2. Multivariate Time Series Analysis

To apply standard estimation or testing procedures in a dynamic time series model, it is typically required that the various variables are stationary, since the majority of economic theory is built upon the assumption of stationarity. For example, regressing a non-stationary variable y_t upon a non-stationary variable x_t may lead to the so-called **spurious regression**, in which estimators and test statistics are misleading. An important exception arises when two or more $I(1)$ variables are **cointegrated**, that is, if there exists a particular linear combination of these non-stationary variables that is ***stationary***. In such cases a long-run relationship between these variables exists. Often economic theory suggests the existence of such long-run or equilibrium relationships. The existence of a long-run relationship also has its implications for the short-run behavior of the $I(1)$ variables, because there also has to be some mechanism that derives the variables to their long-run equilibrium relationship. This mechanism is modelled by an **error correction mechanism**, in which the 'equilibrium error' also derives the short-run dynamics of the series.

Another starting point of multivariate time series analysis is the multivariate generalization of the *ARMA* process where particular emphasis is placed on vector autoregressive models. The existence of cointegrating relationship between variables in the VAR has important implications on the way it can be estimated and represented. This will be followed by discussion on how hypothesis regarding the number of cointegrating relationships can be tested and how an error correction model representing the data can be estimated.

2.1 *Dynamic Models with Stationary Variables*

Considering an economic time series in isolation and applying the techniques of univariate time series may provide good forecasts in many cases. However, it does not allow us to determine what the effects are of, for example, a change in policy variable. To do so, it is possible to include additional variables in the model. Consider two stationary variables y_t and x_t, and assume it holds that

$$y_t = \alpha + \phi y_{t-1} + \theta_0 x_t + \theta_1 x_{t-1} + \epsilon_t \qquad [2.1.1]$$

If we assume that ϵ_t is a white noise process, independent of $x_t, x_{t-1}, \cdots$ and $y_{t-1}, y_{t-2}, \cdots$ the above model is sometimes referred to as an **autoregressive distributed lag** model. To estimate it consistently, we can simply use ordinary least squares. What is interesting in [2.1.1]

is that it describes the dynamic effects of a change in x_t upon current and future values of y_t. Taking partial derivatives, we can derive that the immediate response is given by

$$\partial y_t / \partial x_t = \theta_0 \qquad [2.1.2]$$

Sometimes this is referred to as the **impact multiplier**. The effect after one period is

$$\partial y_{t+1} / \partial x_t = \phi \partial y_t / \partial x_t + \theta_1 = \phi\theta_0 + \theta_1, \qquad [2.1.3]$$

and after two periods

$$\partial y_{t+2} / \partial x_t = \phi \partial y_{t+1} / \partial x_t = \phi(\phi\theta_0 + \theta_1) \qquad [2.1.4]$$

and so on. This shows that after the first period, the effect is decreasing if $|\phi| < 1$. Imposing this so-called stability condition allows us to determine the long-run effect of a unit change in x_t. Thus, the **long-run multiplier** (or equilibrium multiplier) is given by

$$\theta_0 + (\phi\theta_0 + \theta_1) + \phi(\phi\theta_0 + \theta_1) + \cdots = \theta_0 + (1 + \phi + \phi^2 + \cdots)(\phi\theta_0 + \theta_1) = \frac{\theta_0 + \theta_1}{1 - \phi}. \qquad [2.1.5]$$

If the increase in x_t is permanent, the long-run multiplier also has the interpretation of the expected long-run permanent increase in y_t. From [2.1.1] the long-run equilibrium relation between y_t and x_t is given by (imposing $E(y_t) = E(y_{t-1})$ and $E(x_t) = E(x_{t-1})$):

$$E(y_t) = \alpha + \phi E(y_t) + \theta_0 E(x_t) + \theta_1 E(x_t) \qquad [2.1.6]$$

Or

$$E(y_t) = \frac{\alpha}{1 - \phi} + \left(\frac{\theta_0 + \theta_1}{1 - \phi} \right) E(x_t), \qquad [2.1.7]$$

which represents an alternative derivation of the long-run multiplier. There is an alternative way to formulate the autoregressive distributed lag model in [2.1.1]. Subtracting y_{t-1} from both sides of [2.1.1] and some rewriting gives

$$\begin{aligned} \Delta y_t &= \alpha + (\phi - 1) y_{t-1} + \theta_0 \Delta x_t + (\theta_0 + \theta_1) x_{t-1} + \epsilon_t \\ \Delta y_t &= \theta_0 \Delta x_t - (1 - \phi) \left[y_{t-1} - \frac{\alpha}{1 - \phi} - \left(\frac{\theta_0 + \theta_1}{1 - \phi} \right) x_{t-1} \right] + \epsilon_t \end{aligned} \qquad [2.1.8]$$

This formulation is an example of an **error-correction model**. It says that the change in y_t is due to the current change in x_t plus an error correction term. If y_{t-1} is above the equilibrium value that corresponds to x_{t-1}, i.e., if the 'equilibrium error' in square brackets is positive, an

additional negative adjustment in y_t is generated. The speed of adjustment is determined by $1-\phi$, which is the adjustment parameter. Assuming stability ensures that $1-\phi>0$. It is also possible to consistently estimate the error-correction model by ordinary least squares as the residual sum of squares that is minimized in [2.1.8] is the same as that of [2.1.1], the resulting estimates are numerically identical.

Both the autoregressive distributed lag model in [2.1.1] and the error-correction model in [2.1.8] assume that the value of x_t can be treated as given, i.e., as being uncorrelated with the equation's error terms. This is to say that [2.1.1] is appropriately describing the expected value of y_t given its own history and conditional on current and lagged values of x_t. If x_t is simultaneously determined with y_t and $E(x_t\epsilon_t)\neq 0$, OLS in either [2.1.1] or [2.1.8] is inconsistent. The solution in this case is to consider a bivariate model for both y_t and x_t.

Special case of the model in [2.1.1] can be derived from alternative models that have some economic interpretation. For example, if y_t^* denotes the optimal or desired level of y_t and assume that

$$y_t^* = \beta_0 + \beta_1 x_t + \eta_t \qquad [2.1.9]$$

where η_t is an error term independent of $x_t, x_{t-1}, \cdots$. The actual value y_t differs from y_t^* because adjustment to its optimal level corresponding to x_t is not immediate. Suppose the adjustment is partial in the sense that

$$y_t - y_{t-1} = (1-\phi)(y_t^* - y_{t-1}) \qquad [2.1.10]$$

where $0<\phi<1$. Substituting [2.1.9] we obtain

$$\begin{aligned} y_t &= y_{t-1} + (1-\phi)\beta_0 + (1-\phi)\beta_1 x_t - (1-\phi)y_{t-1} + (1-\phi)\eta_t \\ &= \alpha + \phi y_{t-1} + \theta_0 x_t + \epsilon_t \end{aligned} \qquad [2.1.11]$$

where $\alpha=(1-\phi)\beta_0, \theta_0=(1-\phi)\beta_1$ and $\epsilon_t=(1-\phi)\eta_t$. This is a special case of [2.1.1] as it does not include x_{t-1}. The model given by [2.1.9] and [2.1.10] is referred to as a **partial adjustment model**.

The autoregressive distributed lag model in [2.1.1] can easily be generalized. Restricting attention to two variables only, we can write the general form as

$$\phi(L)y_t = \alpha + \theta(L)x_t + \epsilon_t \qquad [2.1.12]$$

where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$
$$\theta(L) = \theta_0 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_p L^q$$

are two lag polynomials. Note that the constant term in $\theta(L)$ is not restricted to be one. Assuming $\phi(L)$ is invertible, we can write

$$y_t = \phi^{-1}(1)\alpha + \phi^{-1}(L)\theta(L)x_t + \phi^{-1}(L)\epsilon_t \qquad [2.1.13]$$

The coefficients in the lag polynomial $\phi^{-1}(L)\theta(L)$ describe the dynamic effects of x_t on current and future values of y_t. The long-run effect of x_t is obtained as

$$\phi^{-1}(1)\theta(1) = \frac{\theta_0 + \theta_1 + \theta_2 + \cdots + \theta_p}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \qquad [2.1.14]$$

which generalizes the results stated above. Recall that invertibility of $\phi(L)$ requires that $1 + \phi_1 + \phi_2 + \cdots + \phi_p < 1$, which guarantees that the denominator in [2.1.14] is nonzero. A special case arise if $\phi(L) = 1$, so that the model in [2.1.13] does not contain any lags of y_t. This is referred to as a **distributed lag model**.

