where $X_{ti} \equiv X_{ti}(\boldsymbol{\beta}_0)$. On the other hand, a typical element of (16.10) is

$$\frac{1}{n}\sum_{t=1}^{n} \hat{u}_t^2 \hat{X}_{ti} \hat{X}_{tj}. \tag{16.12}$$

Because $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$, $\hat{u}_t$ is consistent for u_t, $\hat{u}_t^2$ is consistent for u_t^2, and $\hat{X}_{ti}$ is consistent for X_{ti}. Thus expression (16.12) is asymptotically equal to

$$\frac{1}{n}\sum_{t=1}^{n} u_t^2 X_{ti} X_{tj}. \tag{16.13}$$

Under our assumptions, we can apply a law of large numbers to (16.13); see White (1980, 1984) and Nicholls and Pagan (1983) for some technical details. It follows immediately that (16.13), and so also (16.12), tends in probability to (16.11). Consequently, the matrix

$$\big(n^{-1}\hat{\boldsymbol{X}}^{\top}\hat{\boldsymbol{X}}\big)^{-1}\big(n^{-1}\hat{\boldsymbol{X}}^{\top}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{X}}\big)\big(n^{-1}\hat{\boldsymbol{X}}^{\top}\hat{\boldsymbol{X}}\big)^{-1} \tag{16.14}$$

consistently estimates (16.09). Of course, in practice one ignores the factors of n^{-1} and uses the matrix

$$\big(\hat{\boldsymbol{X}}^{\top}\hat{\boldsymbol{X}}\big)^{-1}\hat{\boldsymbol{X}}^{\top}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{X}}\big(\hat{\boldsymbol{X}}^{\top}\hat{\boldsymbol{X}}\big)^{-1} \tag{16.15}$$

to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$.

Asymptotically valid inferences about $\boldsymbol{\beta}$ may be based on the HCCME (16.15) in the usual way. However, one must be cautious when n is not large. There is a good deal of evidence that this HCCME is somewhat unreliable in finite samples. After all, the fact that (16.14) estimates (16.09) consistently does not imply that the former always estimates the latter very well in finite samples.

It is possible to modify the HCCME (16.15) so that it has better finite-sample properties. The major problem is that the squared least squares residuals $\hat{u}_t^2$ are not unbiased estimates of the squared error terms u_t^2. The easiest way to improve the HCCME is simply to multiply (16.15) by $n/(n-k)$. This is analogous to dividing the sum of squared residuals by $n-k$ rather than n to obtain the OLS variance estimator s^2. A second, and better, approach is to define the t^{th} diagonal element of $\hat{\boldsymbol{\Omega}}$ as $\hat{u}_t^2/(1-\hat{h}_t)$, where $\hat{h}_t \equiv \hat{\boldsymbol{X}}_t(\hat{\boldsymbol{X}}^{\top}\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}_t^{\top}$ is the t^{th} diagonal element of the "hat" matrix $\hat{\boldsymbol{P}}_X$ that projects orthogonally onto the space spanned by the columns of $\hat{\boldsymbol{X}}$. Recall from Section 3.2 that, in the OLS case with constant variance σ^2, the expectation of $\hat{u}_t^2$ is $\sigma^2(1-h_t)$. Thus, in the linear case, dividing $\hat{u}_t^2$ by $1-h_t$ would yield an unbiased estimate of σ^2 if the error terms were actually homoskedastic.

A third possibility is to use a technique called the "jackknife" that we will not attempt to discuss here; see MacKinnon and White (1985). The resulting