When $l > k$, the model is overidentified, and the estimator (9.13) depends on the choice of $\boldsymbol{J}$ or $\boldsymbol{\Lambda}$. The efficient GMM estimator, for a given set of instruments, is defined in terms of the true covariance matrix $\boldsymbol{\Omega}_0$, which is usually unknown. If $\boldsymbol{\Omega}_0$ is known up to a scalar multiplicative factor, so that $\boldsymbol{\Omega}_0 = \sigma^2\boldsymbol{\Delta}_0$, with σ^2 unknown and $\boldsymbol{\Delta}_0$ known, then $\boldsymbol{\Delta}_0$ can be used in place of $\boldsymbol{\Omega}_0$ in either (9.10) or (9.11). This is true because multiplying $\boldsymbol{\Omega}_0$ by a scalar leaves (9.10) invariant, and it also leaves invariant the $\boldsymbol{\beta}$ that minimizes (9.11).

GMM Estimation with Heteroskedasticity of Unknown Form

The assumption that $\boldsymbol{\Omega}_0$ is known, even up to a scalar factor, is often too strong. What makes GMM estimation practical more generally is that, in both (9.10) and (9.11), $\boldsymbol{\Omega}_0$ appears only through the $l \times l$ matrix product $\boldsymbol{W}^\top\boldsymbol{\Omega}_0\boldsymbol{W}$. As we saw first in Section 5.5, in the context of heteroskedasticity consistent covariance matrix estimation, n^{-1} times such a matrix can be estimated consistently if $\boldsymbol{\Omega}_0$ is a diagonal matrix. What is needed is a preliminary consistent estimate of the parameter vector $\boldsymbol{\beta}$, which furnishes residuals that are consistent estimates of the error terms.

The preliminary estimates of $\boldsymbol{\beta}$ must be consistent, but they need not be asymptotically efficient, and so we can obtain them by using any convenient choice of $\boldsymbol{J}$ or $\boldsymbol{\Lambda}$. One choice that is often convenient is $\boldsymbol{\Lambda} = (\boldsymbol{W}^\top\boldsymbol{W})^{-1}$, in which case the preliminary estimator is the generalized IV estimator (8.29). We then use the preliminary estimates $\hat{\boldsymbol{\beta}}$ to calculate the residuals $\hat{u}_t \equiv y_t - \boldsymbol{X}\hat{\boldsymbol{\beta}}$. A typical element of the matrix $n^{-1}\boldsymbol{W}^\top\boldsymbol{\Omega}_0\boldsymbol{W}$ can then be estimated by

$$\frac{1}{n}\sum_{t=1}^{n}\hat{u}_t^2 w_{ti}w_{tj}. \tag{9.14}$$

This estimator is very similar to (5.36), and the estimator (9.14) can be proved to be consistent by using arguments just like those employed in Section 5.5.

The matrix with typical element (9.14) can be written as $n^{-1}\boldsymbol{W}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{W}$, where $\hat{\boldsymbol{\Omega}}$ is an $n \times n$ diagonal matrix with typical diagonal element $\hat{u}_t^2$. Then the **feasible efficient GMM estimator** is

$$\hat{\boldsymbol{\beta}}_{\text{FGMM}} = \big(\boldsymbol{X}^\top\boldsymbol{W}(\boldsymbol{W}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{W}(\boldsymbol{W}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{y}, \tag{9.15}$$

which is just (9.10) with $\boldsymbol{\Omega}_0$ replaced by $\hat{\boldsymbol{\Omega}}$. Since $n^{-1}\boldsymbol{W}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{W}$ consistently estimates $n^{-1}\boldsymbol{W}^\top\boldsymbol{\Omega}_0\boldsymbol{W}$, it follows that $\hat{\boldsymbol{\beta}}_{\text{FGMM}}$ is asymptotically equivalent to (9.10). It should be noted that, in calling (9.15) efficient, we mean that it is asymptotically efficient within the class of estimators that use the given instrument set $\boldsymbol{W}$.

Like other procedures that start from a preliminary estimate, this one can be iterated. The GMM residuals $y_t - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{FGMM}}$ can be used to calculate a new estimate of $\boldsymbol{\Omega}$, which can then be used to obtain second-round GMM estimates, which can then be used to calculate yet another estimate of $\boldsymbol{\Omega}$, and so on. We will refer to this iterative procedure as **continuously updated GMM**, although it is not quite the same as the procedure by that name investigated by Hansen, Heaton, and Yaron (1996). Whether we stop after one round or continue until the procedure converges, the estimates have the same asymptotic distribution if the model is correctly specified. However, there is evidence that performing more iterations improves finite-sample performance. In practice, the covariance matrix is estimated by

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_{\mathrm{FGMM}}) = \big(\boldsymbol{X}^\top\boldsymbol{W}(\boldsymbol{W}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{X}\big)^{-1}. \tag{9.16}$$

It is not hard to see that n times the estimator (9.16) tends to the asymptotic covariance matrix (9.09) as $n \to \infty$.

Fully Efficient GMM Estimation

In choosing to use a particular matrix of instrumental variables $\boldsymbol{W}$, we are choosing a particular representation of the information sets Ω_t appropriate for each observation in the sample. It is required that $\boldsymbol{W}_t \in \Omega_t$ for all t, and it follows from this that any deterministic function, linear or nonlinear, of the elements of $\boldsymbol{W}_t$ also belongs to Ω_t. It is quite clearly impossible to use all such deterministic functions as actual instrumental variables, and so the econometrician must make a choice. What we have established so far is that, once the choice of $\boldsymbol{W}$ is made, (9.08) gives the optimal set of linear combinations of the columns of $\boldsymbol{W}$ to use for estimation. What remains to be seen is how best to choose $\boldsymbol{W}$ out of all the possible valid instruments, given the information sets Ω_t.

In Section 8.3, we saw that, for the model (9.01) with $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$, the best choice, by the criterion of the asymptotic covariance matrix, is the matrix $\bar{\boldsymbol{X}}$ given in (8.18) by the defining condition that $\mathrm{E}(\boldsymbol{X}_t \,|\, \Omega_t) = \bar{\boldsymbol{X}}_t$, where $\boldsymbol{X}_t$ and $\bar{\boldsymbol{X}}_t$ are the t^{th} rows of $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$, respectively. However, it is easy to see that this result does not hold unmodified when $\boldsymbol{\Omega}$ is not proportional to an identity matrix. Consider the GMM estimator (9.10), of which (9.15) is the feasible version, in the special case of exogenous explanatory variables, for which the obvious choice of instruments is $\boldsymbol{W} = \boldsymbol{X}$. If, for notational ease, we write $\boldsymbol{\Omega}$ for the true covariance matrix $\boldsymbol{\Omega}_0$, (9.10) becomes

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{GMM}} &= \big(\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}\\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}\\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}\\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y} = \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}.
\end{aligned}$$

However, we know from the results of Section 7.2 that the efficient estimator is actually the GLS estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{y}, \tag{9.17}$$

which, except in special cases, is different from $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$.