The estimator (17.63) was proposed by Hansen (1982) and White and Domowitz (1984), and was used in some of the earlier published work that employed GMM estimation, such as Hansen and Singleton (1982). From the point of view of theory, it is necessary to let the truncation parameter p, usually referred to as the **lag truncation parameter**, go to infinity at some suitable rate. A typical rate would be $n^{1/4}$, in which case $p = o(n^{1/4})$. This ensures that, for large enough n, all the nonzero $\boldsymbol{\Gamma}(j)$'s are estimated consistently. Unfortunately, this type of result is not of much use in practice, where one typically faces a given, finite n. We will return to this point a little later, and for the meantime suppose simply that we have somehow selected an appropriate value for p.

A much more serious difficulty associated with (17.63) is that, in finite samples, it need not be positive definite or even positive semidefinite. If one is unlucky enough to be working with a data set that yields a nondefinite $\hat{\boldsymbol{\Phi}}$, then (17.63) is unusable. There are numerous ways out of this difficulty. The most widely used was suggested by Newey and West (1987a). It is simply to multiply the $\hat{\boldsymbol{\Gamma}}(j)$'s by a sequence of weights that decrease as $|j|$ increases. Specifically, the estimator that they propose is

$$\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{\Gamma}}(0) + \sum_{j=1}^{p} \left(1 - \frac{j}{p+1}\right) \left(\hat{\boldsymbol{\Gamma}}(j) + \hat{\boldsymbol{\Gamma}}(j)^{\top}\right). \tag{17.64}$$

It can be seen that the weights $1 - j/(p+1)$ decrease linearly with j from a value of 1 for $\hat{\boldsymbol{\Gamma}}(0)$ by steps of $1/(p+1)$ down to a value of $1/(p+1)$ for $|j| = p$. The use of such a set of weights is clearly compatible with the idea that the impact of the autocovariance of order j diminishes with $|j|$.

We will not attempt even to sketch a proof of the consistency of the Newey-West or similar estimators. We have alluded to the sort of regularity conditions needed for consistency to hold: Basically, the autocovariance matrices of the empirical moments must tend to zero quickly enough as p increases. It would also go well beyond the scope of this book to provide a theoretical justification for the Newey-West estimator. It rests on considerations of the so-called "frequency domain representation" of the $\boldsymbol{F}_t$'s and also of a number of notions associated with nonparametric estimation procedures. Interested readers are referred to Andrews (1991b) for a rather complete treatment of many of the issues. This paper suggests some alternatives to the Newey-West estimator and shows that in some circumstances they are preferable. However, the performance of the Newey-West estimator is never greatly inferior to that of the alternatives. Consequently, its simplicity is much in its favor.

Let us now return to the linear IV model with empirical moments given by $\boldsymbol{W}^{\top}(\boldsymbol{y} - \boldsymbol{X\beta})$. In order to be able to use (17.64), we suppose that the true error terms $u_t \equiv y_t - \boldsymbol{X}_t\boldsymbol{\beta}_0$ satisfy an appropriate mixing condition. Then the sample autocovariance matrices $\hat{\boldsymbol{\Gamma}}(j)$ for $j = 0, \ldots, p$, for some given p, are calculated as follows. A preliminary consistent estimate of $\boldsymbol{\beta}$ is first obtained