matrix is block-diagonal between $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. This means that, asymptotically, the vectors $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ are independent. Thus the fact that $\hat{\boldsymbol{\alpha}}$ is estimated jointly with $\hat{\boldsymbol{\beta}}$ can be ignored, and $\hat{\boldsymbol{\beta}}$ will have the same properties asymptotically as the GNLS estimator $\tilde{\boldsymbol{\beta}}$ and the feasible GNLS estimator $\check{\boldsymbol{\beta}}$.

The above argument does not require that the error terms u_t actually be normally distributed. All that we require is that the vectors $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ be asymptotically independent and $O(1)$ under whatever DGP actually generated the data. It can be shown that this is in fact the case under fairly general conditions, similar to the conditions detailed in Chapter 5 for least squares to be consistent and asymptotically normal; see White (1982) and Gouriéroux, Monfort, and Trognon (1984) for fundamental results in this area. As we saw in Section 8.1, when the method of maximum likelihood is applied to a data set for which the DGP was not in fact a special case of the model being estimated, the resulting estimator is called a quasi-ML, or QML, estimator. In practice, of course, almost all the ML estimators we use are actually QML estimators, since some of the assumptions of our models are almost always wrong. It is therefore comforting that in certain common situations, including this one, the properties of QML estimators are very similar to those of genuine ML estimators, although asymptotic efficiency is of course lost.

As a concrete example of GLS, feasible GLS, and ML estimation, consider the model

$$\boldsymbol{y} = \boldsymbol{x}(\boldsymbol{\beta}) + \boldsymbol{u}, \;\; \boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}), \;\; \Omega_{tt} = \sigma^2 w_t^{\alpha}, \;\; \Omega_{ts} = 0 \text{ for all } t \neq s. \tag{9.33}$$