If $\boldsymbol{X}$ is exogenous, the optimal instruments are given by the matrix $\boldsymbol{\Omega}^{-1}\boldsymbol{X}$, and the moment conditions for efficient estimation are $\mathrm{E}\big(\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\big)=\mathbf{0}$, which can also be written as

$$\mathrm{E}\big(\boldsymbol{X}^\top\boldsymbol{\Psi}\boldsymbol{\Psi}^\top(\boldsymbol{y}-\boldsymbol{X\beta})\big)=\mathbf{0}. \tag{9.22}$$

Comparison with (9.21) shows that the optimal choice of $\boldsymbol{Z}$ is $\boldsymbol{\Psi}^\top\boldsymbol{X}$. Even if $\boldsymbol{X}$ is not exogenous, (9.22) is a correct set of moment conditions if

$$\mathrm{E}\big((\boldsymbol{\Psi}^\top\boldsymbol{u})_t\,|\,(\boldsymbol{\Psi}^\top\boldsymbol{X})_t\big)=0. \tag{9.23}$$

But this is not true in general when $\boldsymbol{X}$ is not exogenous. Consequently, we seek a new definition for $\bar{\boldsymbol{X}}$, such that (9.23) becomes true when $\boldsymbol{X}$ is replaced by $\bar{\boldsymbol{X}}$.

In most cases, it is possible to choose $\boldsymbol{\Psi}$ so that $(\boldsymbol{\Psi}^\top\boldsymbol{u})_t$ is an innovation in the sense of Section 4.5, that is, so that $\mathrm{E}\big((\boldsymbol{\Psi}^\top\boldsymbol{u})_t\,|\,\Omega_t\big)=0$. As an example, see the analysis of models with AR(1) errors in Section 7.8, especially the discussion surrounding (7.58). What is then required for condition (9.23) is that $(\boldsymbol{\Psi}^\top\bar{\boldsymbol{X}})_t$ should be predetermined in period t. If $\boldsymbol{\Omega}$ is diagonal, and so also $\boldsymbol{\Psi}$, the old definition of $\bar{\boldsymbol{X}}$ works, because $(\boldsymbol{\Psi}^\top\bar{\boldsymbol{X}})_t=\Psi_{tt}\bar{\boldsymbol{X}}_t$, where Ψ_{tt} is the t^{th} diagonal element of $\boldsymbol{\Psi}$, and this belongs to Ω_t by construction. If $\boldsymbol{\Omega}$ contains off-diagonal elements, however, the old definition of $\bar{\boldsymbol{X}}$ no longer works in general. Since what we need is that $(\boldsymbol{\Psi}^\top\bar{\boldsymbol{X}})_t$ should belong to Ω_t, we instead define $\bar{\boldsymbol{X}}$ implicitly by the equation

$$\mathrm{E}\big((\boldsymbol{\Psi}^\top\boldsymbol{X})_t\,|\,\Omega_t\big)=(\boldsymbol{\Psi}^\top\bar{\boldsymbol{X}})_t. \tag{9.24}$$

This implicit definition must be implemented on a case-by-case basis. One example is given in Exercise 9.5.

By setting $\boldsymbol{Z}=\boldsymbol{\Psi}^\top\bar{\boldsymbol{X}}$, we find that the moment conditions (9.21) become

$$\mathrm{E}\big(\bar{\boldsymbol{X}}^\top\boldsymbol{\Psi}\boldsymbol{\Psi}^\top(\boldsymbol{y}-\boldsymbol{X\beta})\big)=\mathrm{E}\big(\bar{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\big)=\mathbf{0}. \tag{9.25}$$

These conditions do indeed use $\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{X}}$ as instruments, albeit with a possibly redefined $\bar{\boldsymbol{X}}$. The estimator based on (9.25) is

$$\hat{\boldsymbol{\beta}}_{\mathrm{EGMM}}\equiv(\bar{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\bar{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{y}, \tag{9.26}$$

where EGMM denotes "fully efficient GMM." The asymptotic covariance matrix of (9.26) can be computed using (9.09), in which, on the basis of (9.25), we see that $\boldsymbol{W}$ is to be replaced by $\boldsymbol{\Psi}^\top\bar{\boldsymbol{X}}$, $\boldsymbol{X}$ by $\boldsymbol{\Psi}^\top\boldsymbol{X}$, and $\boldsymbol{\Omega}$ by $\mathbf{I}$. We cannot apply (9.09) directly with instruments $\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{X}}$, because there is no reason to suppose that the result (9.02) holds for the untransformed error terms $\boldsymbol{u}$ and the instruments $\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{X}}$. The result is

