form, but they are based on different matrices of residuals. Equation (12.36) and equation (12.09) evaluated at $\hat{\boldsymbol{\Sigma}}^{\rm ML}$, that is

$$\hat{\boldsymbol{\beta}}_\bullet^{\rm ML} = \big(\boldsymbol{X}_\bullet^\top(\hat{\boldsymbol{\Sigma}}_{\rm ML}^{-1} \otimes \mathbf{I}_n)\boldsymbol{X}_\bullet\big)^{-1}\boldsymbol{X}_\bullet^\top(\hat{\boldsymbol{\Sigma}}_{\rm ML}^{-1} \otimes \mathbf{I}_n)\boldsymbol{y}_\bullet, \tag{12.37}$$

together define the ML estimator for the model (12.28).

Equations (12.36) and (12.37) are exactly the ones that are used by the continuously updated GMM estimator to update the estimates of $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}_\bullet$, respectively. It follows that, if the continuous updating procedure converges, it converges to the ML estimator. Consequently, we can estimate the covariance matrix of $\hat{\boldsymbol{\beta}}_\bullet^{\rm ML}$ in the same way as for the GLS or GMM estimator, by the formula

$$\widehat{\rm Var}(\hat{\boldsymbol{\beta}}_\bullet^{\rm ML}) = \big(\boldsymbol{X}_\bullet^\top(\hat{\boldsymbol{\Sigma}}_{\rm ML}^{-1} \otimes \mathbf{I}_n)\boldsymbol{X}_\bullet\big)^{-1}. \tag{12.38}$$

It is also possible to estimate the covariance matrix of the estimated contemporaneous covariance matrix, $\hat{\boldsymbol{\Sigma}}^{\rm ML}$, although this is rarely done. If the elements of $\boldsymbol{\Sigma}$ are stacked in a vector of dimension g^2, a suitable estimator is

$$\widehat{\rm Var}\big(\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_\bullet^{\rm ML})\big) = \frac{2}{n}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_\bullet^{\rm ML}) \otimes \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_\bullet^{\rm ML}). \tag{12.39}$$

Notice that the estimated variance of any diagonal element of $\boldsymbol{\Sigma}$ is just twice the square of that element, divided by n. This is precisely what is obtained for the univariate case in Exercise 10.10. As with that result, the asymptotic validity of (12.39) depends critically on the assumption that the error terms are multivariate normal.

As we saw in Chapter 10, ML estimators are consistent and asymptotically efficient *if* the underlying model is correctly specified. It may therefore seem that the asymptotic efficiency of the ML estimator (12.37) depends critically on the multivariate normality assumption. However, the fact that the ML estimator is identical to the continuously updated efficient GMM estimator means that it is in fact efficient in the same sense as the latter. When the errors are not normal, the estimator is more properly termed a QMLE (see Section 10.4). As such, it is consistent, but not necessarily efficient, under assumptions about the error terms that are no stronger than those needed for feasible GLS to be consistent. Moreover, if the stronger assumptions made in (12.02) hold, even without normality, then the estimator (12.38) of $\text{Var}(\hat{\boldsymbol{\beta}}_\bullet^{\rm ML})$ is asymptotically valid. If the error terms are not normal, it would be necessary to have information about their actual distribution in order to derive an estimator with a smaller asymptotic variance than (12.37).

It is of considerable theoretical interest to concentrate the loglikelihood function (12.33) with respect to $\boldsymbol{\Sigma}$. In order to do so, we use the first-order conditions that led to (12.36) to define $\boldsymbol{\Sigma}(\boldsymbol{\beta}_\bullet)$ as the matrix that maximizes (12.33) for given $\boldsymbol{\beta}_\bullet$. We find that

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}_\bullet) \equiv \frac{1}{n}\boldsymbol{U}^\top(\boldsymbol{\beta}_\bullet)\boldsymbol{U}(\boldsymbol{\beta}_\bullet).$$