As long as it can be assumed that the error term ϵ_t is a white noise process, or more generally stationary and independent of $x_t, x_{t-1}, \cdots$ and $y_{t-1}, y_{t-2}, \cdots$, the distributed lag modes can be estimated consistently by ordinary least squares. Problems may arise, however, if along with y_t and x_t the implied ϵ_t is also non-stationary.

2.2 *Models with Nonstationary Variables*

2.2.1 *Spurious Regressions*

The assumption that y_t and x_t are stationary is crucial for the properties of standard estimation and testing procedures. To show consistency of OLS estimators, for example, we typically use the result that sample (co)variances converge to population (co)variances as the sample size becomes sufficiently large. Unfortunately, when the series are Nonstationary the (co)variances are ill-defined as the series does not fluctuate around a constant mean.

As an illustration, consider two variables, y_t and x_t, generated by two independent random walks,

$$y_t = y_{t-1} + \epsilon_{1t}, \qquad \epsilon_{1t} \sim IID\left(0, \sigma_1^2\right) \tag{2.2.1}$$

$$x_t = x_{t-1} + \epsilon_{2t}, \qquad \epsilon_{2t} \sim IID\left(0, \sigma_2^2\right) \tag{2.2.2}$$

Where ϵ_{1t} and ϵ_{2t} are mutually independent. There is nothing in this data generating mechanism that leads to a relationship between y_t and x_t. A researcher unfamiliar with these process may want to estimate a regression model explaining y_t by x_t and a constant,

$$y_t = \alpha + \beta x_t + \epsilon_t, \tag{2.2.3}$$

The results from this regression are likely to be characterized by a fairly high R^2 statistic, highly autocorrelated residuals and a significant value for β. This phenomenon is the well-known problem of nonsense or **spurious regressions**. Granger and Newbold argued that spurious regressions are characterized by a high R^2 and a low Durbin-Watson (DW) statistic, the usual *t*- and *F*-tests on the regression parameters may be very misleading. The reason being the distributions of the conventional test statistics are very different from those derived under the assumption of stationarity. Granger and Newbold stated that $R^2 > DW$ can used as the simplest rule of thumb to identify a spurious regression.

To illustrate the spurious regression result, two series of size 200 are generated according to [2.2.1] and [2.2.2] with normal error terms, starting with $y_0 = x_0 = 0$ and setting $\sigma_1^2 = \sigma_2^2 = 1$. The results of standard OLS regression of y_t on x_t and a constant are as presented in the following table.

Table 2.1 Spurious Regression: OLS involving two independent random Walks

Dependent Variable: y				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.980115	0.190269	-10.40695	0.0000
x	0.284043	0.027604	10.28995	0.0000

$S = 3.093597 \quad R^2 = 0.348434 \quad \overline{R}^2 = 0.345143 \quad F = 105.8831 \quad DW = 0.155138$

While the parameter estimators in this table would be completely different from one simulation to another, the *t*-ratios, R^2 and DW statistic show a very typical pattern: using the usual significance levels, both the constant term and x_t are highly significant, the R^2 of 35% seems reasonable though the DW statistic is very low. Estimation results like this should not be taken seriously. Because both y_t and x_t contain stochastic trend, the OLS estimator tends to find a significant correlation between the two series, even if they are unrelated. The problem is that ϵ_t is nonstationary.

2.2.2 *Cointegration*

An important exception to the results in the previous subsection arises when the two nonstationary series have the same stochastic trend in common. Consider two series, integrated of order one, y_t and x_t, and suppose that a linear relationship exists between them. If there exist some value of β such that $y_t - \beta x_t$ is $I(0)$, although y_t and x_t are both $I(1)$, in such a case it is said that y_t and x_t are **cointegrated**, and that they share a **common trend**. The relevant asymptotic theory is nonstandard, however, it can be shown that one can consistently estimate β from an OLS regression of y_t on x_t as in [2.2.3]. In this case, the OLS estimator $\hat{\beta}$ is said to be **super consistent** for β, because it converges to β at the rate faster than the conventional asymptotics. In the standard case, $\sqrt{T}\left(\hat{\beta}-\beta\right)$ is asymptotically normal and we say that $\hat{\beta}$ is $\sqrt{T}$ consistent for β. In the cointegration case, $\sqrt{T}\left(\hat{\beta}-\beta\right)$ is degenerate and the appropriate asymptotic distribution is that of $T\left(\hat{\beta}-\beta\right)$. Consequently, conventional inference procedures do not apply.

If y_t and x_t are both $I(1)$ and there exists a β such that $Z_t = y_t - \beta x_t$ is $I(0)$, y_t and x_t are cointegrated, with β being called the cointegrating parameter, or, more generally, $\begin{pmatrix}1 & -\beta\end{pmatrix}'$ being called the **cointegrating vector**. When this occurs, a special constraint operates on the long-run components of y_t and x_t. Since both y_t and x_t are $I(1)$, they will be dominated by 'long-wave' components, but Z_t, being $I(0)$, will not be: y_t and βx_t must therefore have long-run components that virtually cancel out to produce Z_t. This idea is related to the concept of a long-run equilibrium. Suppose that such an equilibrium is defined by the relationship

$$y_t = \alpha + \beta x_t \qquad [2.2.4]$$

Then $z_t = Z_t - \alpha$ is the 'equilibrium error', which measures the extent to which the value of y_t deviates from its 'equilibrium value' $\alpha + \beta x_t$. If z_t is $I(0)$, the equilibrium error is stationary and fluctuating around zero. Consequently, the system will, on average, be in equilibrium. However, if y_t and x_t are not cointegrated and, consequently z_t is $I(1)$, the equilibrium error wander widely and zero crossing would be very rare. Under this circumstance, it does not make sense to refer to $y_t = \alpha + \beta x_t$ as a long-run equilibrium. Consequently, the presence of a cointegrating vector can be interpreted as the presence of long-run equilibrium relationship.

From the discussion above, it is obvious that it is important to distinguish cases where there is a cointegrating relationship between y_t and x_t and spurious regression cases. Suppose we know that y_t and x_t are integrated of order one, and suppose we estimate the 'cointegrating regression'

$$y_t = \alpha + \beta x_t + \epsilon_t \qquad [2.2.5]$$

If y_t and x_t are cointegrated, the error term in [2.2.5] is $I(0)$. If not, ϵ_t will be $I(1)$. Hence one can test for the presence of a cointegrating relationship by testing for a unit root in the OLS residuals e_t form [2.2.5]. To do so, one can run the regression

$$\Delta e_t = \gamma_0 + \gamma_1 e_{t-1} + u_t \qquad [2.2.6]$$

and test whether $\gamma_1 = 0$ (a unit root). There is, however, an additional complication in testing for unit roots in OLS residuals rather than in observed time series. Because the OLS estimator chooses the residuals in the cointegrating regression [2.2.5] to have as small a sample variance as possible, even if the variables are not cointegrated, the OLS estimator will make the residuals look as stationary as possible. Thus, using standard DF or ADF critical values, we may reject the null hypothesis of nonstationary too often. As a result, the appropriate critical values are more negative than those for the standard Dickey-Fuller tests. For appropriate Critical values see Davidson, R. and Mackinnon, J.G., (1993), *Estimation and Inference in Econometrics*, Oxford University Press, Oxford. If e_t is not appropriately described by first order autoregressive process, one should add lagged values of Δe_t in [2.2.6], leading to the augmented Dickey-Fuller (ADF) test, with the same asymptotic critical values. This test can be extended to test for cointegration between three or more variables. If more than one x_t variable is included in the cointegrating regression, the critical values shift further to the left.

An alternative test for cointegration is based on the usual Durbin-Watson statistic from [2.2.5]. Note that the presence of a unit root in ϵ_t asymptotically corresponds to a zero value of the DW statistic. Under the null hypothesis of a unit root, the appropriate test is whether DW is significantly larger than zero. Unfortunately the critical values for this test, commonly referred to as the **cointegrating regression Durbin-Watson test** or **CRDW** test, depend on the process that generated the data. Nevertheless, the value of the DW statistic often suggests the presence or absence of a cointegrating relationship. Note that, when T goes to infinity, and y_t and x_t are not cointegrated, the DW statistic converges to zero in probability.

Although the existence of a long-run relationship between two variables is of interest, it may be even more relevant to analyze the short-run properties of the two series. This can be done using the result that the presence of a cointegrating relationship implies that there exists an error-correction model that describes the short-run dynamics consistently with the long-run relationship.