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\bar{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\Big(\frac{1}{n}\bar{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{X}}\Big)^{-1}\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{X}}\Big)^{-1}. \tag{9.27}$$

By exactly the same argument as that used in (8.20), we find that, for any matrix $\boldsymbol{Z}$ that satisfies $\boldsymbol{Z}_t \in \Omega_t$,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{\Psi}^\top\boldsymbol{X} = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{\Psi}^\top\bar{\boldsymbol{X}}. \tag{9.28}$$

Since $(\boldsymbol{\Psi}^\top\bar{\boldsymbol{X}})_t \in \Omega_t$, this implies that

$$\begin{aligned}\operatorname*{plim}_{n\to\infty} \frac{1}{n}\bar{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X} &= \operatorname*{plim}_{n\to\infty} \frac{1}{n}\bar{\boldsymbol{X}}^\top\boldsymbol{\Psi}\boldsymbol{\Psi}^\top\boldsymbol{X} \\ &= \operatorname*{plim}_{n\to\infty} \frac{1}{n}\bar{\boldsymbol{X}}^\top\boldsymbol{\Psi}\boldsymbol{\Psi}^\top\bar{\boldsymbol{X}} = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\bar{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{X}}.\end{aligned}$$

Therefore, the asymptotic covariance matrix (9.27) simplifies to

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\bar{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{X}}\Big)^{-1}. \tag{9.29}$$

Although the matrix (9.09) is less of a sandwich than (9.07), the matrix (9.29) is still less of one than (9.09). This is a clear indication of the fact that the instruments $\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{X}}$, which yield the estimator $\hat{\boldsymbol{\beta}}_{\rm EGMM}$, are indeed optimal. Readers are asked to check this formally in Exercise 9.7.

In most cases, $\bar{\boldsymbol{X}}$ is not observed, but it can often be estimated consistently. The usual state of affairs is that we have an $n \times l$ matrix $\boldsymbol{W}$ of instruments, such that $\mathcal{S}(\bar{\boldsymbol{X}}) \subseteq \mathcal{S}(\boldsymbol{W})$ and

$$(\boldsymbol{\Psi}^\top\boldsymbol{W})_t \in \Omega_t. \tag{9.30}$$

This last condition is the form taken by the **predeterminedness condition** when $\boldsymbol{\Omega}$ is not proportional to the identity matrix. The theoretical moment conditions used for (overidentified) estimation are then

$$\mathrm{E}\big(\boldsymbol{W}^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})\big) = \mathrm{E}\big(\boldsymbol{W}^\top\boldsymbol{\Psi}\boldsymbol{\Psi}^\top(\boldsymbol{y}-\boldsymbol{X\beta})\big) = \boldsymbol{0}, \tag{9.31}$$

from which it can be seen that what we are in fact doing is estimating the transformed model (9.20) using the transformed instruments $\boldsymbol{\Psi}^\top\boldsymbol{W}$. The result of Exercise 9.8 shows that, if indeed $\mathcal{S}(\bar{\boldsymbol{X}}) \subseteq \mathcal{S}(\boldsymbol{W})$, the asymptotic covariance matrix of the resulting estimator is still (9.29). Exercise 9.9 investigates what happens if this condition is not satisfied.

The main obstacle to the use of the efficient estimator $\hat{\boldsymbol{\beta}}_{\rm EGMM}$ is thus not the difficulty of estimating $\bar{\boldsymbol{X}}$, but rather the fact that $\boldsymbol{\Omega}$ is usually not known. As with the GLS estimators we studied in Chapter 7, $\hat{\boldsymbol{\beta}}_{\rm EGMM}$ cannot be calculated unless we either know $\boldsymbol{\Omega}$ or can estimate it consistently, usually by knowing the form of $\boldsymbol{\Omega}$ as a function of parameters that can be estimated consistently. But whenever there is heteroskedasticity or serial correlation of unknown form, this is impossible. The best we can then do, asymptotically, is to use the feasible efficient GMM estimator (9.15). Therefore, when we later refer to GMM estimators without further qualification, we will normally mean feasible efficient ones.