has mean zero and is uncorrelated with $\hat{\boldsymbol{\beta}}$. The random component simply adds noise to the efficient estimator $\hat{\boldsymbol{\beta}}$. This makes it clear that $\hat{\boldsymbol{\beta}}$ is more efficient than $\tilde{\boldsymbol{\beta}}$. To complete the proof, we note that

$$\begin{aligned} \operatorname{Var}(\tilde{\boldsymbol{\beta}}) &= \operatorname{Var}\big(\hat{\boldsymbol{\beta}} + (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})\big) \\ &= \operatorname{Var}\big(\hat{\boldsymbol{\beta}} + \boldsymbol{C}\boldsymbol{y}\big) \\ &= \operatorname{Var}(\hat{\boldsymbol{\beta}}) + \operatorname{Var}(\boldsymbol{C}\boldsymbol{y}), \end{aligned} \tag{3.41}$$

because, from (3.40), the covariance of $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{C}\boldsymbol{y}$ is zero. Thus the difference between $\operatorname{Var}(\tilde{\boldsymbol{\beta}})$ and $\operatorname{Var}(\hat{\boldsymbol{\beta}})$ is $\operatorname{Var}(\boldsymbol{C}\boldsymbol{y})$. Since it is a covariance matrix, this difference is necessarily positive semidefinite. ■

We will encounter many cases in which an inefficient estimator is equal to an efficient estimator plus a random variable that has mean zero and is uncorrelated with the efficient estimator. The zero correlation ensures that the covariance matrix of the inefficient estimator is equal to the covariance matrix of the efficient estimator plus another matrix that is positive semidefinite, as in the last line of (3.41). If the correlation were not zero, this sort of proof would not work. Observe that, because everything is done in terms of second moments, the Gauss-Markov Theorem does not require any assumption about the normality of the error terms.

The Gauss-Markov Theorem that the OLS estimator is BLUE is one of the most famous results in statistics. However, it is important to keep in mind the limitations of this theorem. The theorem applies only to a correctly specified model with exogenous regressors and error terms that are homoskedastic and serially uncorrelated. Moreover, it does *not* say that the OLS estimator $\hat{\boldsymbol{\beta}}$ is more efficient than every imaginable estimator. Estimators which are nonlinear and/or biased may well perform better than ordinary least squares.

3.6 Residuals and Error Terms

The vector of least-squares residuals, $\hat{\boldsymbol{u}} \equiv \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$, is easily calculated once we have obtained $\hat{\boldsymbol{\beta}}$. The numerical properties of $\hat{\boldsymbol{u}}$ were discussed in Section 2.3. These properties include the fact that $\hat{\boldsymbol{u}}$ is orthogonal to $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ and to every vector that lies in $\mathcal{S}(\boldsymbol{X})$. In this section, we turn our attention to the statistical properties of $\hat{\boldsymbol{u}}$ as an estimator of $\boldsymbol{u}$. These properties are very important, because we will want to use $\hat{\boldsymbol{u}}$ for a number of purposes. In particular, we will want to use it to estimate σ^2, the variance of the error terms. We need an estimate of σ^2 if we are to obtain an estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$. As we will see in later chapters, the residuals can also be used to test some of the strong assumptions that are often made about the distribution of the error terms and to implement more sophisticated estimation methods that require weaker assumptions.