Of course, the last line of (12.08) can be true only for nonsingular, square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The Kronecker product is not commutative, by which we mean that $\boldsymbol{A} \otimes \boldsymbol{B}$ and $\boldsymbol{B} \otimes \boldsymbol{A}$ are different matrices. However, the elements of these two products are the same; they are just laid out differently. In fact, it can be shown that $\boldsymbol{B} \otimes \boldsymbol{A}$ can be obtained from $\boldsymbol{A} \otimes \boldsymbol{B}$ by a sequence of interchanges of rows and columns. Exercise 12.2 asks readers to prove these properties of Kronecker products. For an exceedingly detailed discussion of the properties of Kronecker products, see Magnus and Neudecker (1988).

As we have seen, the system of equations defined by (12.01) and (12.02) is equivalent to the single equation (12.04), with gn observations and error terms that have covariance matrix $\boldsymbol{\Sigma}_\bullet$. Therefore, when the matrix $\boldsymbol{\Sigma}$ is known, we can obtain consistent and efficient estimates of the $\boldsymbol{\beta}_i$, or equivalently of $\boldsymbol{\beta}_\bullet$, simply by using the classical GLS estimator (7.04). We find that

$$\begin{aligned}\hat{\boldsymbol{\beta}}_\bullet^{\mathrm{GLS}} &= (\boldsymbol{X}_\bullet^\top \boldsymbol{\Sigma}_\bullet^{-1} \boldsymbol{X}_\bullet)^{-1} \boldsymbol{X}_\bullet^\top \boldsymbol{\Sigma}_\bullet^{-1} \boldsymbol{y}_\bullet \\ &= \big(\boldsymbol{X}_\bullet^\top (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \boldsymbol{X}_\bullet\big)^{-1} \boldsymbol{X}_\bullet^\top (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \boldsymbol{y}_\bullet, \end{aligned} \tag{12.09}$$

where, to obtain the second line, we have used the last of equations (12.08). This GLS estimator is sometimes called the **SUR estimator**. From the result (7.05) for GLS estimation, its covariance matrix is

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_\bullet^{\mathrm{GLS}}) = \big(\boldsymbol{X}_\bullet^\top (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \boldsymbol{X}_\bullet\big)^{-1}. \tag{12.10}$$

Since $\boldsymbol{\Sigma}$ is assumed to be known, we can use this covariance matrix directly, because there are no variance parameters to estimate.

As in the univariate case, there is a criterion function associated with the GLS estimator (7.04). This criterion function is simply expression (7.06) adapted to the model (12.04), namely,

$$(\boldsymbol{y}_\bullet - \boldsymbol{X}_\bullet \boldsymbol{\beta}_\bullet)^\top (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)(\boldsymbol{y}_\bullet - \boldsymbol{X}_\bullet \boldsymbol{\beta}_\bullet). \tag{12.11}$$

The first-order conditions for the minimization of (12.11) with respect to $\boldsymbol{\beta}_\bullet$ can be written as

$$\boldsymbol{X}_\bullet^\top (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)(\boldsymbol{y}_\bullet - \boldsymbol{X}_\bullet \boldsymbol{\beta}_\bullet) = \mathbf{0}. \tag{12.12}$$

These moment conditions, which are analogous to conditions (7.07) for the case of univariate GLS estimation, can be interpreted as a set of estimating equations that define the GLS estimator (12.09).

In the slightly less unrealistic situation in which $\boldsymbol{\Sigma}$ is assumed to be known only up to a scalar factor, so that $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Delta}$, the form of (12.09) would be unchanged, but with $\boldsymbol{\Delta}$ replacing $\boldsymbol{\Sigma}$, and the covariance matrix (12.10) would become

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_\bullet^{\mathrm{GLS}}) = \sigma^2 \big(\boldsymbol{X}_\bullet^\top (\boldsymbol{\Delta}^{-1} \otimes \mathbf{I}_n) \boldsymbol{X}_\bullet\big)^{-1}.$$

In practice, to estimate $\operatorname{Var}(\hat{\boldsymbol{\beta}}_\bullet^{\rm GLS})$, we replace σ^2 by something that estimates it consistently. Two natural estimators are

$$\hat{\sigma}^2 \equiv \frac{1}{gn}\,\hat{\boldsymbol{u}}_\bullet^\top(\boldsymbol{\Delta}^{-1}\otimes\mathbf{I}_n)\,\hat{\boldsymbol{u}}_\bullet,\ \text{ and}$$
$$s^2 \equiv \frac{1}{(gn-k)}\,\hat{\boldsymbol{u}}_\bullet^\top(\boldsymbol{\Delta}^{-1}\otimes\mathbf{I}_n)\,\hat{\boldsymbol{u}}_\bullet,$$

where $\hat{\boldsymbol{u}}_\bullet$ denotes the vector of error terms from GLS estimation of (12.04). The first of these estimators is analogous to the ML estimator of σ^2 in the linear regression model, and the second is analogous to the GLS estimator.

