fundamental principles on which most hypothesis tests are based: the Wald, Lagrange multiplier, and likelihood ratio principles. Finally, in Section 3.7, we discuss the effects of imposing incorrect restrictions and introduce the notion of preliminary test estimators.

3.2 Covariance Matrix Estimation

In the case of the linear regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{IID}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{3.02}$$

it is well known that when the DGP satisfies (3.02) for specific parameter values $\boldsymbol{\beta}_0$ and σ_0, the covariance matrix of the vector of OLS estimates $\hat{\boldsymbol{\beta}}$ is

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}}) = \sigma_0^2\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}. \tag{3.03}$$

The proof of this familiar result is quite straightforward. The covariance matrix $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ is defined as the expectation of the outer product of $\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})$ with itself, conditional on the independent variables $\boldsymbol{X}$. Starting with this definition and using the fact that $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}_0$, we first replace $\hat{\boldsymbol{\beta}}$ by what it is equal to under the DGP, then take expectations conditional on $\boldsymbol{X}$, and finally simplify the algebra to obtain (3.03):

$$\begin{aligned}
\boldsymbol{V}(\hat{\boldsymbol{\beta}}) &\equiv E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \\
&= E\big((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y} - \boldsymbol{\beta}_0\big)\big((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y} - \boldsymbol{\beta}_0\big)^\top \\
&= E\big((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{X\beta}_0 + \boldsymbol{u}) - \boldsymbol{\beta}_0\big)\big((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{X\beta}_0 + \boldsymbol{u}) - \boldsymbol{\beta}_0\big)^\top \\
&= E\big(\boldsymbol{\beta}_0 + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u} - \boldsymbol{\beta}_0\big)\big(\boldsymbol{\beta}_0 + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u} - \boldsymbol{\beta}_0\big)^\top \\
&= E\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{u}\boldsymbol{u}^\top\boldsymbol{X}\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1} \\
&= \big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\big(\sigma_0^2\mathbf{I}\big)\boldsymbol{X}\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1} \\
&= \sigma_0^2\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{X}\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1} \\
&= \sigma_0^2\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}.
\end{aligned}$$

Deriving an analogous result for the nonlinear regression model (3.01) requires a few concepts of asymptotic analysis that we have not yet developed, plus a certain amount of mathematical manipulation. We will therefore postpone this derivation until Chapter 5 and merely state an approximate result here.

For a nonlinear model, we cannot in general obtain an exact expression for $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ in the finite-sample case. In Chapter 5, on the assumption that the data are generated by a DGP which is a special case of (3.01), we will, however, obtain an asymptotic result which allows us to state that

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}}) \cong \sigma_0^2\big(\boldsymbol{X}^\top(\boldsymbol{\beta}_0)\boldsymbol{X}(\boldsymbol{\beta}_0)\big)^{-1}, \tag{3.04}$$