Chapter 3

Common features

When doing an experiment in which several variables are observed, it is often not just a coincidence that the features like trend, seasonal component and the others (described in previous chapters) contaminating the individual time series are in fact a consequence of the same underlying processes. Then the time series are said to have these features in common and it is natural to build a multivariate model to utilize these connections for reducing the number of parameters and improving the forecast performance. Moreover, investigating the presence of the common features allows us to solve the problem of spurious regressions, as already noted in section 1.6.

The matter of nonsense regression can be easily explained on the example, where two time series variables are independently generated as random walks, i.e. $y_{1,t} = y_{1,t-1} + \varepsilon_{1,t}$ and $y_{2,t} = y_{2,t-1} + \varepsilon_{2,t}$, thus both time series are dominated by smooth, long term trend and it seems suitable to specify a relation by static regression $y_{1,t} = \beta y_{2,t} + u_t$. This often produce significant estimate of the unknown parameter β with a large absolute t-ratio and the coefficient of determination close to unity, and the residuals u_t appear to be stationary. But such an empirical result tells us little of the short run relationship between $y_{1,t}$ and $y_{2,t}$. In fact, if the two series are both I(1) then we will often reject the hypothesis of no relationship between them even if none exists. For there to be such a long run relationship, the variables must be cointegrated. The cointegration analysis is therefore very useful in preventing misleading inference and will be discussed in the following section.

3.1 Cointegration

Before any analytical description of the term cointegration is given, consider following illustration [6]. In a crowded open-field park we observe old man being on an afternoon walk and a young lady walking her dog. Both persons are unrelated to each other and their motions are mutually unaffected except they may collide accidentally for a moment. However, the dog is related physically to his mistress by one of those leashes that has cord rolled up inside the handle on a spring. It is intuitive that information about the old man's location tells us nothing about the woman's location, whereas although she and her dog are both individually on a random walk, they cannot wander too far from one another because of the leash. We say that the random processes describing their paths are cointegrated. In other words, if there exists a stationary linear combination of nonstationary random variables, the variables combined are said to be cointegrated.

In the following, first we try to formulate the problem with different representations, then the most favourite tests for cointegration are described. We continue to use the notation from [15].

Representation

A good representation to start with is the following simple bivariate model similar to the dynamic simultaneous model (1.52), that is,

$$\begin{aligned} y_{1,t} + \delta y_{2,t} &= v_t, \qquad & v_t &= \mu_1^* + \rho_1 v_{t-1} + \varepsilon_{1,t}^*, \\ y_{1,t} + \eta y_{2,t} &= w_t, \qquad & w_t &= \mu_2^* + \rho_2 w_{t-1} + \varepsilon_{2,t}^*, \end{aligned} \tag{3.1}$$

where $0 \leq \rho_i \leq 1$, $i = 1, 2$, and $\delta \neq \eta$. The later restriction prevents δ and η from being equal zero at the same time. The μ_1^* and μ_2^* are intercept terms and $\varepsilon_{1,t}^*$, $\varepsilon_{2,t}^*$ are assumed to be standard white noise error processes mutually independent at all lags. The two equations (3.1) reflect that two distinct linear combinations of $y_{1,t}$ and $y_{2,t}$ can be described by AR(1) models. The interpretation of the two linear combinations depends on the values of ρ_1 and ρ_2. In this bivariate case, we may encounter three relevant cases:

a) $\rho_1 = \rho_2 = 1$; Any linear combination of $y_{1,t}$ and $y_{2,t}$ is a random walk variable (possibly with drift if μ_1^* or μ_2^* is unequal to zero). Also $y_{1,t}$ and $y_{2,t}$ are $I(1)$ variables, nonstationary themselves. Since no linear combination of them is stationary, they do not have stochastic trend in common and they are said not to be cointegrated.

b) $0 \leq \rho_i < 1$, $i = 1, 2$; Any linear combination of $y_{1,t}$ and $y_{2,t}$ is stationary AR(1) process, hence $y_{1,t}$ and $y_{2,t}$ are themselves stationary variables. There is no sense in talking about cointegration.

c) $\rho_1 = 1$ and $0 \leq \rho_2 < 1$ (or vice versa); There is one linear combination of $y_{1,t}$ and $y_{2,t}$ which is stationary AR(1) process, while the another combination is a random walk (with drift). When there is such a stationary relationship between $y_{1,t}$ and $y_{2,t}$, which individually have a stochastic trend, cointegration among $y_{1,t}$ and $y_{2,t}$ implies that both series also have a common stochastic trend.

The model framework (3.1) is useful when we analyse relation of two variables, however generalisation to multivariate case becomes complicated, therefore it is convenient to rewrite the dynamic simultaneous model as VAR model (for now, of order one). Let us summarize (3.1) as

$$\begin{bmatrix} 1 & \delta \\ 1 & \eta \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1^* \\ \mu_2^* \end{bmatrix} + \begin{bmatrix} \rho_1 & \delta\rho_1 \\ \rho_2 & \eta\rho_2 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t}^* \\ \varepsilon_{2,t}^* \end{bmatrix}, \tag{3.2}$$

which multiplying both sides with the inverse of the left-hand side matrix and subtracting the one period lagged $\boldsymbol{y}_{t-1}$ from both sides gives

$$\Delta_1 \boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{e}_t \tag{3.3}$$

with $\boldsymbol{y}_t = (y_{1,t}, y_{2,t})'$ and the error series in $\boldsymbol{e}_t = (e_{1,t}, e_{2,t})'$ are functions of $\varepsilon_{1,t}^*$, $\varepsilon_{2,t}^*$, δ and η, and with

$$\boldsymbol{\mu} = \frac{1}{\eta - \delta} \begin{bmatrix} \eta\mu_1^* - \delta\mu_2^* \\ \mu_2^* - \mu_1^* \end{bmatrix}, \tag{3.4}$$

$$\boldsymbol{\Pi} = \frac{1}{\eta - \delta} \begin{bmatrix} \eta\rho_1 - \delta\rho_2 - \eta + \delta & \eta\delta(\rho_1 - \rho_2) \\ \rho_2 - \rho_1 & \eta\rho_2 - \delta\rho_1 - \eta + \delta \end{bmatrix}. \tag{3.5}$$

Again, when

a) $\rho_1 = \rho_2 = 1$, all elements of $\boldsymbol{\Pi}$ have value zero and hence the rank of $\boldsymbol{\Pi}$ is equal zero,

b) $0 \leq \rho_i < 1$, $i = 1, 2$, the matrix $\boldsymbol{\Pi}$ has full rank, $\text{rank}(\boldsymbol{\Pi}) = 2$,

c) $\rho_1 = 1$ and $0 \leq \rho_2 < 1$ (for instance), this is the cointegration case when the (2×2) matrix $\boldsymbol{\Pi}$ equals the outer product of (2×1) matrices, i.e.

