$\boldsymbol{A}_0$ by $\boldsymbol{A}(\boldsymbol{y})$, and $\boldsymbol{\Phi}_{ij}$ by expression (17.29) without the probability limit. Although this yields a consistent estimate of (17.30), it is often a very noisy one. We will discuss this issue further in Section 17.5, but it is still far from being completely resolved.

It is interesting to illustrate (17.31) for the case of the IV estimator defined by (17.08). The result will enable us to construct a heteroskedasticity-consistent estimate of the covariance matrix of the latter. We merely have to establish some notational equivalences between the IV case and the more general case discussed above. In the IV case, the elements of the matrix $\boldsymbol{F}$ become $f_{ti} = W_{ti}(y_t - \boldsymbol{X}_t\boldsymbol{\beta})$. Therefore,

$$\boldsymbol{D} = -\operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\boldsymbol{W}^\top\boldsymbol{X}\right) \tag{17.33}$$

and

$$\boldsymbol{A}_0 = \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\boldsymbol{W}^\top\boldsymbol{W}\right)^{-1}. \tag{17.34}$$

The matrix $\boldsymbol{\Phi}$ is obtained from (17.29):

$$\boldsymbol{\Phi} = \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\sum_{t=1}^{n}\big(y_t - \boldsymbol{X}_t\boldsymbol{\beta}\big)^2\boldsymbol{W}_t^\top\boldsymbol{W}_t\right) = \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\boldsymbol{W}^\top\boldsymbol{\Omega}\boldsymbol{W}\right), \tag{17.35}$$

where $\boldsymbol{\Omega}$ is the diagonal matrix with typical element $E(y_t - \boldsymbol{X}_t\boldsymbol{\beta})^2$. By substituting (17.33), (17.34), and (17.35) into (17.31), we obtain the following expression for the asymptotic covariance matrix of the IV estimator:

$$\operatorname*{plim}_{n\to\infty}\left(\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{P}_W\boldsymbol{X}\right)^{-1}\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{P}_W\boldsymbol{\Omega}\boldsymbol{P}_W\boldsymbol{X}\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{P}_W\boldsymbol{X}\right)^{-1}\right). \tag{17.36}$$

The matrix (17.36) is clearly analogous for IV estimation to (16.08) for NLS estimation: It provides the asymptotic covariance matrix in the presence of heteroskedasticity of unknown form. Thus we see that HCCMEs of the sort discussed in Section 16.3 are available for the IV estimator. One can use any of the inconsistent estimators $\hat{\boldsymbol{\Omega}}$ suggested there in order to obtain a consistent estimator of $\operatorname{plim}\big(n^{-1}\boldsymbol{X}^\top\boldsymbol{P}_W\boldsymbol{\Omega}\boldsymbol{P}_W\boldsymbol{X}\big)$.

Readers may reasonably wonder why we have obtained a covariance matrix robust *only* to heteroskedasticity and not also to serial correlation of the error terms. The answer is that the covariance matrix $\boldsymbol{V}$ of (17.30) is valid only if condition CLT is satisfied by the contributions to the empirical moments. That condition will *not* be satisfied if the error terms have an arbitrary pattern of correlation among themselves. In Section 17.5, we will discuss methods for dealing with serial correlation, but these will take us out of the asymptotic framework we have used up to now.

17.3 Efficient GMM Estimators

It is not completely straightforward to answer the question of whether GMM estimators are asymptotically efficient, since a number of separate issues are involved. The first issue was raised at the beginning of the last section, in connection with estimation by instrumental variables. We saw there that, for a given set of empirical moments $\boldsymbol{W}^\top(\boldsymbol{y} - \boldsymbol{X\beta})$, a whole family of estimators can be generated by different choices of the weighting matrix $\boldsymbol{A}(\boldsymbol{y})$ used to construct a quadratic form from the moments. Asymptotically, the most efficient of these estimators is obtained by choosing $\boldsymbol{A}(\boldsymbol{y})$ such that it tends to a nonrandom probability limit proportional to the inverse of the limiting covariance matrix of the empirical moments, suitably weighted by an appropriate power of the sample size n. This turns out to be true quite generally, as we now show.

