A univariate AR model has a unit root if the coefficient on the lagged dependent variable is equal to unity. Analogously, as we now show, the VAR model (14.37) has a unit root if an eigenvalue of the matrix $\boldsymbol{\Phi}$ is equal to 1.

Recall from Section 12.8 that the matrix $\boldsymbol{\Phi}$ has an eigenvalue λ and corresponding eigenvector $\boldsymbol{\xi}$ if $\boldsymbol{\Phi\xi} = \lambda\boldsymbol{\xi}$. For a 2×2 matrix, there are two eigenvalues, λ_1 and λ_2. If $\lambda_1 \neq \lambda_2$, there are two corresponding eigenvectors, $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, which are linearly independent; see Exercise 14.18. If $\lambda_1 = \lambda_2$, we assume, with only a slight loss of generality, that there still exist two linearly independent eigenvectors $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$. Then, as in equation (12.117), we can write

$$\boldsymbol{\Phi\Xi} = \boldsymbol{\Xi\Lambda}, \text{ with } \boldsymbol{\Xi} \equiv [\boldsymbol{\xi}_1 \quad \boldsymbol{\xi}_2] \text{ and } \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

It follows that $\boldsymbol{\Phi}^2\boldsymbol{\Xi} = \boldsymbol{\Phi}(\boldsymbol{\Phi\Xi}) = \boldsymbol{\Phi\Xi\Lambda} = \boldsymbol{\Xi\Lambda}^2$. Performing this operation repeatedly shows that, for any positive integer s, $\boldsymbol{\Phi}^s\boldsymbol{\Xi} = \boldsymbol{\Xi\Lambda}^s$.

The solution (14.39) can be rewritten in terms of the eigenvalues and eigenvectors of $\boldsymbol{\Phi}$ as follows:

$$\boldsymbol{\Xi}^{-1}\boldsymbol{z}_t = \sum_{s=1}^{t} \boldsymbol{\Lambda}^{t-s}\boldsymbol{\Xi}^{-1}\boldsymbol{u}_s. \tag{14.40}$$

The inverse matrix $\boldsymbol{\Xi}^{-1}$ exists because $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are linearly independent. It is then not hard to show that the solution (14.40) can be written as

$$\begin{aligned} y_{t1} &= \xi_{11}\sum_{s=1}^{t}\lambda_1^{t-s}e_{s1} + \xi_{12}\sum_{s=1}^{t}\lambda_2^{t-s}e_{s2}, \\ y_{t2} &= \xi_{21}\sum_{s=1}^{t}\lambda_1^{t-s}e_{s1} + \xi_{22}\sum_{s=1}^{t}\lambda_2^{t-s}e_{s2}, \end{aligned} \tag{14.41}$$

where $\boldsymbol{e}_t \equiv [e_{t1} \vdots e_{t2}] \sim \text{IID}(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} \equiv \boldsymbol{\Xi}^{-1}\boldsymbol{\Omega}(\boldsymbol{\Xi}^\top)^{-1}$, and ξ_{ij} is the ij^{th} element of $\boldsymbol{\Xi}$.

It can be seen from equations (14.41) that the series y_{t1} and y_{t2} are both linear combinations of the two series

$$v_{t1} \equiv \sum_{s=1}^{t}\lambda_1^{t-s}e_{s1} \quad \text{and} \quad v_{t2} \equiv \sum_{s=1}^{t}\lambda_2^{t-s}e_{s2}. \tag{14.42}$$

If both eigenvalues are less than 1 in absolute value, then v_{t1} and v_{t2} are I(0). If both eigenvalues are equal to 1, then the two series are random walks, and consequently y_{t1} and y_{t2} are I(1). If one eigenvalue, say λ_1, is equal to 1 while the other is less than 1 in absolute value, then v_{t1} is a random walk, and v_{t2} is I(0). In this case, there is a linear combination of y_{t1} and y_{t2}, namely v_{t2}, that is I(0), even though both y_{t1} and y_{t2} are I(1) unless one of them is actually proportional to v_{t2}. According to the definition we gave above, y_{t1} and y_{t2} are cointegrated in this case. Each differs from a multiple of the random walk v_{t1} by a process that, being I(0), does not diverge and has a finite variance as $t \to \infty$.

Quite generally, if the series y_{t1} and y_{t2} are cointegrated, then there exists a 2–vector $\boldsymbol{\eta}$ with elements η_1 and η_2 such that

$$\nu_t \equiv \boldsymbol{\eta}^\top \boldsymbol{z}_t = \eta_1 y_{t1} + \eta_2 y_{t2} \tag{14.43}$$

is I(0). The vector $\boldsymbol{\eta}$ is called a **cointegrating vector**. It is clearly not unique, since it could be multiplied by any nonzero scalar without affecting anything except the sign and the scale of ν_t.

Equation (14.43) is an example of a **cointegrating regression**. This particular one is unnecessarily restrictive. In practice, we might expect the relationship between y_{t1} and y_{t2} to change gradually over time. We can allow for this by adding a constant term and, perhaps, one or more trend terms, so as to obtain

$$\boldsymbol{\eta}^\top \boldsymbol{z}_t = \boldsymbol{X}_t \boldsymbol{\gamma} + \nu_t, \tag{14.44}$$

where $\boldsymbol{X}_t$ denotes a deterministic row vector that may or may not have any elements. If it does, the first element is a constant, the second, if it exists, is normally a linear time trend, the third, if it exists, is normally a quadratic time trend, and so on. There could also be seasonal dummy variables in $\boldsymbol{X}_t$. Since $\boldsymbol{z}_t$ could contain more than two variables, equation (14.44) is actually a very general way of writing a cointegrating regression. The error term $\nu_t = \boldsymbol{\eta}^\top \boldsymbol{z}_t - \boldsymbol{X}_t \boldsymbol{\gamma}$ that is implicitly defined in equation (14.44) is called the **equilibrium error**.

Unless each of a set of cointegrated variables is I(1), the cointegrating vector is trivial, since it has only one nonzero element, namely, the one that corresponds to the I(0) variable. Therefore, before estimating equations like (14.43) and (14.44), it is customary to test the null hypothesis that each of the series in $\boldsymbol{z}_t$ has a unit root. If this hypothesis is rejected for any of the series, it is pointless to retain it in the set of possibly cointegrated variables.

When there are more than two variables involved, there may be more than one cointegrating vector. For the remainder of this section, however, we will focus on the case in which there is just one such vector. The more general case, in which there are g variables and up to $g-1$ cointegrating vectors, will be discussed in the next section.

It is not entirely clear how to specify the deterministic vector $\boldsymbol{X}_t$ in a cointegrating regression like (14.44). Ordinary t and F tests are not valid, partly because the stochastic regressors are not I(0) and any trending regressors do not satisfy the usual conditions for the matrix $n^{-1}\boldsymbol{X}^\top \boldsymbol{X}$ to tend to a positive definite matrix as $n \to \infty$, and partly because the error terms are likely to display serial correlation. As with unit root tests, investigators commonly use several choices for $\boldsymbol{X}_t$ and present several sets of results.