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' \tag{3.6}$$

with

$$\boldsymbol{\alpha} = \frac{1}{\eta - \delta} \begin{bmatrix} \delta(1 - \rho_2) \\ -(1 - \rho_2) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} 1 \\ \eta \end{bmatrix} \tag{3.7}$$

Note, that this leads to reduction from 4 to 3 parameters in $\mathbf{\Pi}$. In general, cointegration reduces the number of parameters in VAR model. However, the decomposition is not unique, and we can find other nonlinear parameter restrictions on $\mathbf{\Pi}$ which also correspond with cointegration. In this particular case, the characteristic polynomial of the VAR(1) model (3.3), $|\boldsymbol{I} - (\mathbf{\Pi} + \boldsymbol{I})z| = 0$, yields one solution on the unit circle. Hence, both series have a unit root, while the vector series has only a single unit root, which then represents the common stochastic trend.

The parameter vector $\boldsymbol{\beta}$ is said to contain the *cointegration parameters*, and $\boldsymbol{\beta y}_t$ is the *equilibrium* (or long-run) relation between $y_{1,t}$ and $y_{2,t}$. The parameter matrix $\boldsymbol{\alpha}$ contains the so-called *adjustment parameters*, which reflect the speed of adjustment toward equilibrium. This is easily seen from (3.3) with (3.6) when written as equations

$$\begin{aligned}\Delta_1 y_{1,t} &= \mu_1 + \alpha_1(y_{1,t-1} + \eta y_{2,t-1}) + e_{1,t},\\ \Delta_1 y_{2,t} &= \mu_2 + \alpha_2(y_{1,t-1} + \eta y_{2,t-1}) + e_{2,t}.\end{aligned} \tag{3.8}$$

Notice the term in parentheses. It is the equilibrium common for both equations. Each of the above equations is so-called *error correction model* (ECM), together denoted as VECM.

To help fix the idea, consider again the analogy [6] of the young lady (say, Lin) walking the dog (Spike). Now they are seen staggering out of nearby pub and heading home. Lin has too much to drink and her movement away from the pub is obviously erratic. Spike is also prone to wander aimlessly, randomly attracted by various smells. Now, they don't need to be connected by a leash, Lin is still conscious of being the owner of the dog and Spike will respond to his master's voice. Lin's meandering down the street can be modelled as a random walk along the real line $y_{1,t} - y_{1,t-1} = \varepsilon_{1,t}$. The real line in this case can be taken to be a narrow path leading away from the bar through an open field. Suppose we enter the bar and lose Lin from our sight for a moment. On coming out of the bar a short time later, our best prediction for her current location is where we last saw her. Because her movements are a random walk, she is as likely to be on the path as out in the middle of the field. If the coefficient on $y_{1,t-1}$, her last position, was less then one (in absolute value) then she would tend to return to the path no matter how long we remained in the bar. Similarly, Spike's wandering can also be modelled as a random walk along the real line, $y_{2,t} - y_{2,t-1} = \varepsilon_{1,t}$. If in her stupor Lin notices that Spike is not at her side she will call his name. In response he will trot closer to the source of his name. By the same token, Spike will bark when he realizes that he has wandered off from his mistress and Lin will stagger off in the direction of the bark. Hence we have a long run relationship which recognizes the association between Lin and Spike, $y_{1,t} + \eta y_{2,t} = w_t$ with w_t being stationary. Furthermore, Lin and Spike determine their next "step" according to the system of equations (3.8). The series for the change in, say, Lin's position is determined by the extent to which she and Spike have wandered far apart. In fact, Lin's next step closes the deviation from long run equilibrium in the previous period by the amount α_1.

Now consider a more general case, the VAR(p) model expanded by deterministic terms,

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\delta} t + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{3.9}$$

with k-dimensional Gaussian white noise process $\boldsymbol{\varepsilon}_t$ (with mean zero and variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$), which is convenient to rewrite as vector error correction model (VECM)

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\delta} t + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\Pi} \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{3.10}$$

where $\boldsymbol{\Gamma}_i = (\boldsymbol{\Phi}_1 + \cdots + \boldsymbol{\Phi}_i) - \boldsymbol{I}$, for $i = 1, 2, \ldots p-1$, and $\boldsymbol{\Pi} = (\boldsymbol{\Phi}_1 + \cdots + \boldsymbol{\Phi}_p) - \boldsymbol{I}$. VECM combines short-run dynamics in and among the processes together with long-run relationships (contained in parameter matrix $\boldsymbol{\Pi}$), thus construction of VECM allows to investigate these two types of relations separately. Again, there are three alternative events that may occur with the rank of $\boldsymbol{\Pi}$. The first two cases, as listed above, result in zero or full rank, while the reduced rank refers to cointegration relations between processes. The later case is originally treated by the famous Granger's Representation Theorem[1] (see [31]), which shows that a system of cointegrated time series can be formulated as VAR, ECM and VMA. It is interesting to notice that the matrix $\boldsymbol{\Pi}$ has a counterpart in VMA representation, to be denoted as $\boldsymbol{C}$, for which $\boldsymbol{C}\boldsymbol{\Pi} = \boldsymbol{\Pi}\boldsymbol{C} = \boldsymbol{0}$ and $\operatorname{rank}(\boldsymbol{C}) + \operatorname{rank}(\boldsymbol{\Pi}) = k$ holds. If the rank deficiency allows for decomposition $\boldsymbol{C} = \boldsymbol{\alpha}\boldsymbol{\gamma}'$, then $\boldsymbol{\gamma}$ contains parameters of common stochastic trends. The theorem has greatly influenced all later works concerned in testing for cointegration in a system of nonstationary processes.

In the following we discuss two most widely used methods of testing for cointegration (or common stochastic trends), the first is meant for single-equation representation whereas the second one is based on VAR.

[1] Clive W. J. Granger was awarded the Nobel Prize for Economics in 2003 for methods of economic time series analysis with common trends

Testing and parameters estimation

A simple method to test for cointegration between two (and possibly more) variables, that is despite its limitation still popular among practitioners, allows to estimate the cointegrating vector $\boldsymbol{\beta}$ without necessity of modelling the dynamics, until the itself estimation of ECM. The so-called *Engle-Granger two step method* involves testing the residuals from static regression for stationarity and then using them as equilibrium to estimate the parameters of error correction model. More formally,

1. we first perform the static regression

   $$y_{1,t} = \beta_0 + \beta_1 y_{2,t} + u_t$$

   and use OLS estimates $\hat{u}_t$ of long-run relationship residuals in auxiliary test regression (with an AR(q) model)

   $$\Delta\hat{u}_t = \theta_0 + \rho\hat{u}_{t-1} + \theta_1\Delta\hat{u}_{t-1} + \cdots + \theta_q\Delta\hat{u}_{t-q} + \upsilon_t$$

   to carry out the augmented Dickey-Fuller test, that is to evaluate the t-test for the significance of ρ. When $\rho = 0$, $\hat{u}_t$ has a unit root and hence the linear combination is a nonstationary time series. When $\rho < 0$, i.e. $t(\hat{\rho})$ is significantly negative, $y_{1,t}$ and $y_{2,t}$ are cointegrated. The only departure from standard unit root testing is that the tables of ADF test critical values are not applicable since the series tested has been obtained from regression. Table 3.2 provides correct critical values for analysis of up to 4 variables (source: [38]). Note, that above we described the case with only a constant included in regressions. The critical values are available also for the case with both constant and trend. Finally, the choice of which variable to regress on which in the static regression can be assisted by maximising R^2.