Theorem 17.3. A Necessary Condition for Efficiency

> A necessary condition for the estimator obtained by minimizing the quadratic form (17.13) to be asymptotically efficient is that it should be asymptotically equal to the estimator defined by minimizing (17.13) with $\boldsymbol{A}(\boldsymbol{y})$ independent of $\boldsymbol{y}$ and equal to the inverse of the asymptotic covariance matrix of the empirical moments $n^{-1/2}\boldsymbol{F}^\top(\boldsymbol{\theta})\boldsymbol{\iota}$.

Note that, when the necessary condition holds, the form of the asymptotic covariance matrix of the GMM estimator $\hat{\boldsymbol{\theta}}$ becomes much simpler. For arbitrary limiting weighting matrix $\boldsymbol{A}_0$, that matrix was given by (17.31). If the necessary condition is satisfied, then $\boldsymbol{A}_0$ in (17.31) may be replaced by the inverse of $\boldsymbol{\Phi}$, which, according to its definition (17.29), is the asymptotic covariance of the empirical moments. Substituting $\boldsymbol{A}_0 = \boldsymbol{\Phi}^{-1}$ into (17.31) gives the simple result that

$$\boldsymbol{V}\big(n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big) = \big(\boldsymbol{D}^\top\boldsymbol{\Phi}^{-1}\boldsymbol{D}\big)^{-1}.$$

Theorem 17.3 will be proved if we can show that, for all symmetric, positive definite matrices $\boldsymbol{A}_0$, the difference

$$\big(\boldsymbol{D}^\top\boldsymbol{A}_0\boldsymbol{D}\big)^{-1}\boldsymbol{D}^\top\boldsymbol{A}_0\boldsymbol{\Phi}\boldsymbol{A}_0\boldsymbol{D}\,\big(\boldsymbol{D}^\top\boldsymbol{A}_0\boldsymbol{D}\big)^{-1} - \big(\boldsymbol{D}^\top\boldsymbol{\Phi}^{-1}\boldsymbol{D}\big)^{-1} \tag{17.37}$$

is positive semidefinite. To show this, we rewrite (17.37) as

$$\big(\boldsymbol{D}^\top\boldsymbol{A}_0\boldsymbol{D}\big)^{-1}\boldsymbol{D}^\top\boldsymbol{A}_0\Big(\boldsymbol{\Phi} - \boldsymbol{D}\big(\boldsymbol{D}^\top\boldsymbol{\Phi}^{-1}\boldsymbol{D}\big)^{-1}\boldsymbol{D}^\top\Big)\boldsymbol{A}_0\boldsymbol{D}\big(\boldsymbol{D}^\top\boldsymbol{A}_0\boldsymbol{D}\big)^{-1}. \tag{17.38}$$

Since the matrix $\boldsymbol{D}^\top\boldsymbol{A}_0\boldsymbol{D}$ is nonsingular, (17.38) is positive definite if the matrix in large parentheses is. Since $\boldsymbol{\Phi}$ is a positive definite, symmetric $l \times l$ matrix, we can find another positive definite, symmetric $l \times l$ matrix $\boldsymbol{\Psi}$ such that $\boldsymbol{\Psi}^2 = \boldsymbol{\Phi}^{-1}$. In terms of $\boldsymbol{\Psi}$, the matrix in large parentheses becomes

$$\boldsymbol{\Psi}^{-1}\big(\mathbf{I} - \boldsymbol{P}_{\Psi D}\big)\boldsymbol{\Psi}^{-1} = \boldsymbol{\Psi}^{-1}\boldsymbol{M}_{\Psi D}\boldsymbol{\Psi}^{-1}, \tag{17.39}$$