The next step, as in Section 8.3, is to choose $\boldsymbol{J}$ so as to minimize the covariance matrix (9.07). We may reasonably expect that, with such a choice of $\boldsymbol{J}$, the covariance matrix would no longer have the form of a sandwich. The simplest choice of $\boldsymbol{J}$ that eliminates the sandwich in (9.07) is

$$\boldsymbol{J} = (\boldsymbol{W}^\top \boldsymbol{\Omega}_0 \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{X}; \tag{9.08}$$

notice that, in the special case in which $\boldsymbol{\Omega}_0$ is proportional to $\mathbf{I}$, this expression reduces to the result (8.24) that we found in Section 8.3 as the solution for that special case. We can see, therefore, that (9.08) is the appropriate generalization of (8.24) when $\boldsymbol{\Omega}$ is not proportional to an identity matrix. With $\boldsymbol{J}$ defined by (9.08), the covariance matrix (9.07) becomes

$$\operatorname*{plim}_{n\to\infty} \Big(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{W} (\boldsymbol{W}^\top \boldsymbol{\Omega}_0 \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{X}\Big)^{-1}, \tag{9.09}$$

and the **efficient GMM estimator** is

$$\hat{\boldsymbol{\beta}}_{\text{GMM}} = \big(\boldsymbol{X}^\top \boldsymbol{W} (\boldsymbol{W}^\top \boldsymbol{\Omega}_0 \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{X}\big)^{-1} \boldsymbol{X}^\top \boldsymbol{W} (\boldsymbol{W}^\top \boldsymbol{\Omega}_0 \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{y}. \tag{9.10}$$

When $\boldsymbol{\Omega}_0 = \sigma^2 \mathbf{I}$, this estimator reduces to the generalized IV estimator (8.29). In Exercise 9.1, readers are invited to show that the difference between the covariance matrices (9.07) and (9.09) is a positive semidefinite matrix, thereby confirming (9.08) as the optimal choice for $\boldsymbol{J}$. The estimator $\hat{\boldsymbol{\beta}}_{\text{GMM}}$ is efficient in the class of estimators defined by the moment conditions (9.05), but we will see that a more efficient estimator is available if we know $\boldsymbol{\Omega}_0$ and are prepared to exploit that knowledge.

The GMM Criterion Function

With both GLS and IV estimation, we showed that the efficient estimators could also be derived by minimizing an appropriate criterion function; this function was (7.06) for GLS and (8.30) for IV. Similarly, the efficient GMM estimator (9.10) minimizes the **GMM criterion function**

$$Q(\boldsymbol{\beta}, \boldsymbol{y}) \equiv (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{W} (\boldsymbol{W}^\top \boldsymbol{\Omega}_0 \boldsymbol{W})^{-1} \boldsymbol{W}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{9.11}$$

as can be seen at once by noting that the first-order conditions for minimizing (9.11) are

$$\boldsymbol{X}^\top \boldsymbol{W} (\boldsymbol{W}^\top \boldsymbol{\Omega}_0 \boldsymbol{W})^{-1} \boldsymbol{W}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}.$$

If $\boldsymbol{\Omega}_0 = \sigma_0^2 \mathbf{I}$, (9.11) reduces to the IV criterion function (8.30), divided by σ_0^2. In Section 8.6, we saw that the minimized value of the IV criterion function, divided by an estimate of σ^2, serves as the statistic for the Sargan test for overidentification. We will see in Section 9.4 that the GMM criterion function (9.11), with the usually unknown matrix $\boldsymbol{\Omega}_0$ replaced by a suitable estimate, can also be used as a test statistic for overidentification.