2.2.3 *Cointegration and Error-correction Mechanisms*

The Granger representation theorem states that, if a set of variables are cointegrated, then there exists a valid **error-correction representation** of the data. Thus, if y_t and x_t are both $I(1)$ and have a cointegrating vector $(1 \quad -\beta)'$, there exists an error-correction representation, with $Z_t = y_t - \beta x_t$, of the form

$$\phi(L)\Delta y_t = \delta + \theta(L)\Delta x_{t-1} - \gamma Z_{t-1} + \alpha(L)\epsilon_t \qquad [2.2.7]$$

where ϵ_t is white noise[1] and where $\phi(L)$, $\theta(L)$ and $\alpha(L)$ are polynomials in the lag operator L (with $\phi_0 \equiv 1$). Let us consider a special case of [2.2.7],

$$\Delta y_t = \delta + \theta_1 \Delta x_{t-1} - \gamma(y_{t-1} - \beta x_{t-1}) + \epsilon_t \qquad [2.2.8]$$

If y_t and x_t are both $I(1)$ but have a long-run relationship, there must be some force that pulls the equilibrium error back towards zero. The error-correction model does exactly this: it describes how y_t and x_t behave in the short-run consistent with the long-run cointegrating relationship. If the cointegrating parameter β is known, all terms in [2.2.8] are $I(0)$ and no inferential problems arise: we can estimate it by OLS in the usual way.

When $\Delta y_t = \Delta x_{t-1} = 0$, we obtain the 'no change' steady state equilibrium

$$y_{t-1} - \beta x_{t-1} = \frac{\delta}{\gamma} \qquad [2.2.9]$$

which corresponds to [2.2.4] if $\alpha = \delta/\gamma$. In this case the error-correction model can be written as

$$\Delta y_t = \theta_1 \Delta x_{t-1} - \gamma(y_{t-1} - \alpha - \beta x_{t-1}) + \epsilon_t \qquad [2.2.10]$$

If, however, the error-correction model [2.2.8] contains a constant that equals $\delta = \alpha\gamma + \lambda$, with $\lambda \neq 0$, this implies deterministic trends in both y_t and x_t and the long-run equilibrium corresponds to a steady state growth path with $\Delta y_t = \Delta x_{t-1} = \lambda/(1-\theta_1)$.

[1] The white noise term ϵ_t is assumed to be independent of both $y_{t-1}, y_{t-2}, \cdots$ and $x_{t-1}, x_{t-2}, \cdots$.

In some cases it makes sense to assume that the cointegrating vector is known a priori. If β is unknown, the cointegrating vector can be estimated (super) consistently from the cointegrating regression [2.2.5]. Consequently, with standard $\sqrt{T}$ asymptotics, one can ignore the fact that β is estimated and apply conventional theory to the estimation of parameters in [2.2.7]. Note that the precise lag structure in [2.2.7] is not specified by the theorem, so we need to do some specification search. Moreover, the theory is symmetric in its treatment of y_t and x_t, so that there should also exist an error-correction representation with Δx_t as the left-hand side variable. Because at least one of the variables has to adjust to deviations from the long-run equilibrium, at least one of the adjustment parameters γ in the two error-correction equations has to be nonzero. If x_t does not adjust to the equilibrium error (has zero adjustment parameter), it is weakly exogenous for β. This means that we can include Δx_t in the right-hand side of [2.2.8] without affecting the error-correction term $-\gamma(y_{t-1} - \beta x_{t-1})$. That is, we can condition upon x_t in the error-correction model for y_t.

The representation theorem also holds conversely, i.e., if y_t and x_t are both $I(1)$ and have an error-correction representation, then they are necessarily cointegrated. It is important to realize the concept of cointegration can be applied to (nonstationary) integrated time series only. If y_t and x_t are both $I(0)$, the generating process can always be written in an error-correction form.

2.3 *Vector Autoregressions (VARs)*

The autoregressive moving average models can readily be extended to the multivariate case, in which the stochastic process that generates the vector of time series variables is modelled. The most common approach is to consider a **vector autoregressive (*VAR*)** model. A VAR describes the dynamic evolution of a number of variables from their common history. A multivariate time series $\boldsymbol{y}_t$ is a vector process $m \times 1$. Let $\mathcal{F}_{t-1} = \{\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots\}$ be all lagged information at time *t*. The typical goal is to find the conditional expectation $\mathbb{E}(\boldsymbol{y}_t | \mathcal{F}_{t-1})$. Note that since $\boldsymbol{y}_t$ is a vector, this conditional expectation is also a vector. A *VAR* model specifies that the conditional mean is a function of only a finite number of lags:

$$\mathbb{E}(\boldsymbol{y}_t | \mathcal{F}_{t-1}) = \mathbb{E}(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}, \cdots, \boldsymbol{y}_{t-k}).$$

A linear *VAR* specifies that this conditional mean is linear in the arguments:

$$\mathbb{E}\left(\boldsymbol{y}_t \middle| \boldsymbol{y}_{t-1}, \cdots, \boldsymbol{y}_{t-k}\right) = \boldsymbol{a}_0 + \mathbf{A}_1 \boldsymbol{y}_{t-1} + \mathbf{A}_2 \boldsymbol{y}_{t-2} + \cdots + \mathbf{A}_k \boldsymbol{y}_{t-k} \qquad [2.3.1]$$

Observe that $\mathbf{a}_0$ is $m \times 1$, and each of $\mathbf{A}_1$ through $\mathbf{A}_k$ are $m \times m$ matrices.

Defining the $m \times 1$ regression error

$$\boldsymbol{e}_t = \boldsymbol{y}_t - \mathbb{E}\left(\boldsymbol{y}_t \middle| \mathcal{F}_{t-1}\right), \qquad [2.3.2]$$

we have the *VAR* model

$$\begin{aligned} &\boldsymbol{y}_t = \boldsymbol{a}_0 + \mathbf{A}_1 \boldsymbol{y}_{t-1} + \mathbf{A}_2 \boldsymbol{y}_{t-2} + \cdots + \mathbf{A}_k \boldsymbol{y}_{t-k} + \boldsymbol{e}_t \\ &\left(\boldsymbol{e}_t \middle| \mathcal{F}_{t-1}\right) = 0 \end{aligned} \qquad [2.3.3]$$

As in the univariate case, we can use the lag operator to define a matrix lag polynomial and write the VAR model in [2.3.3] as

$$\boldsymbol{A}(L)\boldsymbol{y}_t = \boldsymbol{a}_0 + \boldsymbol{e}_t, \qquad [2.3.4]$$

where $\mathbf{A}(L) = I_m - \mathbf{A}_1 L - \mathbf{A}_2 L^2 - \cdots - \mathbf{A}_k L^k$ and I_m is an m-dimensional identity matrix.

Extensions to vectorial ARMA (VARMA) model is obtained by premultiplying $\boldsymbol{e}_t$ with a (matrix) lag polynomial. The advantage of considering the components simultaneously include that the model may be more parsimonious and includes fewer lags, and that more accurate forecasting is possible, because the information set is extended also to include the history of the other variables. From a different perspective, Sims(1980) has advocated the use of VAR models instead of structural simultaneous models because the distinction between endogenous and exogenous variables does not have to be made a priori, and 'arbitrary' constraints to ensure identification are not required. Like a reduced form, a VAR is always identified.

The unconditional expected value of $\boldsymbol{y}_t$ can be determined if we impose stationarity and the elements of the vector error term are a white noise process as follows:

$$\mathbb{E}\left(\boldsymbol{y}_t\right) = \boldsymbol{a}_0 + \mathbf{A}_1 \mathbb{E}\left(\boldsymbol{y}_t\right) + \mathbf{A}_2 \mathbb{E}\left(\boldsymbol{y}_t\right) + \cdots + \mathbf{A}_k \mathbb{E}\left(\boldsymbol{y}_t\right)$$

or $\mathbb{E}\left(\boldsymbol{y}_t\right) = \left(I_m - \mathbf{A}_1 - \mathbf{A}_2 - \cdots - \mathbf{A}_k\right)^{-1} \boldsymbol{a}_0$.

This shows that stationarity will require that the $m \times m$ matrix $\boldsymbol{A}(1)$ is invertible.

