$\mathrm{E}\big(\eta(x)\big) \neq \eta\big(\mathrm{E}(x)\big)$. Thus, it is often very easy to calculate plims in circumstances where it would be difficult or impossible to calculate expectations.

However, working with plims can be a little bit tricky. The problem is that many of the stochastic quantities we encounter in econometrics do not have probability limits unless we divide them by n or, perhaps, by some power of n. For example, consider the matrix $\boldsymbol{X}^\top\boldsymbol{X}$, which appears in the formula (3.04) for $\hat{\boldsymbol{\beta}}$. Each element of this matrix is a scalar product of two of the columns of $\boldsymbol{X}$, that is, two n–vectors. Thus it is a sum of n numbers. As $n \to \infty$, we would expect that, in most circumstances, such a sum would tend to infinity as well. Therefore, the matrix $\boldsymbol{X}^\top\boldsymbol{X}$ does not generally have a plim. However, it is not at all unreasonable to assume that

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X} = \boldsymbol{S}_{\boldsymbol{X}^\top\boldsymbol{X}}, \tag{3.17}$$

where $\boldsymbol{S}_{\boldsymbol{X}^\top\boldsymbol{X}}$ is a finite nonstochastic matrix with full rank k, because each element of the matrix on the left-hand side of equation (3.17) is now an average of n numbers:

$$\Big(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\Big)_{ij} = \frac{1}{n}\sum_{t=1}^{n} x_{ti}\,x_{tj}.$$

In effect, when we write (3.17), we are implicitly making some assumption sufficient for a LLN to hold for the sequences generated by the squares of the regressors and their cross-products. Thus there should not be too much dependence between $x_{ti}x_{tj}$ and $x_{si}x_{sj}$ for $s \neq t$, and the variances of these quantities should not differ too much as t and s vary.

The OLS Estimator Is Consistent

We can now show that, under plausible assumptions, the least-squares estimator $\hat{\boldsymbol{\beta}}$ is consistent. When the DGP is a special case of the regression model (3.03) that is being estimated, we saw in (3.05) that

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}. \tag{3.18}$$

To demonstrate that $\hat{\boldsymbol{\beta}}$ is consistent, we need to show that the second term on the right-hand side here has a plim of zero. This term is the product of two matrix expressions, $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ and $\boldsymbol{X}^\top\boldsymbol{u}$. Neither $\boldsymbol{X}^\top\boldsymbol{X}$ nor $\boldsymbol{X}^\top\boldsymbol{u}$ has a probability limit. However, we can divide both of these expressions by n without changing the value of this term, since $n \cdot n^{-1} = 1$. By doing so, we convert them into quantities that, under reasonable assumptions, have nonstochastic plims. Thus the plim of the second term in (3.18) becomes

$$\Big(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\Big)^{-1} \operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{u} = \big(\boldsymbol{S}_{\boldsymbol{X}^\top\boldsymbol{X}}\big)^{-1} \operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{u} = \boldsymbol{0}. \tag{3.19}$$