employed. We know that variables which are $I(1)$ tend to diverge as $n \to \infty$, because their unconditional variances are proportional to n. Thus it might seem that such variables could never be expected to obey any sort of long-run equilibrium relationship. But in fact it is possible for two or more variables to be $I(1)$ and yet for certain linear combinations of those variables to be $I(0)$. If that is the case, the variables are said to be **cointegrated**. If two or more variables are cointegrated, they must obey an equilibrium relationship in the long run, although they may diverge substantially from equilibrium in the short run. The concept of cointegration is fundamental to the understanding of long-run relationships among economic time series. It is also quite recent. The earliest reference is probably Granger (1981), the best-known paper is Engle and Granger (1987), and two relatively accessible articles are Hendry (1986) and Stock and Watson (1988a).

Suppose, to keep matters simple, that we are concerned with just two variables, y_{t1} and y_{t2}, each of which is known to be $I(1)$. Then, in the simplest case, y_{t1} and y_{t2} would be cointegrated if there exists a vector $\boldsymbol{\eta} \equiv [1 \quad -\eta_2]^\top$ such that, when the two variables are in equilibrium,

$$[\boldsymbol{y}_1 \quad \boldsymbol{y}_2]\boldsymbol{\eta} \equiv \boldsymbol{y}_1 - \eta_2 \boldsymbol{y}_2 = \boldsymbol{0}. \tag{20.20}$$

Here $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ denote n–vectors with typical elements y_{t1} and y_{t2}, respectively. The 2–vector $\boldsymbol{\eta}$ is called a **cointegrating vector**. It is clearly not unique, since it could be multiplied by any nonzero scalar without affecting the equality in (20.20).

Realistically, one might well expect y_{t1} and y_{t2} to be changing systematically as well as stochastically over time. Thus one might expect (20.20) to contain a constant term and perhaps one or more trend terms as well. If we write $\boldsymbol{Y} = [\boldsymbol{y}_1 \quad \boldsymbol{y}_2]$, (20.20) can be rewritten to allow for this possibility as

$$\boldsymbol{Y}\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}, \tag{20.21}$$

where, as in (20.14), $\boldsymbol{X}$ denotes a nonstochastic matrix that may or may not have any elements. If it does, the first column will be a constant, the second, if it exists, will be a linear time trend, the third, if it exists, will be a quadratic time trend, and so on. Since $\boldsymbol{Y}$ could contain more than two variables, (20.21) is actually a very general way of writing a cointegrating relationship among any number of variables.

At any particular time t, of course, an equality like (20.20) or (20.21) cannot be expected to hold exactly. We may therefore define the **equilibrium error** ν_t as

$$\nu_t = \boldsymbol{Y}_t\boldsymbol{\eta} - \boldsymbol{X}_t\boldsymbol{\beta}, \tag{20.22}$$

where $\boldsymbol{Y}_t$ and $\boldsymbol{X}_t$ denote the t^{th} rows of $\boldsymbol{Y}$ and $\boldsymbol{X}$, respectively. In the special case of (20.20), this equilibrium error would simply be $y_{t1} - \eta_2 y_{t2}$. The m variables y_{t1} through y_{tm} are said to be cointegrated if there exists a vector $\boldsymbol{\eta}$ such that ν_t in (20.22) is $I(0)$.