At this point, a word of warning is in order. Although the GLS estimator (12.09) has quite a simple form, it can be expensive to compute when gn is large. In consequence, no sensible regression package would actually use this formula. We can proceed more efficiently by working directly with the estimating equations (12.12). Writing them out explicitly, we obtain

$$\begin{aligned}
&\boldsymbol{X}_\bullet^\top(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_n)(\boldsymbol{y}_\bullet-\boldsymbol{X}_\bullet\hat{\boldsymbol{\beta}}_\bullet)\\
&\quad=\begin{bmatrix}\boldsymbol{X}_1^\top & \cdots & \mathbf{O}\\ \vdots & \ddots & \vdots\\ \mathbf{O} & \cdots & \boldsymbol{X}_g^\top\end{bmatrix}\begin{bmatrix}\sigma^{11}\mathbf{I}_n & \cdots & \sigma^{1g}\mathbf{I}_n\\ \vdots & \ddots & \vdots\\ \sigma^{g1}\mathbf{I}_n & \cdots & \sigma^{gg}\mathbf{I}_n\end{bmatrix}\begin{bmatrix}\boldsymbol{y}_1-\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1^{\rm GLS}\\ \vdots\\ \boldsymbol{y}_g-\boldsymbol{X}_g\hat{\boldsymbol{\beta}}_g^{\rm GLS}\end{bmatrix}\\
&\quad=\begin{bmatrix}\sigma^{11}\boldsymbol{X}_1^\top & \cdots & \sigma^{1g}\boldsymbol{X}_1^\top\\ \vdots & \ddots & \vdots\\ \sigma^{g1}\boldsymbol{X}_g^\top & \cdots & \sigma^{gg}\boldsymbol{X}_g^\top\end{bmatrix}\begin{bmatrix}\boldsymbol{y}_1-\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1^{\rm GLS}\\ \vdots\\ \boldsymbol{y}_g-\boldsymbol{X}_g\hat{\boldsymbol{\beta}}_g^{\rm GLS}\end{bmatrix}=\mathbf{0},
\end{aligned}\tag{12.13}$$

where σ^{ij} denotes the $ij^{\,\rm th}$ element of the matrix $\boldsymbol{\Sigma}^{-1}$. By solving the k equations (12.13) for the $\hat{\boldsymbol{\beta}}_i$, we find easily enough (see Exercise 12.5) that

$$\hat{\boldsymbol{\beta}}_\bullet^{\rm GLS}=\begin{bmatrix}\sigma^{11}\boldsymbol{X}_1^\top\boldsymbol{X}_1 & \cdots & \sigma^{1g}\boldsymbol{X}_1^\top\boldsymbol{X}_g\\ \vdots & \ddots & \vdots\\ \sigma^{g1}\boldsymbol{X}_g^\top\boldsymbol{X}_1 & \cdots & \sigma^{gg}\boldsymbol{X}_g^\top\boldsymbol{X}_g\end{bmatrix}^{-1}\begin{bmatrix}\sum_{j=1}^g\sigma^{1j}\boldsymbol{X}_1^\top\boldsymbol{y}_j\\ \vdots\\ \sum_{j=1}^g\sigma^{gj}\boldsymbol{X}_g^\top\boldsymbol{y}_j\end{bmatrix}. \tag{12.14}$$

Although this expression may look more complicated than (12.09), it is much less costly to compute. Recall that we grouped all the linearly independent explanatory variables of the entire SUR system into the $n\times l$ matrix $\boldsymbol{X}$. By computing the matrix product $\boldsymbol{X}^\top\boldsymbol{X}$, we may obtain all the blocks of the form $\boldsymbol{X}_i^\top\boldsymbol{X}_j$ merely by selecting the appropriate rows and corresponding columns of this product. Similarly, if we form the $n\times g$ matrix $\boldsymbol{Y}$ by stacking the g dependent variables horizontally rather than vertically, so that

$$\boldsymbol{Y}\equiv[\,\boldsymbol{y}_1\quad\cdots\quad\boldsymbol{y}_g\,],$$