Alternatively the VAR model in [2.3.3] can be written as, defining the $(mk+1)\times 1$ vector

$$\boldsymbol{x}_t = \begin{pmatrix} 1 \\ \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \vdots \\ \boldsymbol{y}_{t-k} \end{pmatrix}$$

and the $m \times (mk+1)$ matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_k \end{pmatrix},$$

then

$$\boldsymbol{y}_t = \mathbf{A}\boldsymbol{x}_t + \boldsymbol{e}_t$$

The *VAR* model is a system of *m* equations. One way to write this is to let a_j' be the j[th] row of **A.** Then the *VAR* system can be written as the equations

$$Y_{jt} = \boldsymbol{a}_j'\boldsymbol{x}_t + \boldsymbol{e}_{jt}.$$

Unrestricted VARs were introduced to econometrics by Sims (1980). If $\boldsymbol{A}(L)$ in [2.3.4] is invertible, it means that we can write the vector autoregressive model as a vector moving average (VMA) model by premultiplying by $\boldsymbol{A}^{-1}(L)$.

$$\boldsymbol{y}_t = \boldsymbol{A}^{-1}(1)\boldsymbol{a}_0 + \boldsymbol{A}^{-1}(L)\boldsymbol{e}_t = \boldsymbol{\mu} + \boldsymbol{A}^{-1}(L)\boldsymbol{e}_t$$

This describes each element in $\boldsymbol{y}_t$ as a weighted sum of current and past shocks in the system. Writing $\boldsymbol{A}^{-1}(L) = I_m + \boldsymbol{B}_1 L + \boldsymbol{B}_2 L^2 + \cdots$, we have

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{e}_t + \boldsymbol{B}_1\boldsymbol{e}_{t-1} + \boldsymbol{B}_2\boldsymbol{e}_{t-2} + \cdots$$

If the vector error term $\boldsymbol{e}_t$ increases by a vector $\boldsymbol{\delta}$, the effect upon $\boldsymbol{y}_{t+s}$ $(s>0)$ is given by $\boldsymbol{B}_s\boldsymbol{\delta}$. Thus the matrix

$$\boldsymbol{B}_s = \frac{\partial \boldsymbol{y}_{t+s}}{\partial \boldsymbol{e}_t'}$$

has the interpretation that its (i,j) element measures the effect of a one-unit increase in e_{jt} upon $y_{i,t+s}$. If only the first element e_{1t} of $\boldsymbol{e}_t$ changes, the effects are given by the first column of $\boldsymbol{B}_s$. The dynamic effects upon the j[th] variable of such a one-unit increase are given by the elements in the first column and j[th] row of $I_m, \boldsymbol{B}_1, \boldsymbol{B}_2, \cdots$. A plot of these elements as a function of *s* is called the **impulse-response function**. It measures the response of an increase in $y_{i,t+s}$ to an impulse in y_{1t}, keeping constant all other variables dated *t* and before.

2.3.1 Estimation

Consider the moment conditions

$$\mathbb{E}\left(\boldsymbol{x}_t \boldsymbol{e}_{jt}\right) = \mathbf{0},$$

$j = 1, \cdots, m$. These are implied by the VAR model, either as a regression, or, as a linear projection.

The GMM estimator corresponding to these moment conditions is equation by equation OLS estimator given by

$$\hat{\boldsymbol{a}}_j = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}_j$$

where

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_T' \end{pmatrix} \qquad (T \text{ x } (mk+1)) \text{ matrix}$$

and an alternative to compute this is as follows. Note that

$$\hat{\boldsymbol{a}}_j' = \boldsymbol{y}_j'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

and if we stack these to create the estimator $\widehat{\mathbf{A}}$, we find

$$\widehat{\mathbf{A}} = \begin{pmatrix} \boldsymbol{y}_1' \\ \boldsymbol{y}_2' \\ \vdots \\ \boldsymbol{y}_m' \end{pmatrix} \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \quad = \quad \boldsymbol{Y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1},$$

where $Y = (\boldsymbol{y}_1 \quad \boldsymbol{y}_2 \quad \cdots \quad \boldsymbol{y}_m)$ is the $T \times m$ matrix of the stacked $\boldsymbol{y}_t'$.

This (system) estimator is known as the SUR (Seemingly Unrelated Regressions) estimator, and was originally derived by Zellner (1962).

2.3.2 Restricted VARs

The unrestricted *VAR* is a system of *m* equations, each with the same set of regressors. A restricted *VAR* imposes restrictions on the system. For example, some regressors may be excluded from some the equations. Restrictions may be imposed on individual equations, or across equations. The GMM framework gives a convenient method to impose such restrictions on estimation.

2.3.3 *Single equation from a VAR*

Often, we are only interested in a single equation out of a VAR system. This takes the form

$$y_{jt} = \boldsymbol{a}_j'\boldsymbol{x}_t + e_{jt}$$

and $\boldsymbol{x}_t$ consists of lagged values of y_{jt} and the other $y_{lt}'s$. In this case, it is convenient to re-define the variables as follows. Let $y_t = y_{jt}$, and $\boldsymbol{z}_t$ be the other variables. Let $e_t = e_{jt}$, and $\boldsymbol{\beta} = \boldsymbol{a}_j$. Then the single equation takes the form

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t, \qquad [2.3.5]$$

and $\boldsymbol{x}_t = \left[\left(1 \quad y_{t-1} \quad \cdots \quad y_{t-k} \quad \boldsymbol{z}_{t-1}' \quad \cdots \quad \boldsymbol{z}_{t-k}'\right)'\right]$.

This is just a conventional regression with time series data.

2.3.4 *Testing for Omitted Serial Correlation*

Consider the problem of testing for omitted serial correlation in equation [2.3.5]. Suppose that e_t is an *AR(1)*. Then

$$\begin{aligned} y_t &= \boldsymbol{x}_t'\boldsymbol{\beta} + e_t \\ e_t &= \theta e_{t-1} + u_t \\ \mathbb{E}\left(u_t \middle| \mathcal{F}_{t-1}\right) &= 0 \end{aligned} \qquad [2.3.6]$$

Then the null and the alternative hypothesis are

$$\mathbb{H}_0 : \theta = 0 \qquad \text{and} \qquad \mathbb{H}_1 : \theta \neq 0$$

Take the equation $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t$, and subtract off the equation once lagged multiplied by θ, to get

$$\begin{aligned} y_t - \theta y_{t-1} &= \left(\boldsymbol{x}_t'\boldsymbol{\beta} + e_t\right) - \theta\left(\boldsymbol{x}_{t-1}'\boldsymbol{\beta} + e_{t-1}\right) \\ &= \boldsymbol{x}_t'\boldsymbol{\beta} - \theta\boldsymbol{x}_{t-1}'\boldsymbol{\beta} + e_t - \theta e_{t-1} \end{aligned}$$

or

$$y_t = \theta y_{t-1} + \boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{x}_{t-1}'\boldsymbol{\gamma} + u_t \qquad [2.3.7]$$

which is a valid regression model.

So testing $\mathbb{H}_0$ versus $\mathbb{H}_1$ is equivalent to testing for the significance of adding $\left(y_{t-1}, \boldsymbol{x}_{t-1}\right)$ to the regression. This can be done by a Wald test. We see that an appropriate, general, and simple way to test for omitted serial correlation is to test for the significance of extra lagged values of the dependent variable and regressors.

You may recall of the Durbin Watson's test for omitted serial correlation, which once was very popular, and still routinely reported by conventional regression packages. The DW test is appropriate only when regression $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t$ is not dynamic (has no lagged values on the RHS), and e_t is $i.i.d.N\left(0, \sigma^2\right)$. Otherwise it is invalid.

Another interesting fact is that [2.3.6] is a special case of [2.3.7], under the restriction $\boldsymbol{\gamma} = -\theta\boldsymbol{\beta}$. This restriction, which is called a common factor restriction, may be tested if desired. If valid, the model [2.3.6] may be estimated by iterated GLS. (A simple version of this estimator is called Cochrane-Orcutt.) Since the common factor restriction appears arbitrary, and is

typically rejected empirically, direct estimation of [2.3.6] is uncommon in recent applications.

2.3.5 *Selection of Lag Length in a VAR*

If you want a data-dependent rule to pick the lag length k in a *VAR*, you may either use a testing based approach (using, for example, the Wald statistic), or an information criterion approach. The formula for *AIC* and *BIC* are

$$AIC(k) = \log\det\left(\widehat{\boldsymbol{\Omega}}(k)\right) + 2\frac{p}{T}$$
$$BIC(k) = \log\det\left(\widehat{\boldsymbol{\Omega}}(k)\right) + 2\frac{p\log(T)}{T}$$
$$\widehat{\boldsymbol{\Omega}}(k) = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{e}}_t(k)\hat{\boldsymbol{e}}_t(k)'$$
$$p = m(mk+1)$$

where p is the number of parameters in the model, and $\hat{\boldsymbol{e}}_t(t)$ is the OLS residual vector from the model with k lags. The log determinant is the criterion from the multivariate normal likelihood.