2. Once the variables has been ascertained to be cointegrated, the residual estimates $\hat{u}_t$ are used in ECM

   $$\Delta y_{1,t} = c + \sum_{i=1}^{p_1} a_i \Delta y_{1,t-i} + \sum_{j=1}^{p_2} b_j \Delta y_{2,t-j} + \alpha_1(y_{1,t-1} - \beta_0 - \beta_1 y_{2,t-1}),$$

   instead of the error correction term (in parentheses) to estimate the parameters $(a_1, \ldots, b_1, \ldots, c, \alpha_1)$. Lag orders p_1, p_2 are chosen using standard diagnostic techniques.

The limitation of Engle-Granger procedure comes from imposing a common factor restriction on the dynamics of the relationship between variables. Invalidity of such assumption leads to significant loss of power. There are also several other disadvantages, however from practical point of view, we may found the method being less useful for analysis of more than two variables as the most limiting feature. As the number of possible cointegration relations increases with the number of time series and so does the ambiguity in determining the empirical validity of single equation models, this suggests that multivariate methods may be more appropriate.

One such *multivariate method*, which has been used extensively in applied work, came from *Johansen's approach* and provides a unified framework for estimation and testing in the context of a multivariate vector autoregressive model in error correction form (VECM) with normal errors. The normality assumption allows a neat application of maximum likelihood theory, which produces both the test statistics and the maximum likelihood estimators of the parameters in model. Now consider the VECM (3.10), for which the goal remains twofold: to determine the number of cointegrating vectors in $\boldsymbol{\beta}$ and to estimate parameters $(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Gamma}_1, \ldots \boldsymbol{\Gamma}_{p-1}, \boldsymbol{\Pi}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$ of this model. Because the topic is interesting to understand in more details, we follow [31] and [2] in describing the estimation background. For simplicity of notation, let n be the length of time series with the initial values $\boldsymbol{y}_0, \ldots \boldsymbol{y}_{-p}$ excluded, then let

$$\begin{aligned} \boldsymbol{z}_{0t} &= \Delta \boldsymbol{y}_t \\ \boldsymbol{z}_{1t} &= (1, t, \Delta \boldsymbol{y}'_{t-1}, \ldots \Delta \boldsymbol{y}'_{t-p+1})', \\ \boldsymbol{z}_{2t} &= \boldsymbol{y}_{t-p}, \end{aligned} \tag{3.11}$$

and $\boldsymbol{\Gamma} = (\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Gamma}_1, \ldots \boldsymbol{\Gamma}_{p-1})$ is $k \times [(p-1)k + 2]$ parameter matrix. Thus model (3.10) can be expressed as

$$\boldsymbol{z}_{0t} = \boldsymbol{\Gamma} \boldsymbol{z}_{1t} + \boldsymbol{\Pi} \boldsymbol{z}_{2t} + \boldsymbol{\varepsilon}_t, \tag{3.12}$$

and the normal equations for this model become

$$\boldsymbol{M}_{01} = \boldsymbol{\Gamma} \boldsymbol{M}_{11} + \boldsymbol{\Pi} \boldsymbol{M}_{21}, \tag{3.13}$$

where $\boldsymbol{M}_{ij} = n^{-1} \sum_{t=1}^{n} \boldsymbol{z}_{it} \boldsymbol{z}'_{jt}$, $i, j = 0, 1, 2$, are product moment matrices. Thus parameter matrix

$$\boldsymbol{\Gamma} = \boldsymbol{M}_{01} \boldsymbol{M}_{11}^{-1} - \boldsymbol{\Pi} \boldsymbol{M}_{21} \boldsymbol{M}_{11}^{-1} \tag{3.14}$$

incorporated back into the model leads to the equation

$$\boldsymbol{r}_{0t} = \boldsymbol{\Pi} \boldsymbol{r}_{2t} + \boldsymbol{\varepsilon}_t \tag{3.15}$$

where $\boldsymbol{r}_{it} = \boldsymbol{z}_{it} - \boldsymbol{M}_{i1}\boldsymbol{M}_{11}^{-1}\boldsymbol{z}_{1t}$, $i = 0, 2$, are the residuals we would obtain by regressing $\Delta\boldsymbol{y}_t$ and $\boldsymbol{y}_{t-p}$ (respectively) on $1, t, \Delta\boldsymbol{y}'_{t-1}, \dots \Delta\boldsymbol{y}_{t-p+1}$. The parameters $\boldsymbol{\Pi}$ and $\boldsymbol{\Sigma_\varepsilon}$ are to be estimated via maximizing the logarithm of conditional likelihood function

$$\ln L(\boldsymbol{\Pi}, \boldsymbol{\Sigma_\varepsilon}) = -\frac{pn}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma_\varepsilon}| - \frac{1}{2}\sum_{t=1}^{n}(\boldsymbol{r}_{0t} - \boldsymbol{\Pi}\boldsymbol{r}_{2t})'\boldsymbol{\Sigma_\varepsilon}(\boldsymbol{r}_{0t} - \boldsymbol{\Pi}\boldsymbol{r}_{2t}), \tag{3.16}$$

such that

$$\hat{\boldsymbol{\Pi}} = \boldsymbol{S}_{02}\boldsymbol{S}_{22}^{-1} \tag{3.17}$$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}} = \boldsymbol{S}_{00} - \boldsymbol{S}_{02}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{20} \tag{3.18}$$

where $\boldsymbol{S}_{ij} = n^{-1}\sum_{t=1}^{n}\boldsymbol{r}_{it}\boldsymbol{r}'_{jt} = \boldsymbol{M}_{ij} - \boldsymbol{M}_{i1}\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{1j}$, $i, j = 0, 2$, are sums of squared residuals. Because the likelihood function is maximal for $\sum_{t=1}^{n}(\boldsymbol{r}_{0t} - \boldsymbol{\Pi}\boldsymbol{r}_{2t})'\boldsymbol{\Sigma_\varepsilon}(\boldsymbol{r}_{0t} - \boldsymbol{\Pi}\boldsymbol{r}_{2t}) = 0$, its maximum can be written as

$$L_{\max}^{-2/n} = |\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}| \tag{3.19}$$

omitting constant term $(2\pi)^p$. The estimate of $\boldsymbol{\Pi}$ inserted into (3.14) gives the estimate of $\boldsymbol{\Gamma}$.

