As expected, the feasible GLS estimator (12.18) and the estimated covariance matrix (12.19) have precisely the same forms as their full GLS counterparts, which are (12.09) and (12.10), respectively.

Because we divided by n in (12.17), $\hat{\boldsymbol{\Sigma}}$ must be a biased estimator of $\boldsymbol{\Sigma}$. If k_i is the same for all i, then it would seem natural to divide by $n-k_i$ instead, and this would produce unbiased estimates of the diagonal elements if $\boldsymbol{X}_\bullet$ were exogenous. But we cannot do that when k_i is not the same in all equations. If we were to divide different elements of $\hat{\boldsymbol{U}}^\top\hat{\boldsymbol{U}}$ by different quantities, the resulting estimate of $\boldsymbol{\Sigma}$ would not necessarily be positive definite.

Replacing $\boldsymbol{\Sigma}$ with an estimator $\hat{\boldsymbol{\Sigma}}$ based on OLS estimates, or indeed any other estimator, inevitably degrades the finite-sample properties of the GLS estimator. In general, we would expect the performance of the feasible GLS estimator, relative to that of the GLS estimator, to be especially poor when the sample size is small and the number of equations is large. Under the strong assumption that all the regressors are exogenous, exact inference based on the normal and χ^2 distributions is possible whenever the error terms are normally distributed and $\boldsymbol{\Sigma}$ (or $\boldsymbol{\Delta}$) is known, but this is not the case when $\boldsymbol{\Sigma}$ has to be estimated. Not surprisingly, there is evidence that bootstrapping can yield more reliable inferences than using asymptotic theory for SUR models; see, among others, Rilstone and Veall (1996) and Fiebig and Kim (2000).

Cases in Which OLS Estimation Is Efficient

The SUR estimator (12.09) is efficient under the assumptions we have made, because it is just a special case of the GLS estimator (7.04), the efficiency of which was proved in Section 7.2. In contrast, the OLS estimator (12.06) is, in general, inefficient. The reason is that, unless the matrix $\boldsymbol{\Sigma}$ is proportional to an identity matrix, the error terms of equation (12.04) are not IID. Nevertheless, there are two important special cases in which the OLS estimator is numerically identical to the SUR estimator, and therefore just as efficient.

In the first case, the matrix $\boldsymbol{\Sigma}$ is diagonal, although the diagonal elements need not be the same. This implies that the error terms of equation (12.04) are heteroskedastic but serially independent. It might seem that this heteroskedasticity would cause inefficiency, but that turns out not to be the case. If $\boldsymbol{\Sigma}$ is diagonal, then so is $\boldsymbol{\Sigma}^{-1}$, which means that $\sigma^{ij} = 0$ for $i \neq j$. In that case, the estimating equations (12.16) simplify to

$$\sigma^{ii}\boldsymbol{X}_i^\top(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i^{\rm GLS}) = \boldsymbol{0}, \quad i = 1, \ldots, g.$$

The factors σ^{ii}, which must be nonzero, have no influence on the solutions to the above equations, which are therefore the same as the solutions to the g independent sets of equations $\boldsymbol{X}_i^\top(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i) = \boldsymbol{0}$ which define the equation-by-equation OLS estimator (12.06). Thus, if the error terms are uncorrelated across equations, the GLS and OLS estimators are numerically identical. The "seemingly" unrelated equations are indeed unrelated in this case.