2.3.6 *Granger Causality*

Partition the data vector into $(\boldsymbol{y}_t, \boldsymbol{z}_t)$. Define the two information sets

$$\mathcal{F}_{1t} = (\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \cdots)$$
$$\mathcal{F}_{2t} = (\boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{y}_{t-1}, \boldsymbol{z}_{t-1}, \boldsymbol{y}_{t-2}, \boldsymbol{z}_{t-2} \cdots)$$

The information set $\mathcal{F}_{1t}$ is generated only by the history of $\boldsymbol{y}_t$, and the information set $\mathcal{F}_{2t}$ is generated by both $\boldsymbol{y}_t$ and $\boldsymbol{z}_t$. The latter has more information.

We say that $\boldsymbol{z}_t$ does not Granger-cause $\boldsymbol{y}_t$, if

$$\mathbb{E}\left(\boldsymbol{y}_t \middle| \mathcal{F}_{1t-1}\right) = \mathbb{E}\left(\boldsymbol{y}_t \middle| \mathcal{F}_{2t-1}\right)$$

That is, conditional on information in lagged $\boldsymbol{y}_t$, lagged $\boldsymbol{z}_t$ does not help to forecast $\boldsymbol{y}_t$. If this condition does not hold, then we say that $\boldsymbol{z}_t$ Granger-cause $\boldsymbol{y}_t$.

The reason why we call this "Granger-Causality" rather than "Causality" is because this is not physical or structure definition of causality. If $\boldsymbol{z}_t$ is some sort of forecast of the future, such as future prices, then $\boldsymbol{z}_t$ may help to forecast $\boldsymbol{y}_t$ even though it does not "cause" $\boldsymbol{y}_t$. This definition of causality is developed by Granger (1969) and Sims (1972).

In a linear *VAR*, the equation for $\boldsymbol{y}_t$ is

$$\boldsymbol{y}_t = \alpha + \rho_1 \boldsymbol{y}_{t-1} + \cdots + \rho_k \boldsymbol{y}_{t-k} + \boldsymbol{z}_{t-1}' \boldsymbol{\gamma}_1 + \cdots + \boldsymbol{z}_{t-k}' \boldsymbol{\gamma}_\kappa + e_t$$

In this equation $\boldsymbol{z}_t$ does not Granger-cause $\boldsymbol{y}_t$ if and only if

$$\mathbb{H}_0 : \boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \cdots = \boldsymbol{\gamma}_k = \mathbf{0}.$$

This may be tested using an exclusion (Wald) test.

This idea can be applied to blocks of variables. That is, $\boldsymbol{y}_t$ and/or $\boldsymbol{z}_t$ can be vectors. The hypothesis can be tested using the appropriate multivariate Wald test.

If it is found that $\boldsymbol{z}_t$ does not Granger-cause $\boldsymbol{y}_t$, then we deduce that our time-series model of $\mathbb{E}\left(\boldsymbol{y}_t \middle| \mathcal{F}_{1t-1}\right)$ does not require the use of $\boldsymbol{z}_t$. Note, however, that $\boldsymbol{z}_t$ may still be useful to explain other feature of $\boldsymbol{y}_t$, such as the conditional variance.

2.3.7 *Cointegration: the Multivariate case*

The idea of cointegration is due to Granger (1981), and was articulated in detail by Engle and Granger (1987).

Definition 2.3.1 The $m \times 1$ series $\boldsymbol{y}_t$ is **cointegrated** if $\boldsymbol{y}_t$ is $I(1)$ yet there exists $\boldsymbol{\beta}$, $m \times \mathrm{r}$, of rank *r*, such that $\boldsymbol{z}_t = \boldsymbol{\beta}'\boldsymbol{y}_t$ is $I(0)$. The *r* vectors in $\boldsymbol{\beta}$ are called the **cointegrating vectors**.

If the series $\boldsymbol{y}_t$ is not cointegrated, then $r = 0$. If $r = m$, then $\boldsymbol{y}_t$ is $I(0)$. For $0 < r < m$, $\boldsymbol{y}_t$ is $I(1)$ and cointegrated.

In some cases, it may be believed that $\boldsymbol{\beta}$ is known a priori. Often $\boldsymbol{\beta} = \begin{pmatrix} 1 & -1 \end{pmatrix}'$. For example, if $\boldsymbol{y}_t$ is a pair of interest rates, then $\boldsymbol{\beta} = \begin{pmatrix} 1 & -1 \end{pmatrix}'$ specifies that the spread (the difference in returns) is stationary. If $\boldsymbol{y} = \begin{pmatrix} \log(consumption) & \log(income) \end{pmatrix}'$, then $\boldsymbol{\beta} = \begin{pmatrix} 1 & -1 \end{pmatrix}'$ specifies that $\log(consumption/income)$ is stationary. In other cases, $\boldsymbol{\beta}$ may not be known.

If $\boldsymbol{y}_t$ is cointegrated with a single cointegrating vector $(r = 1)$, then it turns out that $\boldsymbol{\beta}$ can be consistently estimated by an OLS regression of one component of $\boldsymbol{y}_t$ on the others. Thus $\boldsymbol{y}_t = \begin{pmatrix} y_{1t} & y_{2t} \end{pmatrix}'$ and $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix}'$ and normalize $\beta_1 = 1$. Then $\hat{\beta}_2 = \left(\boldsymbol{y}_2'\boldsymbol{y}_2\right)^{-1}\boldsymbol{y}_2'\boldsymbol{y}_1 \xrightarrow{p} \beta_2$. Furthermore, this estimation is super-consistent: $T\left(\hat{\beta}_2 - \beta_2\right) \xrightarrow{d} Limit$, as first shown by

Stock(1987). This is not, in general, a good method to estimate $\boldsymbol{\beta}$, but it is useful in the construction of alternative estimators and tests.

We are often interested in testing the null hypothesis of no cointegration:

$$\mathbb{H}_0 : r = 0$$
$$\mathbb{H}_1 : r > 0$$

Suppose that $\boldsymbol{\beta}$ is known, so that $z_t = \boldsymbol{\beta}' \boldsymbol{y}_t$ is known. Then under $\mathbb{H}_0$ z_t is $I(1)$, yet under $\mathbb{H}_1$ z_t is $I(0)$. Thus $\mathbb{H}_0$ can be tested using a univariate ADF test on z_t.

When $\boldsymbol{\beta}$ is unknown, Engle and Granger (1987) suggested using ADF test on the estimated residuals $\hat{z}_t = \hat{\boldsymbol{\beta}}' \boldsymbol{y}_t$, from OLS of y_{1t} on y_{2t}. Their justification was Stock's result that $\hat{\boldsymbol{\beta}}$ is super-consistent under $\mathbb{H}_1$. Under $\mathbb{H}_0$, however, $\hat{\boldsymbol{\beta}}$ is not consistent, so the ADF critical values are not appropriate. The asymptotic distribution was worked out by Phillips and Ouliaries (1990).

When the data have time trends, it may be necessary to include a time trend in the estimated cointegrating regression. Whether or not the time trend is included, the asymptotic distribution of the test is affected by the presence of the time trend. The asymptotic distribution was worked out in B. Hansen (1992).

2.3.8 *Cointegrated VARs*

We can write a *VAR* as

$$\mathbf{A}(\mathrm{L})\boldsymbol{y}_t = \boldsymbol{e}_t$$
$$\mathbf{A}(\mathrm{L}) = \boldsymbol{I} - \mathbf{A}_1 \mathrm{L} - \mathbf{A}_2 \mathrm{L}^2 - \cdots - \mathbf{A}_k \mathrm{L}^k$$

Or alternatively as

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \mathbf{D}(\mathrm{L}) \Delta \boldsymbol{y}_{t-1} + \boldsymbol{e}_t$$

where

$$\boldsymbol{\Pi} = -\mathbf{A}(1)$$
$$= -\mathbf{I} + \mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_k$$

Theorem 2.3.1 Granger Representation Theorem $\boldsymbol{y}_t$ is cointegrated with $m \times r$ $\boldsymbol{\beta}$ if and only if $\operatorname{rank}(\boldsymbol{\Pi}) = r$ and $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ where α is $m \times r$, $\operatorname{rank}(\alpha) = r$.