Now assume the model where $\text{rank}(\boldsymbol{\Pi}) = r < k$ makes the long run parameter matrix $\boldsymbol{\Pi}$ to be a product of two $k \times r$ matrices, i.e. $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, thus containing r cointegration vectors. The regression (3.15) becomes $\boldsymbol{r}_{0t} = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{r}_{2t} + \boldsymbol{\varepsilon}_t$ and for fixed $\boldsymbol{\beta}$ the least square estimates will be

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}) = \boldsymbol{S}_{02}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{22}\boldsymbol{\beta})^{-1}, \tag{3.20}$$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}(\boldsymbol{\beta}) = \boldsymbol{S}_{00} - \boldsymbol{S}_{02}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{22}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{S}_{20} = \boldsymbol{S}_{00} - \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})(\boldsymbol{\beta}'\boldsymbol{S}_{22}\boldsymbol{\beta})^{-1}\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})', \tag{3.21}$$

and the maximized likelihood function

$$L_{\max}^{-2/n}(\boldsymbol{\beta}) = |\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}(\boldsymbol{\beta})| = |\boldsymbol{S}_{00} - \boldsymbol{S}_{02}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{22}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{S}_{20}| = \tag{3.22}$$

$$= |\boldsymbol{S}_{00}||\boldsymbol{\beta}'\boldsymbol{S}_{22}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{S}_{20}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{02}\boldsymbol{\beta})|/|\boldsymbol{\beta}'\boldsymbol{S}_{22}\boldsymbol{\beta}| = \tag{3.23}$$

$$= |\boldsymbol{S}_{00}||\boldsymbol{\beta}'(\boldsymbol{S}_{22} - \boldsymbol{S}_{20}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{02})\boldsymbol{\beta}|/|\boldsymbol{\beta}'\boldsymbol{S}_{22}\boldsymbol{\beta}| \tag{3.24}$$

The maximum likelihood estimator of $\boldsymbol{\beta}$ is found minimizing $L_{\max}^{-2/n}(\boldsymbol{\beta})$, that is by solving the eigenvalue problem[2]

$$|\lambda\boldsymbol{S}_{22} - \boldsymbol{S}_{20}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{02}| = 0. \tag{3.25}$$

[2] The generalized eigenvalue problem (3.25) can be reduced to simpler form $|\lambda\boldsymbol{I} - \boldsymbol{A}| = \boldsymbol{0}$, where $\boldsymbol{A}$ is symmetric, first by decomposition $\boldsymbol{S}_{22} = \boldsymbol{C}\boldsymbol{C}'$ for some non-singular $k \times k$ matrix $\boldsymbol{C}$, then by solving $|\lambda\boldsymbol{I} - \boldsymbol{C}^{-1}\boldsymbol{S}_{20}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{02}\boldsymbol{C}'^{-1}| = 0$. We get the same eigenvalues, but different eigenvectors $\boldsymbol{w}_1, \dots \boldsymbol{w}_k$, for which the equality $\hat{\boldsymbol{v}}_i = \boldsymbol{C}'^{-1}\boldsymbol{w}_i$ holds.

This gives eigenvalues $\hat{\lambda}_1 > \cdots > \hat{\lambda}_k$ and eigenvectors $\hat{V} = (\hat{\boldsymbol{v}}_1, \ldots \hat{\boldsymbol{v}}_k)$ normalized to satisfy $\hat{V}'\boldsymbol{S}_{22}\hat{V} = \boldsymbol{I}$. Then

$$\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{v}}_1, \ldots \hat{\boldsymbol{v}}_r) \tag{3.26}$$

and

$$L_{\max}^{-2/n} = |\boldsymbol{S}_{\mathbf{00}}| \prod_{i=1}^{r} (1 - \hat{\lambda}_i). \tag{3.27}$$

The estimates of other parameters are found by inserting $\hat{\boldsymbol{\beta}}$ into the corresponding equations, e.g., $\hat{\boldsymbol{\alpha}} = \boldsymbol{S}_{02}\hat{\boldsymbol{\beta}}$. One can interpret $\hat{\lambda}_i$ as a squared canonical correlation between $\Delta\boldsymbol{y}_t$ and $\boldsymbol{y}_{t-p}$ conditional on $\Delta\boldsymbol{y}_{t-1}, \ldots \Delta\boldsymbol{y}_{t-p+1}$. Thus the estimate of the 'most stable' relations between the levels are those that correlate most with the stationary process $\Delta\boldsymbol{y}_t$ corrected for lagged differences and deterministic terms.

To test for (the order of) cointegration means testing the rank of $\boldsymbol{\Pi}$ or the number of cointegration vectors. Let H_r denote the hypothesis that (or the model (3.10) in which) rank($\boldsymbol{\Pi}$) $= r$, so that H_k means system consisting purely of $I(0)$ time series. Note that in model H_r the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not identified, since $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\boldsymbol{\xi}^{-1}(\boldsymbol{\beta}\boldsymbol{\xi}')'$ for any $\boldsymbol{\xi}$ of full rank, but that one can estimate the spaces spanned by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, and the parameters in $\boldsymbol{\beta}$ can be estimated if they are identified or normalised suitably. Thus cointegration analysis is formulated as the problem of making inference on the cointegration space, sp($\boldsymbol{\beta}$), and the adjustment space, sp($\boldsymbol{\alpha}$). If we want to estimate individual coefficients it is necessary to normalise $\boldsymbol{\beta}$ or impose restrictions so that the parameters become identified (see [32] for further details).

To find r, we formulate a nested sequence of hypotheses

$$H_0 \subset \cdots \subset H_r \subset \cdots \subset H_k$$

so the test that there are (at most) r cointegrating relations is the test of H_r in H_k, performed by comparing likelihood functions (3.27). The likelihood ratio (LR) statistic is then

$$Q^{-2/n}(H_r|H_k) = \frac{L_{\max}^{-2/n}(H_r)}{L_{\max}^{-2/n}(H_k)} = \frac{|\boldsymbol{S}_{\mathbf{00}}| \prod_{i=1}^{r}(1 - \hat{\lambda}_i)}{|\boldsymbol{S}_{\mathbf{00}}| \prod_{i=1}^{k}(1 - \hat{\lambda}_i)} = \frac{1}{\prod_{i=r+1}^{k}(1 - \hat{\lambda}_i)} \tag{3.28}$$

which after taking a logarithm gives the so-called *trace test* statistic

$$LR_{Trace} = -2\ln Q(H_r|H_k) = -n \sum_{i=r+1}^{k} \ln(1 - \hat{\lambda}_i). \tag{3.29}$$

We begin with testing whether there is no cointegration (H_0) versus the general alternative of k such relations or less (H_k). If this is rejected, we test H_1 (of at most 1 relation) against H_k, and so on. The rank of $\mathbf{\Pi}$ is estimated as r if H_r is the first hypothesis which cannot be rejected. Another useful test is given by testing H_r in H_{r+1}, i.e. the significance of the estimated eigenvalues themselves. This yields so-called λ-*max test* statistic

$$LR_{\lambda-max} = -2\ln Q(H_r|H_{r+1}) = -n\ln(1-\hat{\lambda}_{r+1}) \tag{3.30}$$

The asymptotic null distribution of these statistics, expressed in terms of vector Brownian motion functionals, is derived and summarised in, e.g., [31][33], and depends on the specification of deterministic terms. Some of the critical values for both test statistics are provided here in Table 3.2 (source: [39],[15]).

Unsurprisingly, in small samples the distribution of the LR test is not well approximated by limiting distribution. There were suggested several ways of corrections to the test statistic, some authors propose degrees of freedom correction, e.g., $LR_{Trace} = -(n-pk)\sum_{i=r+1}^{k}\ln(1-\hat{\lambda}_i)$, while for example Johansen in his more recent work used the idea of Bartlett correction.