The criterion function (9.11) is a quadratic form in the vector $\boldsymbol{W}^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ of sample moments and the inverse of the matrix $\boldsymbol{W}^\top \boldsymbol{\Omega}_0 \boldsymbol{W}$. Equivalently, it is a quadratic form in $n^{-1/2}\boldsymbol{W}^\top(\boldsymbol{y}-\boldsymbol{X\beta})$ and the inverse of $n^{-1}\boldsymbol{W}^\top\boldsymbol{\Omega}_0\boldsymbol{W}$, since the powers of n cancel. Under the sort of regularity conditions we have used in earlier chapters, $n^{-1/2}\boldsymbol{W}^\top(\boldsymbol{y}-\boldsymbol{X\beta}_0)$ satisfies a central limit theorem, and so tends, as $n\to\infty$, to a normal random variable, with mean vector $\mathbf{0}$ and covariance matrix the limit of $n^{-1}\boldsymbol{W}^\top\boldsymbol{\Omega}_0\boldsymbol{W}$. It follows that (9.11) evaluated using the true $\boldsymbol{\beta}_0$ and the true $\boldsymbol{\Omega}_0$ is asymptotically distributed as χ^2 with l degrees of freedom; recall Theorem 4.1, and see Exercise 9.2.

This property of the GMM criterion function is simply a consequence of its structure as a quadratic form in the sample moments used for estimation and the inverse of the asymptotic covariance matrix of these moments evaluated at the true parameters. As we will see in Section 9.4, this property is what makes the GMM criterion function useful for testing. The argument leading to (9.10) shows that this same property of the GMM criterion function leads to the asymptotic efficiency of the estimator that minimizes it.

Provided the instruments are predetermined, so that they satisfy the condition that $\mathrm{E}(u_t\,|\,\boldsymbol{W}_t)=0$, we still obtain a consistent estimator, even when the matrix $\boldsymbol{J}$ used to select linear combinations of the instruments is different from (9.08). Such a consistent, but in general inefficient, estimator can also be obtained by minimizing a quadratic criterion function of the form

$$(\boldsymbol{y}-\boldsymbol{X\beta})^\top\boldsymbol{W\Lambda W}^\top(\boldsymbol{y}-\boldsymbol{X\beta}), \tag{9.12}$$

where the **weighting matrix** $\boldsymbol{\Lambda}$ is $l\times l$, positive definite, and must be at least asymptotically nonrandom. Without loss of generality, $\boldsymbol{\Lambda}$ can be taken to be symmetric; see Exercise 9.3. The inefficient GMM estimator is

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^\top\boldsymbol{W\Lambda W}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{W\Lambda W}^\top\boldsymbol{y}, \tag{9.13}$$

from which it can be seen that the use of the weighting matrix $\boldsymbol{\Lambda}$ corresponds to the implicit choice $\boldsymbol{J}=\boldsymbol{\Lambda W}^\top\boldsymbol{X}$. For a given choice of $\boldsymbol{J}$, there are various possible choices of $\boldsymbol{\Lambda}$ that give rise to the same estimator; see Exercise 9.4.

When $l=k$, the model is exactly identified, and $\boldsymbol{J}$ is a nonsingular square matrix which has no effect on the estimator. This is most easily seen by looking at the moment conditions (9.05), which are equivalent, when $l=k$, to those obtained by premultiplying them by $(\boldsymbol{J}^\top)^{-1}$. Similarly, if the estimator is defined by minimizing a quadratic form, it does not depend on the choice of $\boldsymbol{\Lambda}$ whenever $l=k$. To see this, consider the first-order conditions for minimizing (9.12), which, up to a scalar factor, are

$$\boldsymbol{X}^\top\boldsymbol{W\Lambda W}^\top(\boldsymbol{y}-\boldsymbol{X\beta})=\mathbf{0}.$$

If $l=k$, $\boldsymbol{X}^\top\boldsymbol{W}$ is a square matrix, and the first-order conditions can be premultiplied by $\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}^\top\boldsymbol{W})^{-1}$. Therefore, the estimator is the solution to the equations $\boldsymbol{W}^\top(\boldsymbol{y}-\boldsymbol{X\beta})=\mathbf{0}$, independently of $\boldsymbol{\Lambda}$. This solution is just the simple IV estimator defined in (8.12).