Thus cointegration imposes a restriction upon the parameters of a VAR. The restricted model can be written as

$$\begin{aligned}\Delta \boldsymbol{y}_t &= \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \mathbf{D}(\mathrm{L})\Delta \boldsymbol{y}_{t-1} + \boldsymbol{e}_t \\ \Delta \boldsymbol{y}_t &= \boldsymbol{\alpha}\boldsymbol{z}_{t-1} + \mathbf{D}(\mathrm{L})\Delta \boldsymbol{y}_{t-1} + \boldsymbol{e}_t\end{aligned} \qquad [2.3.8]$$

If $\boldsymbol{\beta}$ is known, this can be estimated by OLS of $\Delta \boldsymbol{y}_t$ on $\boldsymbol{z}_{t-1}$ and the lags of $\Delta \boldsymbol{y}_t$. If $\boldsymbol{\beta}$ is unknown, then estimation is done by "reduced rank regression", which is least-squares subject to the stated restriction. Equivalently, this is the MLE of the restricted parameters under the assumption that $\boldsymbol{e}_t$ is iid $N(\mathbf{0},\boldsymbol{\Omega})$. The linear combinations $\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$ represent the *r* cointegrating relationships. The coefficients in $\boldsymbol{\alpha}$ measure how the elements in $\Delta \boldsymbol{y}_t$ are adjusted to the *r* 'equilibrium errors' $\boldsymbol{z}_{t-1} = \boldsymbol{\beta}'\boldsymbol{y}_{t-1}$. Thus, [2.3.8] is a generalization of [2.2.8] and is referred to as a **vector error-correction model (*VECM*)**.

One difficulty is that $\boldsymbol{\beta}$ is not identified without normalization. When $r = 1$, we typically just normalize one element to equal to unity. When $r > 1$, this does not work, and different authors have adopted different identification schemes.

In the context of a cointegrated VAR estimated by reduced rank regression, it is simple to test for cointegration by testing the rank of $\boldsymbol{\Pi}$. These tests are constructed as likelihood ratio (LR) tests. As they were discovered by Johansen (1988, 1991, 1995), they are typically called the "Johansen Max and Trace" tests. Their asymptotic distributions are non-standard, are similar to the Dickey-Fuller distributions.

2.3.9 *Cointegration in a Bivariate VAR*

Consider the case where $m = 2$. In this case the number of cointegrating vectors may be zero or one $(r = 0,1)$. Let us consider a first-order (nonstationary) VAR for $\boldsymbol{y}_t = (y_t \quad x_t)'$. That is,

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix},$$

where, for simplicity, we dropped the intercept terms. The matrix $\boldsymbol{\Pi}$ is given by

$$\boldsymbol{\Pi} = -\mathbf{A}(1) = \begin{pmatrix} a_{11}-1 & a_{12} \\ a_{21} & a_{22}-1 \end{pmatrix},$$

This matrix is a zero matrix if $a_{11} = a_{22} = 1$ and $a_{12} = a_{21} = 0$. This corresponds to the case where y_t and x_t are two random walks. The matrix $\boldsymbol{\Pi}$ has reduced rank if

$$(a_{11}-1)(a_{22}-1) - a_{12}a_{21} = 0 \qquad [2.3.9]$$

If the is the case,

$$\boldsymbol{\beta}' = (a_{11}-1 \quad a_{12})$$

is a cointegrating vector (where we have chosen an arbitrary normalization) and we write

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' = \begin{pmatrix} 1 \\ a_{21}/(a_{11}-1) \end{pmatrix} \begin{pmatrix} a_{11}-1 & a_{12} \end{pmatrix}.$$

Using this, we can write the model in an error-correction form. First, write

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} a_{11}-1 & a_{12} \\ a_{21} & a_{22}-1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

Next, we write this as

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} 1 \\ a_{21}/(a_{11}-1) \end{pmatrix} \left[(a_{11}-1) y_{t-1} + a_{12} x_{t-1} \right] + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \qquad [2.3.10]$$

The error-correction form is thus quite simple, as it excludes any dynamics. Note that both y_t and x_t adjust to the equilibrium error, because $a_{21} = 0$ is excluded. Also note that $a_{21} = 0$ would imply $a_{11} = a_{22} = 1$ and no cointegration.

The fact that the linear combination $z_t = (a_{11}-1) y_t + a_{12} x_t$ is $I(0)$ also follows from this result. Note that we can write

$$\Delta z_t = \begin{pmatrix} a_{11}-1 & a_{12} \end{pmatrix} \begin{pmatrix} 1 \\ a_{21}/(a_{11}-1) \end{pmatrix} z_{t-1} + \begin{pmatrix} a_{11}-1 & a_{12} \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

or using [2.3.9]

$$z_t = z_{t-1} + (a_{11}-1+a_{22}-1) z_{t-1} + v_t = (a_{11}+a_{22}-1) z_{t-1} + v_t,$$

where $v_t = (a_{11}-1)\epsilon_{1t} + a_{12}\epsilon_{2t}$ is a white noise error term. Consequently, z_t is described by a stationary *AR(1)* process unless $a_{11} = 1$ and $a_{22} = 1$, which is excluded.

2.3.10 *Testing for Cointegration*

If it is known that there exists at most one cointegrating vector, a simple approach to testing for the existence of cointegration is the Engle-Granger two-step approach described in section 2.2.2. This requires running a regression of y_{1t} (being the first element of $\boldsymbol{y}_t$) on the other m-1 variables $y_{2t}, y_{3t}, \cdots, y_{mt}$ and testing for a unit root in the residuals. This can be done using the ADF tests on the OLS residuals, applying appropriate critical values. If the unit root hypothesis is rejected, the hypothesis of no cointegration is also rejected. In this case the static regression gives consistent estimates of the cointegrating vector, while in a second stage the error-correction model can be estimated using the estimated cointegrating vector from the first stage.

There are some problems with this Engle-Granger approach. First, the results of the tests are

sensitive to the left-hand side variable of the regression, i.e., to the normalization applied to the cointegrating vector. Second, if the cointegrating vector happens not to involve y_{1t}, but only $y_{2t}, y_{3t}, \cdots, y_{mt}$, the test is not appropriate and the cointegrating vector will not be consistently estimated by the regression of y_{1t} on $y_{2t}, y_{3t}, \cdots, y_{mt}$. Third, the residual based test tends to lack power because it does not exploit all the available information about the dynamic interactions of the variables. Fourth, it is possible that more than one cointegrating relationship exists between the variables $y_{1t}, y_{2t}, \cdots, y_{mt}$. If, for example, two distinct cointegrating relationships exist, OLS typically estimates a linear combination of them. Fortunately, as the null hypothesis for the cointegration tests is that there is no cointegration, the tests are still appropriate for this purpose.

An alternative approach that does not suffer from these drawbacks was proposed by Johansen (1988), who developed the maximum likelihood estimation procedure that also allows one to test for the number of cointegrating relationships.

Johansen's Maximum Likelihood Approach: (Journal of Economic Dynamics and Control, 1988, Econometrica, 1991)

Based upon VARS: k^{th} order VAR in a vector $(m \times 1)$ $\boldsymbol{y}_t$

$$\boldsymbol{y}_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{bmatrix} \quad ; \quad t = 1, 2, \cdots, T.$$

$$\boldsymbol{y}_t = \boldsymbol{a}_0 + \mathbf{A}_1 \boldsymbol{y}_{t-1} + \mathbf{A}_2 \boldsymbol{y}_{t-2} + \cdots + \mathbf{A}_k \boldsymbol{y}_{t-k} + \boldsymbol{\epsilon}_t \qquad [2.3.11]$$

where $\boldsymbol{\epsilon}_t \sim IN(\mathbf{0}, \boldsymbol{\Sigma})$.

Normality assumption is critical for Johansen's approach and let each $y_{it} \sim I(1)$. $\left[\boldsymbol{y}_t \sim I(1)\right]$. If we formulate equation [2.3.11] as a VAR in first differences $\Delta \boldsymbol{y}_t \left\{\Delta \boldsymbol{y}_t \sim I(0) \text{ if } \left[\boldsymbol{y}_t \sim I(1)\right]\right\}$

$$\Delta \boldsymbol{y}_t = \boldsymbol{a}_0 + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_2 \Delta \boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\Gamma}_{k-1} \Delta \boldsymbol{y}_{t-k+1} + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t \qquad [2.3.12]$$

where $\boldsymbol{\Gamma}_j = -\left[\mathbf{A}_{j+1} + \cdots + \mathbf{A}_k\right] \quad ; \quad j = 1, \cdots, k-1$ and $\boldsymbol{\Pi} = -\left[\mathbf{I} - \mathbf{A}_1 - \mathbf{A}_2 - \cdots - \mathbf{A}_\kappa\right]$.

We need $\boldsymbol{\Pi} \boldsymbol{y}_{t-1}$ to be $I(0)$ for this formulation to be sensible, i.e., defines linear combination of $I(1)$ process to give $I(0)$ process.

Error-Correction representation suggests

$$\Delta \boldsymbol{y}_t = \boldsymbol{a}_0 + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_2 \Delta \boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\Gamma}_{k-1} \Delta \boldsymbol{y}_{t-k+1} + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t \qquad [2.3.13]$$

where $\boldsymbol{y}_t \sim I(1)$ and $\boldsymbol{\epsilon}_t \sim IN(\mathbf{0},\boldsymbol{\Sigma})$.