Deterministic terms

A characteristic feature of the error-correction formulation is the inclusion of both differences ($\Delta\boldsymbol{y}_t$) and levels ($\boldsymbol{y}_t$) in the same model, allowing us to investigate both short-run and long-run effects in the data. However, the interpretation of the coefficients in terms of dynamic effects is difficult, and this is true also for the trend and the constant term, as well as other deterministic terms (like dummy variables for modelling seasonal or exceptional effects) included in the model. In the following we will discuss the dual role of trend and constant in cointegrated VAR model, which is important to understand, partly because the asymptotic distributions of the cointegration tests are not invariant to the specifications of these components, and furthermore, the properties of the resulting formulation may prove undesirable for (say) forecasting, by inadvertently retaining unwanted components such as quadratic trends ([30]).

To illustrate the idea, recall (3.10) with constant and trend component decomposed into

$$\boldsymbol{\mu} = \boldsymbol{\mu}_d + \boldsymbol{\alpha}\boldsymbol{\mu}_c, \tag{3.31}$$

$$\boldsymbol{\delta} = \boldsymbol{\delta}_d + \boldsymbol{\alpha}\boldsymbol{\delta}_c, \tag{3.32}$$

so that VECM attains the form

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_d + \boldsymbol{\delta}_d t + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\alpha}(\boldsymbol{\mu}_c + \boldsymbol{\delta}_c t + \boldsymbol{\beta}' \boldsymbol{y}_{t-p}) + \boldsymbol{\varepsilon}_t \tag{3.33}$$

or

$$\boldsymbol{z}_{0t} = (\boldsymbol{\mu}_d, \boldsymbol{\delta}_d, \boldsymbol{\Gamma}_1, \dots \boldsymbol{\Gamma}_{p-1}) \boldsymbol{z}_{1t} + \boldsymbol{\alpha}(\boldsymbol{\mu}_c, \boldsymbol{\delta}_c, \boldsymbol{\beta}') \boldsymbol{z}_{2t} + \boldsymbol{\varepsilon}_t \tag{3.34}$$

where again $\boldsymbol{z}_{0t} = \Delta \boldsymbol{y}_t$ and $\boldsymbol{z}_{1t} = (1, t, \Delta \boldsymbol{y}_1', \dots \Delta \boldsymbol{y}_{p-1}')$, but now

$$\boldsymbol{z}_{2t} = (1, t, \boldsymbol{y}_{t-p}')'. \tag{3.35}$$

We can always choose $\boldsymbol{\mu}_c$ and $\boldsymbol{\delta}_c$ such that equilibrium error (the term in parentheses) has mean zero, so that equation $\mathrm{E}[\Delta \boldsymbol{y}_t] = \boldsymbol{\mu}_d + \boldsymbol{\delta}_d t$ allows us to see that $\boldsymbol{\mu}_d \neq 0$ corresponds to constant growth in the variables $\boldsymbol{y}_t$, whereas $\boldsymbol{\delta}_d \neq 0$ corresponds to linear trends in growth, and so quadratic trends in the variables. To correctly interpret the cointegrated model, one has to understand the dual role of constant term and deterministic linear trend therein, i.e. the distinction between the part of the deterministic component that belongs to the cointegration relations, and the part that belongs to the differences. Below we list five of the most frequently used models arising from restricting the deterministic components in (3.33):

1. No restrictions on trend and intercept. With *unrestricted* parameters $\boldsymbol{\mu}$ and $\boldsymbol{\delta}$ the model is consistent with linear trend in the differenced series $\Delta \boldsymbol{y}_t$ and, thus, quadratic trends in $\boldsymbol{y}_t$. Although quadratic trends may sometimes improve the fit within sample, forecasting outside the sample is likely to produce implausible results, therefore it is preferable to treat this case with care, find out what induced the apparent quadratic growth and, if possible, increase the information set of the model (e.g., by including appropriate exogenous variable).

2. $\boldsymbol{\delta}_d = 0$. The trend is *restricted* to lie in the cointegration space, but the constant is unrestricted in the model. This allows linear, but precludes quadratic, trends in the levels of data ($\boldsymbol{y}_t$). Because $\boldsymbol{\delta}_c \neq 0$, these linear trends in the variables do not cancel in the cointegrating relations, so the model contains *trend-stationary* relations. Such model also include the case when a single variable is trend-stationary.

3. $\boldsymbol{\delta} = 0$. Since the constant term $\boldsymbol{\mu}$ is unrestricted, there are still linear trends in the data, but no deterministic trends in any cointegration relations.

Table 3.1: Restrictions on deterministic terms

Case	Terms in VECM		Determ. components in		Regressors	
	$\boldsymbol{\mu}$	$\boldsymbol{\delta}$	variables	equilibrium	$\boldsymbol{z}_{1t}$	$\boldsymbol{z}_{2t}$
1.	unrest.	unrest.	quadr.	linear	$(1, t, \Delta\boldsymbol{y}'_{t-1}, \ldots \Delta\boldsymbol{y}'_{t-p+1})'$	$(1, t, \boldsymbol{y}'_{t-p})'$
2.	unrest.	$\boldsymbol{\delta}_d = 0$	linear	linear	$(1, \Delta\boldsymbol{y}'_{t-1}, \ldots \Delta\boldsymbol{y}'_{t-p+1})'$	$(1, t, \boldsymbol{y}'_{t-p})'$
3.	unrest.	absent	linear	constant	$(1, \Delta\boldsymbol{y}'_{t-1}, \ldots \Delta\boldsymbol{y}'_{t-p+1})'$	$(1, \boldsymbol{y}'_{t-p})'$
4.	$\boldsymbol{\mu}_d = 0$	absent	constant	constant	$(\Delta\boldsymbol{y}'_{t-1}, \ldots \Delta\boldsymbol{y}'_{t-p+1})'$	$(1, \boldsymbol{y}'_{t-p})'$
5.	absent	absent	zero	zero	$(\Delta\boldsymbol{y}'_{t-1}, \ldots \Delta\boldsymbol{y}'_{t-p+1})'$	$\boldsymbol{y}_{t-p}$

4. $\boldsymbol{\delta} = 0$, $\boldsymbol{\mu}_d = 0$. The constant term is restricted to lie in the cointegration space, i.e. only equilibrium means are different from zero.

5. $\boldsymbol{\delta} = 0$, $\boldsymbol{\mu} = 0$. The model excludes all deterministic components in the data, with both $\mathrm{E}[\Delta\boldsymbol{y}_t] = 0$ and $\mathrm{E}[\boldsymbol{\beta}'\boldsymbol{y}_t] = 0$. Since an intercept is generally needed to account for the initial level of measurements, the restriction $\boldsymbol{\mu} = 0$ can be justified only in exceptional cases.

