Since it is rather complicated to work out the first-order conditions for the maximization of (12.80) with respect to the $\boldsymbol{\beta}_{2i}$, we leave this derivation to the appendix. These conditions can be expressed as

$$\boldsymbol{Y}_\bullet^\top(\boldsymbol{B},\boldsymbol{\Gamma})(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_n)(\boldsymbol{y}_\bullet-\boldsymbol{X}_\bullet\boldsymbol{\beta}_\bullet)=\mathbf{0}, \tag{12.83}$$

where the $gn\times\sum_i k_{2i}$ matrix $\boldsymbol{Y}_\bullet(\boldsymbol{B},\boldsymbol{\Gamma})$ is again defined in terms of diagonal blocks. Block i is the $n\times k_{2i}$ matrix $\boldsymbol{Y}_i(\boldsymbol{B},\boldsymbol{\Gamma})$, which is the submatrix of $\boldsymbol{W}\boldsymbol{B}\boldsymbol{\Gamma}^{-1}$ formed by selecting the columns that correspond to the columns of the matrix $\boldsymbol{Y}_i$ of included endogenous explanatory variables in equation i. The conditions (12.82) and (12.83) can be grouped together as

$$\boldsymbol{X}_\bullet^\top(\boldsymbol{B},\boldsymbol{\Gamma})(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_n)(\boldsymbol{y}_\bullet-\boldsymbol{X}_\bullet\boldsymbol{\beta}_\bullet)=\mathbf{0}, \tag{12.84}$$

where the i^{th} diagonal block of $\boldsymbol{X}_\bullet(\boldsymbol{B},\boldsymbol{\Gamma})$ is the $n\times k_i$ matrix $[\boldsymbol{Z}_i \quad \boldsymbol{Y}_i(\boldsymbol{B},\boldsymbol{\Gamma})]$. There are $k=\sum_i k_{1i}+\sum_i k_{2i}$ equations in (12.84).

With (12.81) and (12.84), we have assembled all of the first-order conditions that define the FIML estimator. We write them here as a set of estimating equations:

$$\begin{aligned}&\boldsymbol{X}_\bullet^\top(\hat{\boldsymbol{B}}_{\text{ML}},\hat{\boldsymbol{\Gamma}}_{\text{ML}})(\hat{\boldsymbol{\Sigma}}_{\text{ML}}^{-1}\otimes\mathbf{I}_n)(\boldsymbol{y}_\bullet-\boldsymbol{X}_\bullet\hat{\boldsymbol{\beta}}_\bullet^{\text{ML}})=\mathbf{0},\ \text{and}\\ &\hat{\boldsymbol{\Sigma}}_{\text{ML}}=\frac{1}{n}(\boldsymbol{Y}\hat{\boldsymbol{\Gamma}}_{\text{ML}}-\boldsymbol{W}\hat{\boldsymbol{B}}_{\text{ML}})^\top(\boldsymbol{Y}\hat{\boldsymbol{\Gamma}}_{\text{ML}}-\boldsymbol{W}\hat{\boldsymbol{B}}_{\text{ML}}).\end{aligned} \tag{12.85}$$

Solving these equations, which must of course be done numerically, yields the FIML estimator.

There are many numerical methods for obtaining FIML estimates. One of them is to make use of the artificial regression

$$(\boldsymbol{\Psi}^\top\otimes\mathbf{I}_n)(\boldsymbol{y}_\bullet-\boldsymbol{X}_\bullet\boldsymbol{\beta}_\bullet)=(\boldsymbol{\Psi}^\top\otimes\mathbf{I}_n)\boldsymbol{X}_\bullet(\boldsymbol{B},\boldsymbol{\Gamma})\boldsymbol{b}+\text{residuals}, \tag{12.86}$$

where, as usual, $\boldsymbol{\Psi}\boldsymbol{\Psi}^\top=\boldsymbol{\Sigma}^{-1}$. This is analogous to the multivariate GNR (12.53). If we start from initial consistent estimates, this artificial regression can be used to update the estimates of $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$, and equation (12.81) can be used to update the estimate of $\boldsymbol{\Sigma}$. Like other artificial regressions, (12.86) can also be used to compute test statistics and covariance matrices.

Another approach is to concentrate the loglikelihood function with respect to $\boldsymbol{\Sigma}$. As readers are asked to show in Exercise 12.24, the concentrated loglikelihood function can be written as

$$-\frac{gn}{2}(\log 2\pi+1)+n\log|\det\boldsymbol{\Gamma}|-\frac{n}{2}\log\left|\frac{1}{n}(\boldsymbol{Y}\boldsymbol{\Gamma}-\boldsymbol{W}\boldsymbol{B})^\top(\boldsymbol{Y}\boldsymbol{\Gamma}-\boldsymbol{W}\boldsymbol{B})\right|, \tag{12.87}$$

which is the analog of (12.41) and (12.51). Expression (12.87) may be maximized directly with respect to $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ to yield $\hat{\boldsymbol{B}}_{\text{ML}}$ and $\hat{\boldsymbol{\Gamma}}_{\text{ML}}$. This approach may or may not be easier numerically than solving equations (12.85).