For this to be sensible $\boldsymbol{\Pi} = -\left[\mathbf{I} - \mathbf{A}_1 - \mathbf{A}_2 - \cdots - \mathbf{A}_\kappa\right]$ is of rank say $r < m$ as $\boldsymbol{\Pi}\boldsymbol{y}_{t-1}$ is taking linear combinations of the vector $\boldsymbol{y}_{t-1}$ to make $\boldsymbol{\Pi}\boldsymbol{y}_{t-1} \sim I(0)$. These linear combinations of $I(1)$ variables are identified with cointegrating relationships among the $I(1)$ variables.

Johansen represents the hypothesis that $rank(\boldsymbol{\Pi}) = r < m$ as

$$\mathcal{H}_0 : \boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

where both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(m \times r)$ matrices $(r < m)$, then $\boldsymbol{\Pi}\boldsymbol{y}_{t-1} = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$.

Interpretations of α and β: *r* rows of $\boldsymbol{\beta}'$ are the *r* cointegrating vectors such that they define *r* linear combinations of the *m* components of the vector $\boldsymbol{y}_{t-1}$ that give $I(0)$. The elements of the $(m \times r)$ matrix $\boldsymbol{\alpha}$ represent the weights (loading) of the *r* cointegrating vectors in each of the *m* equations.

Example: Consider the case where $m = 3$ and $r = 2$.

$$\boldsymbol{\Pi}\boldsymbol{y}_{t-1} = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta 23 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{3t-1} \\ y_{3t-1} \end{bmatrix}$$

The 1st row of $\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$ is $\underbrace{\left(\beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + \beta_{13}y_{3t-1}\right)}_{ECM_1} \sim I(0)$ is a cointegrating vector.

The 2st row of $\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$ is $\underbrace{\left(\beta_{21}y_{1t-1} + \beta_{22}y_{2t-1} + \beta_{23}y_{3t-1}\right)}_{ECM_2} \sim I(0)$ is also a cointegrating vector.

$$\boldsymbol{\Pi}\boldsymbol{y}_{t-1} = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \underbrace{\begin{bmatrix} ECM_1 \\ ECM_2 \end{bmatrix}}_{I(0)} \qquad [2.3.14]$$

Thus, using [2.3.14] the 1st equation of error correction model can now be written as, $+\cdots+\alpha_{11}ecm_{1t-1} + \alpha_{12}ecm_{2t-1} + \epsilon_{1t}$. The Johansen's approach enables to

- determine the number of cointegrating vectors, *r*.
- estimate the cointegrating vectors.

Estimation of Π proceeds by ML

Suppose we have the k[th] order VAR represented as in an error correction of the form

$$\Delta\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_2\Delta\boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\Gamma}_{k-1}\Delta\boldsymbol{y}_{t-k+1} + \boldsymbol{\Pi}\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t \qquad [2.3.15]$$

Given $\boldsymbol{\epsilon}_t \sim IN(\mathbf{0},\boldsymbol{\Sigma})$, we can write the likelihood function corresponding to equation [2.3.15] by concentrating with respect to $\boldsymbol{\Gamma}_1,\cdots,\boldsymbol{\Gamma}_{k-1}$. This can be done by regressions of $\Delta\boldsymbol{y}_t$ and $\boldsymbol{y}_{t-1}$

each in turn on $\Delta \boldsymbol{y}_{t-1},\cdots,\Delta \boldsymbol{y}_{t-k+1}$ gives $(m\times 1)$ vector of residuals labelled by $\mathbf{r}_{0t}$, $\mathbf{r}_{1t}$, respectively. Define residual product moment matrices as

$$\boldsymbol{S}_{ij}=\frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_{it}\mathbf{r}_{jt}' \qquad ; \qquad i,j=0,1.$$

This gives $\boldsymbol{S}_{00}$, $\boldsymbol{S}_{11}$, $\boldsymbol{S}_{01}$, $\mathbf{S}_{10}$. The concentrated likelihood function corresponds to the regression

$$\mathbf{r}_{0t}=-\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{r}_{1t}+\boldsymbol{\epsilon}_t$$

For a fixed matrix $\boldsymbol{\beta}$, we can solve for $\boldsymbol{\alpha}$ by regressing $\mathbf{r}_{0t}$ on $-\boldsymbol{\beta}'\mathbf{r}_{1t}$ and obtain $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})=-\boldsymbol{S}_{01}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}$ and $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta})=\boldsymbol{S}_{00}-\boldsymbol{S}_{01}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{S}_{10}$ and Johansen demonstrated that maximizing the likelihood function is obtained by minimizing $\left|\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta})\right|$ with respect to β, this in turn apparently related to solving the eigen values, i.e., solve

$$\left|\lambda\boldsymbol{S}_{11}-\boldsymbol{S}_{01}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{10}\right|=0$$

That is, find λ (*m* eigenvalues) or (Characteristic roots), ordered so that $\hat{\lambda}_1>\hat{\lambda}_2>\cdots>\hat{\lambda}_m>0$. With these eigenvalues, there are *m* corresponding eigenvectors $\hat{\mathbf{V}}=\left[\hat{\mathbf{v}}_1 \quad \hat{\mathbf{v}}_2 \quad \cdots \quad \hat{\mathbf{v}}_m\right]$ normalized so that $\hat{\mathbf{V}}'\mathbf{S}_{11}\hat{\mathbf{V}}=\mathbf{I}$.

Johansen's estimation presumes knowledge of *r* (the number of cointegrating relations) but has **two advantages relative to Engle-Granger**:

- allows $r\geq 1$: non-uniqueness of the cointegrating vectors
- provides a framework (*ML-based*) for testing the number *r* of cointegrating vectors.

Recall that ML estimation involved getting *m* eigenvalues $\hat{\lambda}_1>\hat{\lambda}_2>\cdots>\hat{\lambda}_m>0$. These are used directly in testing the dimension *r*. Johansen suggested two tests:

i). based on testing the $(m-r)$ smallest eigenvalues are jointly zero. This test statistic often called the trace statistic is given by

$$\lambda_{trace}(r_0)=-T\sum_{i=r_0+1}^{m}\ln(1-\hat{\lambda}_i)$$

This statistic can be used to test:

$\mathcal{H}_0: r\leq r_0$ (at most r_0 cointegrating vectors) versus the alternative hypothesis

$\mathcal{H}_a: r_0<r\leq m$ (more than r_0 cointegrating vectors).

ii). The second is the "maximal eigenvalue" statistic based on the estimated $(r_0+1)^{th}$ largest eigenvalue given by

$$\lambda_{\max}(r_0) = -T\ln\left(1-\hat{\lambda}_{r_0+1}\right)$$

This statistic can be used to test:

$\mathcal{H}_0 : r \le r_0$ (at most r_0 cointegrating vectors) versus the alternative

$\mathcal{H}_a : r = r_0 + 1$ (exactly $r_0 + 1$ cointegrating vectors).

For both test statistics, asymptotic critical values have been tabulated for various different cases (*i.e.*, constant/trend). All these critical values are obtained by simulation: e.g. in Johansen (Journal of Economic Dynamics and Control, 1988). More comprehensive set are provided by Johansen and Juselius (Oxford Bulletin of Economics and Statistics, 1990). The two tests described here are actually likelihood ratio tests, but do not have the usual Chi-squared distributions. Instead, the appropriate distributions are multivariate extensions of the Dickey-Fuller distributions. As with the unit root tests, the percentiles of the distributions of these test statistics depend on whether a constant and a time trend are included.

Illustration using Long-run Purchasing Power Parity

Consider the existence of one or more cointegrating relationships between three variables S_t, P_t and P_t^* where S_t is the spot exchange rate (home currency price of a unit of foreign exchange), P_t is the aggregate price in domestic currency and P_t^* is the price of the foreign country) using the Johansen's technique discussed above. The first step in this procedure is the determination of *k*, the maximum order of the lags in the autoregressive representation given in [2.3.11]. It appears that, in general too few lags in the model leads to rejection of the null hypothesis too easily, while too many lags in the model decrease the power of the tests. This indicates that there is some optimal lag length. In addition to *k*, we have to decide upon whether to include a time trend in [2.3.11] or not. In the absence of a time trend, an intercept is automatically included in the cointegrating relationship(s). The optimal lag length in the autoregressive representation given in [2.3.11] for the dataset used in Verbeek is $k = 2$, which can be determined using the BIC. Let us consider the case with $k = 2$ excluding a time trend. The first step in Johansen's procedure yields the trace statistics in table 2.2 and maximum eigenvalue statistics in table 2.3 below. These results present the estimated eigenvalues $\hat{\lambda}_1, \hat{\lambda}_2$ and $\hat{\lambda}_3$ in decreasing order. Each nonzero eigenvalue corresponds to a cointegrating vector. A range of test statistics base on these estimated eigenvalues is given as well.