In empirical work, usually it is clear whether there is linear deterministic trend in some (or all) of the variables. It might, however, be more difficult to know if they cancel in the cointegrating relations or not. Fortunately, we do not need to know beforehand, all the above (1.–5.) cases – being expressed as linear restrictions on the deterministic components of VAR model – can be tested. Table 3.1 summarises the restrictions and is helpful in performing test and estimation procedures.

Denote the hypothesis of r cointegration relations for particular case as $H_r^{(case)}$. According to [11], a consistent test procedure follows the idea of testing $H_r^{(1)}$ if $H_r^{(2)}$ has been rejected. That means testing the hypotheses

$$H_0^{(2)}, H_0^{(1)}, H_1^{(2)}, H_1^{(1)}, \ldots H_{k-1}^{(2)}, H_{k-1}^{(1)},$$

sequentially against the unrestricted alternative and stopping whenever the hypothesis is "accepted". Correspondingly, if $H_r^{(3)}$ appears more appropriate, the hypotheses testing sequence is

$$H_0^{(4)}, H_0^{(3)}, H_1^{(4)}, H_1^{(3)}, \ldots H_{k-1}^{(4)}, H_{k-1}^{(3)}.$$

It is possible, but not very likely, that an insignificant value is followed by a significant statistic. An example would be: reject $H_0^{(2)}$, accept $H_1^{(2)}$, and reject $H_2^{(2)}$. This could be indicative of more general model mis-specification.

Table 3.2: Asymptotic critical values for cointegration tests at 5% significance level

Engle-Granger method

Deterministic terms in regressions	Number of variables 2	3	4
constant	-3.34	-3.74	-4.12
constant and trend	-3.78	-4.10	-4.40

Johansen's method $(Trace/\lambda - max)$

Case	$k-r$ 1	2	3	4	5
1.	11.55/11.55	23.37/18.04	39.04/23.97	58.57/30.31	82.18/36.65
2.	12.52/12.52	25.86/19.38	42.92/25.83	63.87/32.12	88.79/38.32
3.	8.18/8.18	17.95/14.90	31.52/21.07	48.28/27.14	70.60/33.32
4.	9.17/9.17	20.25/15.88	35.19/22.30	54.09/28.58	76.96/34.80
5.	4.13/4.13	12.32/11.23	24.28/17.80	40.17/24.16	60.06/30.42

Restriction on cointegrating vector

In the previous we talked out testing hypotheses on certain restrictions upon deterministic terms. Sometimes, it is reasonable to restrict also long-run relationships represented by $\boldsymbol{\beta}$ to comply with certain theoretical assumptions. Whether the assumptions really hold, we may verify by testing hypotheses on parameters $\boldsymbol{\beta}$.

One such hypothesis can be defined as

$$H^a : \ \boldsymbol{\beta} = \boldsymbol{H}\boldsymbol{\varphi}, \tag{3.36}$$

where $(k \times s)$ matrix $\boldsymbol{H}$, $r \leq s < k$, reduces $\boldsymbol{\beta}$ to the $(s \times r)$ parameter matrix $\boldsymbol{\varphi}$. The hypothesis corresponds to restriction $\boldsymbol{H}'_\perp \boldsymbol{\beta} = 0$, where $\boldsymbol{H}_\perp$ is orthogonal complement of $\boldsymbol{H}$ (i.e. $\boldsymbol{H}'_\perp \boldsymbol{H} = \boldsymbol{0}$). In this way, various hypothesis can be stated, e.g, if one of the columns of $\boldsymbol{H}$ is of form $(1, -1, -1, 0, 0)'$, this means that in every cointegration vector the first three parameters are of equal magnitude but opposed signs. Another example: by leaving i-th row in matrix $\boldsymbol{H}$ zero valued we may prevent $y_{i,t}$ from entering the cointegration relationships.

To test the hypothesis (3.36), we compare the estimated eigenvalues $\hat{\zeta}_i$, $i = 1, \dots r$, from

$$|\zeta \boldsymbol{H}'\boldsymbol{S}_{22}\boldsymbol{H} - \boldsymbol{H}'\boldsymbol{S}_{20}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{02}\boldsymbol{H}| = 0 \tag{3.37}$$

with $\hat{\lambda}_i$ from (3.25) via the likelihood ratio test statistic

$$LR = -2\ln Q(H^a|H_r) = n\sum_{i=1}^{r}\ln\left[(1-\hat{\zeta}_i)/(1-\hat{\lambda}_i)\right]. \tag{3.38}$$

Under the null hypothesis and conditional on correct value of r, the test statistic in (3.38) asymptotically follows the $\chi^2(r(k-s))$ distribution.

The hypothesis that some cointegrating vectors are known can be formulated as

$$H^b : \boldsymbol{\beta} = (\boldsymbol{G}, \boldsymbol{\psi}) \tag{3.39}$$

where $\boldsymbol{G}$ is known $(k \times r_1)$ and $\boldsymbol{\psi}$ unknown $(k \times r_2)$ matrix, while $r_1 + r_2 = r$. In particular, we may test that individual (i-th) variable is stationary by defining $\boldsymbol{G}$ as unity vector with 1 in i-th row. Notice that the same is testable by putting the unity vector into $\boldsymbol{H}$ from previous hypothesis. Thus the stationarity of a single component of $\boldsymbol{y}_t$ is a special case of cointegration.

Under hypothesis H^b the eigenvalue problem to be solved is

$$|\zeta\boldsymbol{S}_{22.G} - \boldsymbol{S}_{20.G}\boldsymbol{S}_{00.G}^{-1}\boldsymbol{S}_{02.G}| = 0 \tag{3.40}$$

where $S_{ij.G} = S_{i2}\boldsymbol{G}(\boldsymbol{G}'S_{22}\boldsymbol{G})^{-1}\boldsymbol{G}'S_{2j},\ i,j = 0,2$, and the resulting normalised eigenvectors constitute parameter matrix $\boldsymbol{\psi}$. The eigenvalues ζ are, naturally, different from that by H^a. Now likelihood ratio test statistic $LR = -2\ln Q(H^b|H_r)$ asymptotically follows χ^2 distribution with $r_1(k-r)$ degrees of freedom.

There can be formulated many other and more general linear hypotheses, obviously we could benefit also from testing restrictions on adjustment parameter $\boldsymbol{\alpha}$. If interested, further information can be found in lots of publications, e.g., [32] or [2].

Common stochastic trends

Finding r cointegration relations imply that there are $k-r$ common stochastic trends in the system. Their estimation can sometimes be of practical interest, particularly in order to get insight into the driving non-stationary forces. Among several works of various authors, Gonzalo and Granger proposed a method, which explicitly exploits the duality of cointegration and stochastic trends. The canonical correlation approach, used to find those combinations of the elements of $\boldsymbol{y}_t$ which have maximum partial correlation with the stationary variables, can also be reversed to find those combinations, which have minimum correlation. The relevant eigenvalue problem, which is dual version of (3.25), then becomes

$$|\lambda\boldsymbol{S}_{00} - \boldsymbol{S}_{02}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{20}| = 0 \tag{3.41}$$

and the solutions are the same eigenvalues as before, but now we obtain different eigenvectors $\hat{\boldsymbol{w}}_1, \ldots \hat{\boldsymbol{w}}_k$. In case of r cointegration relations, stochastic trend variables can be constructed as $\hat{\boldsymbol{w}}'_{r+1}\boldsymbol{y}_t$ to $\hat{\boldsymbol{w}}'_k\boldsymbol{y}_t$.