The FIML estimator is not defined if the matrix $(\boldsymbol{Y\Gamma} - \boldsymbol{WB})^\top(\boldsymbol{Y\Gamma} - \boldsymbol{WB})$ that appears in (12.87) does not have full rank for all admissible values of $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$, and this requires that $n \geq g+k$. This result suggests that n may have to be substantially greater than $g+k$ if FIML is to have good finite-sample properties; see Sargan (1975) and Brown (1981).

Comparison with Three-Stage Least Squares

Even though the FIML and 3SLS estimators are asymptotically equivalent, the FIML estimator is not, in general, equal to the continuously updated 3SLS estimator. In order to study the relationship between the two estimators, we write out explicitly the estimating equations for 3SLS and compare them with the estimating equations (12.85) for FIML. Equations (12.58) and (12.17) imply that the continuously updated version of the 3SLS estimator is defined by the equations

$$\begin{aligned}&\hat{\boldsymbol{X}}_\bullet^\top(\hat{\boldsymbol{\Sigma}}_{\text{3SLS}}^{-1} \otimes \mathbf{I}_n)(\boldsymbol{y}_\bullet - \boldsymbol{X}_\bullet\hat{\boldsymbol{\beta}}_\bullet^{\text{3SLS}}) = \mathbf{0}, \text{ and}\\ &\hat{\boldsymbol{\Sigma}}_{\text{3SLS}} = \frac{1}{n}(\boldsymbol{Y}\hat{\boldsymbol{\Gamma}}_{\text{3SLS}} - \boldsymbol{W}\hat{\boldsymbol{B}}_{\text{3SLS}})^\top(\boldsymbol{Y}\hat{\boldsymbol{\Gamma}}_{\text{3SLS}} - \boldsymbol{W}\hat{\boldsymbol{B}}_{\text{3SLS}}).\end{aligned} \tag{12.88}$$

The second of these equations has exactly the same form as the second equation of (12.85). The first equation is also very similar to the first equation of (12.85), but there is one difference. In (12.85), the leftmost matrix on the left-hand side of the first equation is the transpose of $\boldsymbol{X}_\bullet(\hat{\boldsymbol{B}}_{\text{ML}}, \hat{\boldsymbol{\Gamma}}_{\text{ML}})$, of which the typical diagonal block is $[\,\boldsymbol{Z}_i \quad \boldsymbol{Y}_i(\hat{\boldsymbol{B}}_{\text{ML}}, \hat{\boldsymbol{\Gamma}}_{\text{ML}})\,]$. In contrast, the corresponding matrix in the first equation of (12.88) is the transpose of $\hat{\boldsymbol{X}}_\bullet$, of which the typical diagonal block is, from (12.57), $[\,\boldsymbol{Z}_i \quad \boldsymbol{P_W Y}_i\,]$.

In both cases, the matrix is an estimate of the matrix of optimal instruments for equation i, that is, the matrix of the expectations of the explanatory variables conditional on all predetermined information. It is clear from the RRF (12.70) that this matrix is $[\,\boldsymbol{Z}_i \quad \boldsymbol{Y}_i(\boldsymbol{B}, \boldsymbol{\Gamma})\,]$, where $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are the true parameters of the DGP. FIML uses the FIML estimates of $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ in place of the true values, while 3SLS estimates $\boldsymbol{Y}_i(\boldsymbol{B}, \boldsymbol{\Gamma})$ by $\boldsymbol{P_W Y}_i$, that is, by the fitted values from estimation of the *unrestricted* reduced form (12.71). The latter is, in general, less efficient than the former.

If the restricted and unrestricted reduced forms are equivalent, as they must be if all the equations of the system are just identified, then the estimating equations (12.88) and (12.85) are also equivalent, and the 3SLS and FIML estimators must coincide. In this case, as we saw in the last section, 3SLS is also the same as 2SLS, that is, equation-by-equation IV estimation. Thus all the estimators we have considered are identical in the just-identified case. When there are overidentifying restrictions, and 3SLS is used without continuous updating, then the 3SLS estimators of $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are replaced by the 2SLS ones in the second equation of (12.88). Solving this equation yields the classical 3SLS estimator (12.78), which is evidently much easier to compute than the FIML estimator.