Table 2.2 Unrestricted Cointegration Rank Test (Trace)

Null Hypothesis	Alternative	Eigenvalue	λ_{trace} − Statistic	5% Critical Value	Prob.**
$H_0: r = 0^*$	$H_a: 0 < r \leq 3$	0.385964	115.6063	35.19275	0.0000
$H_0: r \leq 1^*$	$H_a: 1 < r \leq 3$	0.102029	25.86934	20.26184	0.0076
$H_0: r \leq 2$	$H_a: 2 < r \leq 3$	0.032439	6.067643	9.164546	0.1856

Trace test indicates 2 cointegrating equations at 5% level of significance.
*Denotes rejection of the null hypothesis at 5% level of significance.
**MacKinnon-Haug-Michelis (1999) p-values, lag length *k=2*, *T=184*, intercept included.

Table 2.2 above reports the results of the trace test statistic and as can be seen from the table:

1). The null hypothesis of no cointegration $(r = 0)$ has to be rejected at 5% level when tested against the alternative hypothesis of more than zero cointegrating vectors $(0 < r \leq 3)$, because 115.61 exceeds the critical value of 35.19.

2). The null hypothesis of zero or one cointegrating vector $(r \leq 1)$ also has to be rejected in favour of the alternative hypothesis of more than one cointegrating vectors $(r = 2)$.

3). The null hypothesis of two or fewer cointegrating vectors cannot be rejected against the alternative of hypothesis of more than two cointegrating vectors $(2 < r \leq 3)$.

Table 2.3 Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Null Hypothesis	Alternative	Eigenvalue	$\lambda_{\max}$ − Statistic	5% Critical Value	Prob.**
$H_0: r = 0*$	$H_a: r = 1$	0.385964	89.73699	22.29962	0.0000
$H_0: r \leq 1*$	$H_a: r = 2$	0.102029	19.80170	15.89210	0.0115
$H_0: r \leq 2$	$H_a: r = 3$	0.032439	6.067643	9.164546	0.1856

Maximum-eigenvalue test indicates 2 cointegrating equations at 5% level of significance.
*Denotes rejection of the null hypothesis at 5% level of significance.
**MacKinnon-Haug-Michelis (1999) p-values, lag length *k=2*, *T=184*, intercept included.

Table 2.3 above reports the results of the maximum eigenvalue test statistic and as can be seen from this table:

1). The null hypothesis of no cointegration $(r = 0)$ has to be rejected at 5% level when tested against the alternative hypothesis of one cointegrating vectors $(r = 1)$, because 89.74 exceeds the critical value of 22.30.

2). The null hypothesis of zero or one cointegrating vector $(r \leq 1)$ also has to be rejected in favour of the alternative hypothesis of two cointegrating vectors $(r = 2)$.

3). The null hypothesis of two or fewer cointegrating vectors cannot be rejected against the alternative of hypothesis of three cointegrating vectors $(r = 3)$. Recall that $r = 3$ corresponds to stationarity of each of the three series.

The following table reports cointegrating and the adjustment coefficients.

Table 2.4 Unrestricted Cointegrating Coefficients (normalized by $\hat{\mathbf{V}}'\mathbf{S}_{11}\hat{\mathbf{V}} = \mathbf{I}$):

s_t	p_t	p_t^*	C
-3.155508	28.48416	-55.76099	144.4973
9.448938	-46.05347	73.97055	-178.8928
-13.30384	35.41067	-48.51006	133.4533

where $s_t = \ln(S_t)$, $p_t = \ln(P_t)$ and $p_t^* = \ln(P_t^*)$

Unrestricted Adjustment Coefficients (α):

Δs_t	-0.000107	-0.000741	0.003567
Δp_t	0.001182	0.000710	8.44E-05
Δp_t^*	0.001309	-0.000292	2.21E-05

1 Cointegrating Equation: Log likelihood =2180.411

Normalized cointegrating coefficients (standard error in parentheses)

s_t	p_t	p_t^*	C
1.000000	-9.026808	17.67100	-45.79210
	(1.18380)	(2.12351)	(4.46623)

Adjustment coefficients (standard error in parentheses)

Δs_t	0.000336
	(0.00469)
Δp_t	-0.003728
	(0.00064)
Δp_t^*	-0.004132
	(0.00045)

2 Cointegrating Equations: Log likelihood = 2190.312

Normalized cointegrating coefficients (standard error in parentheses)

s_t	p_t	p_t^*	C
1.000000	0.000000	-3.723041	12.59047
		(0.62307)	(2.93895)
0.000000	1.000000	-2.370057	6.467687
		(0.09208)	(0.43436)

Adjustment coefficients (standard error in parentheses)		
Δs_t	-0.006664	0.031082
	(0.01479)	(0.08042)
Δp_t	0.002977	0.000972
	(0.00195)	(0.01058)
Δp_t^*	-0.006893	0.050756
	(0.00139)	(0.00754)

The Johansen's tests indicate the presence of two cointegrating vectors. Tables 2.5 and 2.6 below show what happens if we repeat the above procedure with a lag length of $k = 12$, motivated by the fact that we have monthly data.

Table 2.5 Unrestricted Cointegration Rank Test (Trace)

Null Hypothesis	Alternative	Eigenvalue	λ_{trace} − Statistic	5% Critical Value	Prob.**
$H_0 : r = 0*$	$H_a : 0 < r \leq 3$	0.106125	42.13841	35.19275	0.0076
$H_0 : r \leq 1*$	$H_a : 1 < r \leq 3$	0.090143	22.61742	20.26184	0.0233
$H_0 : r \leq 2$	$H_a : 2 < r \leq 3$	0.034894	6.179980	9.164546	0.1773

Trace test indicates 2 cointegrating equations at 5% level of significance.
*Denotes rejection of the null hypothesis at 5% level of significance.
**MacKinnon-Haug-Michelis (1999) p-values, lag length *k=12*, *T=174*, intercept included.

The first test that considers the null hypothesis of no cointegration $(r = 0)$ versus the alternative hypothesis of more than zero cointegrating relationship(s) $(0 < r \leq 3)$ leads to the rejection of the null using the trace-statistic. The second test that considers the null hypothesis of zero or one cointegrating relationship $(r \leq 1)$ versus the alternative hypothesis of more than one cointegrating relationship(s) $(1 < r \leq 3)$ also leads to the rejection of the null. Thus, the trace test statistic indicates the presence of two cointegrating relationships at 5% level of significance.

Table 2.6 Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Null Hypothesis	Alternative	Eigenvalue	$\lambda_{\max}$ − Statistic	5% Critical Value	Prob.**
$H_0 : r = 0$	$H_a : r = 1$	0.106125	19.52099	22.29962	0.1168
$H_0 : r \leq 1*$	$H_a : r = 2$	0.090143	16.43744	15.89210	0.0411
$H_0 : r \leq 2$	$H_a : r = 3$	0.034894	6.179980	9.164546	0.1773

Maximum-eigenvalue test indicates no cointegration at 5% level of significance.
*Denotes rejection of the null hypothesis at 5% level of significance.
**MacKinnon-Haug-Michelis (1999) p-values, lag length *k=12*, *T=174*, intercept included.

The first test that considers the null hypothesis of no cointegration $(r = 0)$ versus the alternative hypothesis of one cointegrating relationship $(r = 1)$ does not lead to the rejection of

the null using the maximum eigenvalue test statistic. The second test implies a marginal rejection of the hypothesis of the existence of zero or one cointegrating vector. Suppose we decide that the number of cointegrating vector is equal to one$(r=1)$, the estimated cointegrating vector $\boldsymbol{\beta}$, in this case, is given in table 2.7 below.

Table 2.7 Unrestricted and Normalized Cointegrating Vectors

	s_t	p_t	p_t^*	C
Unrestricted CV	1.210735	-7.684815	17.86432	-51.26381
Normalized CV	1.000000	-6.347233	14.75495	-42.34108

The normalized cointegrating vector corresponds to the cointegrating equation given by

$$s_t = 6.437\, p_t - 14.755\, p_t^*$$

which does not seem to correspond to an economically interpretable long-run relationship. The conclusion that there exists one cointegrating relationship between the three variables is most probably incorrect. To test for long-run purchasing power parity via Johansen's procedure, we will probably need longer time series. Alternatively, we may use several set of countries and apply panel data cointegration techniques. Another problem may lie in measuring the two price indices in an accurate way, comparable across the two countries.