3.2 Deterministic trend

Although common deterministic trend has already been treated in the previous section - within the context of common stochastic trend (recall the case 2 of restricting deterministic terms) - we still owe to say some remarks on this topic. Firstly in the single equations analysis we consider interesting to assign a geometrical aspect to common feature like trend, assuming it is indeed deterministic. Secondly, a test whether (deterministic) trends in a system of multivariate time series have statistically the same slope is given.

Geometrical aspect

As geodesy deals with geometric variables to a great extent, it often uses Cartesian coordinate system and various transformation rules within to analyse and to display observables and computed quantities. Consider bivariate observations $\boldsymbol{y}_t = (y_{1,t}, y_{2,t})'$, that represent position of a point in certain horizontal (topocentric) coordinate system. Interested ourselves in process that causes data to display trending behaviour, we try to decompose $\boldsymbol{y}_t$ into components according to the stationarity, in other words we look for linear combination

$$\begin{aligned} u_{1,t} &= c_{11}y_{1,t} + c_{12}y_{2,t} \\ u_{2,t} &= c_{21}y_{1,t} + c_{22}y_{2,t} \end{aligned} \tag{3.42}$$

such that $u_{1,t}$ represents a common trend direction and $u_{2,t}$ is a stationary trend-free variable, orthogonal to $u_{1,t}$. In the light of our geometrical application, it's easy to rewrite the general common trend problem into familiar transformation

$$\begin{aligned} u_{1,t} &= \quad \cos(\alpha)\, y_{1,t} + \sin(\alpha)\, y_{2,t} \\ u_{2,t} &= -\sin(\alpha)\, y_{1,t} + \cos(\alpha)\, y_{2,t} \end{aligned} \tag{3.43}$$

as shown in Figure 3.1. The angle α can be determined via regression

$$y_{1,t} = a_1 + b_1 t + e_{1,t}\ , \qquad y_{2,t} = a_2 + b_2 t + e_{2,t} \tag{3.44}$$

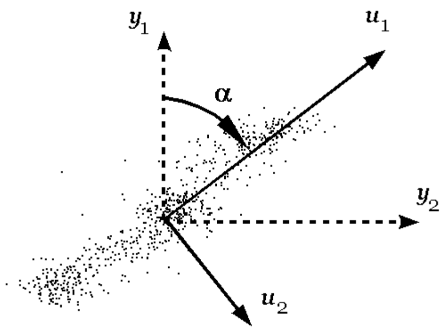

Figure 3.1: Transformation into common trend direction

with regression parameters $a_1, \ldots b_2$. If we place (3.44) into (3.43) and focus on series $u_{2,t}$, which is supposed to be trend-free, then

$$\begin{aligned} u_{2,t} &= -(a_1 + b_1 t + e_{1,t}) \sin\alpha + (a_2 + b_2 t + e_{2,t}) \cos\alpha, \\ &= (a_2 \cos\alpha - a_1 \sin\alpha) + \underbrace{(b_2 \cos\alpha - b_1 \sin\alpha)}_{0}\ t + (e_{2,t} \cos\alpha - e_{1,t} \sin\alpha) \end{aligned}$$

(linear trend term in $u_{2,t}$ is eliminated), so

$$\tan\alpha = \frac{b_2}{b_1} \tag{3.45}$$

and the angle α is used in (3.43) to obtain the new variable $\boldsymbol{u}_t = (u_{1,t}, u_{2,t})'$.

If stochastic trend is present, the above decomposition is applicable as well, however the angle is then determined from

$$y_{2,t} = a_0 + b_0 y_{1,t}\ , \qquad \tan\alpha = b_0 \tag{3.46}$$

and it looses its former geometric interpretation.

We would get $\boldsymbol{u}_t$ equally well from cointegration analysis of bivariate time series as $(\boldsymbol{v}_1, \boldsymbol{w}_2)'\boldsymbol{y}_t$, where $\boldsymbol{v}_1$ and $\boldsymbol{w}_2$ are eigenvectors from solution of (3.25) and (3.41), respectively.

Testing for common deterministic trend slopes

If time series $\boldsymbol{y}_t$ was tested for the presence of stochastic trend and no random walk but linear deterministic trend was detected, it may be of interest to examine if two or more of such a trend-stationary time series have the same deterministic trend. Such a hypothesis can be written as linear restrictions on the slope parameters across the series and we can apply the multivariate linear trend tests [18].

Consider the multivariate trend model

$$y_{1,t} = \mu_1 + \delta_1 t + e_{1,t} \tag{3.47}$$
$$y_{2,t} = \mu_2 + \delta_2 t + e_{2,t} \tag{3.48}$$
$$\vdots \tag{3.49}$$
$$y_{k,t} = \mu_k + \delta_k t + e_{k,t} \tag{3.50}$$

that can be compactly written as $\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\delta} t + \boldsymbol{e}_t$, where $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ are classical constant and linear trend parameters, $\boldsymbol{e}_t$ denotes residuals and k is again the number of time series. We are interested in testing hypotheses of the form

$$H_0 : \boldsymbol{R\delta} = \boldsymbol{r}\,, \qquad H_1 : \boldsymbol{R\delta} \neq \boldsymbol{r}, \tag{3.51}$$

where $\boldsymbol{R}$ is $q \times k$ matrix and $\boldsymbol{r}$ is a $q \times 1$ vector of known constants. The linear hypotheses of (3.51) are quite general, they include linear hypotheses on slopes within given trend equations ($q = k - 1$) as well as joint trend hypotheses across equations ($q = k$). According to [18] we apply two F-tests, both test statistics are functions of the following HAC (heteroskedasticity autocorrelation) variance covariance matrix estimator. Let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\delta}}$ denote the stacked single equation OLS estimates and $\hat{\boldsymbol{u}}_t = \boldsymbol{y}_t - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\delta}} t$ be the residuals. Define

$$\hat{\boldsymbol{\Omega}}_{HAC} = \hat{\boldsymbol{\Gamma}}_0 + \sum_{j=1}^{n-1} (1 - \frac{j}{L})(\hat{\boldsymbol{\Gamma}}_j + \hat{\boldsymbol{\Gamma}}_j')\,, \tag{3.52}$$

which is the Bartlett kernel estimator, where $\hat{\boldsymbol{\Gamma}}_j = \frac{1}{n}\sum_{t=j+1}^{n} \hat{\boldsymbol{e}}_t \hat{\boldsymbol{e}}_{t-j}'$ and L is the truncation lag or bandwidth. Usually a consistent $\hat{\boldsymbol{\Omega}}_{HAC}$ is needed, yet [18] offers an alternative, where $L = n$. Although it does not result in consistent estimator, valid testing is still possible because of asymptotic proportionality and moreover it has certain advantage coming from the choice of bandwidth. It holds that

