It seems reasonable that, if $\hat{\beta}_2$ is biased, so must be $\hat{\beta}_1$. The equivalent of the second line of (3.12) is

$$\hat{\beta}_1 = \beta_{10} + (\boldsymbol{\iota}^\top \boldsymbol{M}_{\boldsymbol{y}_1} \boldsymbol{\iota})^{-1} \boldsymbol{\iota}^\top \boldsymbol{M}_{\boldsymbol{y}_1} \boldsymbol{u}, \tag{3.13}$$

where the notation should be self-explanatory. Once again, because $\boldsymbol{y}_1$ depends on $\boldsymbol{u}$, we cannot employ the methods that we used in (3.07) or (3.09) to prove that the second term on the right-hand side of (3.13) has mean zero. In fact, it does not have mean zero, and $\hat{\beta}_1$ is consequently biased, as readers are also asked to demonstrate in Exercise 3.1.

The problems we have just encountered when dealing with the autoregressive model (3.11) evidently affect every regression model with random regressors for which the exogeneity assumption (3.08) does not hold. Thus, for all such models, the least-squares estimator of the parameters of the regression function is biased. Assumption (3.08) cannot possibly hold when the regressor matrix $\boldsymbol{X}$ contains lagged dependent variables, and it probably fails to hold for most other models that involve time-series data.

## 3.3 Are OLS Parameter Estimators Consistent?

Unbiasedness is by no means the only desirable property that we would like an estimator to possess. Another very important property is **consistency**. A **consistent** estimator is one for which the estimate tends to the quantity being estimated as the size of the sample tends to infinity. Thus, if the sample size is large enough, we can be confident that the estimate is close to the true value. Happily, the least-squares estimator $\hat{\boldsymbol{\beta}}$ is often consistent even when it is biased.

In order to define consistency, we have to specify what it means for the sample size $n$ to tend to infinity or, in more compact notation, $n \to \infty$. At first sight, this may seem like a very odd notion. After all, any given data set contains a fixed number of observations. Nevertheless, we can certainly imagine simulating data and letting $n$ become arbitrarily large. In the case of a pure time-series model like (3.11), we can easily generate any sample size we want, just by letting the simulations run on for long enough. In the case of a model with cross-section data, we can pretend that the original sample is taken from a population of infinite size, and we can imagine drawing more and more observations from that population. Even in the case of a model with fixed regressors, we can think of ways to make $n$ tend to infinity. Suppose that the original $\boldsymbol{X}$ matrix is of dimension $m \times k$. Then we can create $\boldsymbol{X}$ matrices of dimensions $2m \times k$, $3m \times k$, $4m \times k$, and so on, simply by stacking as many copies of the original $\boldsymbol{X}$ matrix as we like. By simulating error vectors of the appropriate dimension, we can then generate $n$–vectors $\boldsymbol{y}$ for any $n$ that is an integer multiple of $m$. Thus, in all these cases, we can reasonably think of letting $n$ tend to infinity.

### Probability Limits

In order to say what happens to a stochastic quantity that depends on $n$ as $n \to \infty$, we need to introduce the concept of a **probability limit**. The probability limit, or **plim** for short, generalizes the ordinary concept of a limit to quantities that are stochastic. If $\boldsymbol{a}(\boldsymbol{y}^n)$ is some vector function of the random vector $\boldsymbol{y}^n$, and the plim of $\boldsymbol{a}(\boldsymbol{y}^n)$ as $n \to \infty$ is $\boldsymbol{a}_0$, we may write

$$\operatorname*{plim}_{n\to\infty} \boldsymbol{a}(\boldsymbol{y}^n) = \boldsymbol{a}_0. \tag{3.14}$$

We have written $\boldsymbol{y}^n$ here, instead of just $\boldsymbol{y}$, to emphasize the fact that $\boldsymbol{y}^n$ is an $n$–vector and that the value of $n$ is not fixed. The superscript is often omitted in practice. In econometrics, we are almost always interested in taking probability limits as $n \to \infty$. Thus, when there can be no ambiguity, we will often simply use notation like $\operatorname{plim} \boldsymbol{a}(\boldsymbol{y})$ rather than more precise notation like that of (3.14).

Formally, the random vector $\boldsymbol{a}(\boldsymbol{y}^n)$ tends in probability to the limiting random vector $\boldsymbol{a}_0$ if, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\big(\|\boldsymbol{a}(\boldsymbol{y}^n) - \boldsymbol{a}_0\| < \varepsilon\big) = 1. \tag{3.15}$$

Here $\|\cdot\|$ denotes the Euclidean norm of a vector (see Section 2.2), which simplifies to the absolute value when its argument is a scalar. Condition (3.15) says that, for any specified tolerance level $\varepsilon$, no matter how small, the probability that the norm of the discrepancy between $\boldsymbol{a}(\boldsymbol{y}^n)$ and $\boldsymbol{a}_0$ is less than $\varepsilon$ goes to unity as $n \to \infty$.

Although the probability limit $\boldsymbol{a}_0$ was defined above to be a random variable (actually, a vector of random variables), it may in fact be an ordinary nonrandom vector or scalar, in which case it is said to be nonstochastic. Many of the plims that we will encounter in this book are in fact nonstochastic. A simple example of a **nonstochastic plim** is the limit of the proportion of heads in a series of independent tosses of an unbiased coin. Suppose that $y_t$ is a random variable equal to 1 if the coin comes up heads, and equal to 0 if it comes up tails. After $n$ tosses, the proportion of heads is just

$$p(\boldsymbol{y}^n) \equiv \frac{1}{n}\sum_{t=1}^{n} y_t.$$

If the coin really is unbiased, $\mathrm{E}(y_t) = {}^1\!/_2$. Thus it should come as no surprise to learn that $\operatorname{plim} p(\boldsymbol{y}^n) = {}^1\!/_2$. Proving this requires a certain amount of effort, however, and we will therefore not attempt a proof here. For a detailed discussion and proof, see Davidson and MacKinnon (1993, Section 4.2).

The coin-tossing example is really a special case of an extremely powerful result in probability theory, which is called a **law of large numbers**, or **LLN**.