$$\hat{\boldsymbol{\Omega}}_{L=n} = \frac{2}{n^2} \sum_{t=1}^{n} \hat{\boldsymbol{S}}_t \hat{\boldsymbol{S}}_t'\,, \tag{3.53}$$

where $\hat{\boldsymbol{S}}_t = \sum_{j=1}^{t} \hat{\boldsymbol{e}}_j$. It is also convenient to express an element of $\hat{\boldsymbol{\delta}}$ as

$$\hat{\delta}_i = \left(\sum_{t=1}^{n} \tilde{t}^2 \right)^{-1} \left(\sum_{t=1}^{n} \tilde{t} y_{i,t} \right) \qquad \text{for } i = 1, 2, \ldots, k\,, \tag{3.54}$$

where $\bar{t} = \frac{1}{n}\sum_{t=1}^{n} t$ and $\tilde{t} = t - \bar{t}$. Now the first of test statistics can be defined

$$F_1 = (\boldsymbol{R}\hat{\delta} - \boldsymbol{r})' \left[\boldsymbol{R}(\sum_{t=1}^{n} \tilde{t}^2)^{-1} \hat{\boldsymbol{\Omega}}_{L=n} \boldsymbol{R}' \right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\delta}} - \boldsymbol{r})/q\,. \tag{3.55}$$

Following [18] we also consider an alternative to $\hat{\boldsymbol{\Omega}}_{L=n}$ which is constructed using $\tilde{t}\hat{\boldsymbol{e}}_t$ instead of $\hat{\boldsymbol{e}}_t$. Because $\tilde{t}\hat{\boldsymbol{e}}_t$ is not a vector of stationary time series, establishing consistency of HAC estimator would be difficult if even feasible, yet again if we use $L = n$, the asymptotic behaviour of the HAC estimator can be derived. We can write

$$\tilde{\boldsymbol{\Omega}}_{L=n} = \frac{2}{n^2} \sum_{t=1}^{n} \tilde{\boldsymbol{S}}_t \tilde{\boldsymbol{S}}_t'\,, \tag{3.56}$$

where $\tilde{\boldsymbol{S}}_t = \sum_{j=1}^{t} (j - \bar{t})\hat{\boldsymbol{e}}_j$, and then the second test statistic is

$$F_2 = n(\boldsymbol{R}\hat{\boldsymbol{\delta}} - \boldsymbol{r})' \left[\boldsymbol{R}(\frac{1}{n}\sum_{t=1}^{n} \tilde{t}^2)^{-1} \tilde{\boldsymbol{\Omega}}_{L=n} (\frac{1}{n}\sum_{t=1}^{n} \tilde{t}^2)^{-1} \boldsymbol{R}' \right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\delta}} - \boldsymbol{r})/q\,. \tag{3.57}$$

The null hypothesis in (3.51) is rejected if test statistic F_1 (F_2) exceeds critical value given for q restrictions in [18], Table 3 (Table 2, alternatively). It is worth noting that due to practical reasons indices of the F-statistics has been swapped in our work.

The asymptotic distribution theory for these F statistics is nonstandard and was developed for the case where the errors are covariance stationary. Simulation evidence reported by [18] suggests that the F-tests suffers much less from over-rejection problem caused by strong positive serial correlation than the compared standard alternative, whereas the power of F-s is slightly lower. Finite sample simulation evidence in [18] also suggested that the performance of the tests are improved when $\hat{\boldsymbol{\Omega}}$ estimator is computed using VAR(1) prewhitening. However, this we do not do here.

The standard alternative to F_1 and F_2 is a Wald test based on consistent $\hat{\Omega}_{HAC}$ estimator, which uses the same Bartlett kernel. For $\hat{\Omega}_{HAC}$ to be consistent, the bandwidth L must increase as the sample increases but at the slower rate. As referred in [18], the rate $\sqrt[3]{n}$ minimizes the approximate mean square error for $\hat{\Omega}$ and considering this in (3.52), the Wald test is defined as

$$W = (\boldsymbol{R}\hat{\boldsymbol{\delta}} - \boldsymbol{r})' \left[\boldsymbol{R}(\sum_{t=1}^{n} \tilde{t}^2)^{-1} \hat{\boldsymbol{\Omega}}_{HAC} \boldsymbol{R}' \right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\delta}} - \boldsymbol{r})\,. \tag{3.58}$$

Asymptotic distribution of the Wald test is χ^2 with q degrees of freedom.

For illustration we recommend to see an interesting application of this theory in [20] too.

3.3 Seasonality

Investigating common trends across observed variables may often be influenced by seasonality. Instead of preprocessing through seasonal adjustment methods, it is most sensible to use seasonally unadjusted data to study common long-run non-seasonal trends, which furthermore may provoke us to ask if also the seasonal pattern detected individually within each component time series (see section 1.4) can be a consequence of one and the same underlying process. However, the issue can be quite exhaustive and as was said it's far behind the scope of this thesis to cope with all types of seasonal models. Therefore if stochastic seasonality is found either through seasonal unit roots or periodic properties, interested reader is kindly referred to [15] and [16] (to see also references therein), we presume here the case when data contain seasonal deterministics. This assumption mostly comes true when dealing with natural processes. In either case, the motivation for investigating common cycles remains the same as with trends: to reduce the number of estimated parameters and to improve forecasting performance.

At first we consider simple case when we may impose cross-equation parameter restrictions. Inspired by the example in [15] let's define model

$$\begin{aligned} y_{1,t} &= \delta_1 t + \sum_{s=1}^{S} \omega_{1,s} D_{s,t} + \phi_1 y_{1,t-1} + \varepsilon_{1,t} \\ y_{2,t} &= \delta_2 t + \sum_{s=1}^{S} \omega_{2,s} D_{s,t} + \phi_2 y_{2,t-1} + \varepsilon_{2,t} \end{aligned} \tag{3.59}$$

where the dummy variables $D_{s,t}$ are defined as in (1.46). These two series would have their deterministic seasonality in common when $\omega_{2,s} = \psi\omega_{1,s}$ for some non-zero value of ψ. This common feature amounts to S parameter restrictions and can be seen to substantially decrease the number of parameters. If joint estimation of (3.59) gives the covariance matrix $\hat{\mathbf{\Sigma}}_u$ while imposing the restrictions results in $\hat{\mathbf{\Sigma}}_r$, then its appropriateness can be judged by likelihood ratio test, $n(\ln|\hat{\mathbf{\Sigma}}_r| - \ln|\hat{\mathbf{\Sigma}}_u|) \sim \chi^2(S)$.

In [14] a general approach to testing for common seasonality is proposed. Consider the multivariate regression

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Omega} \boldsymbol{v}_t + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \Delta \boldsymbol{y}_{t-i} + \mathbf{\Pi} \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t \tag{3.60}$$

where $\boldsymbol{v}_t$ is an m-dimensional vector of deterministic seasonal variables at period t, $\boldsymbol{\Omega}$ is $k \times m$ parameter matrix and the other terms are denoted as in (3.10). A variable is said to have a deterministic seasonal feature if after incorporating the effects of lagged dependent variables, the variable will have a corresponding row of $\boldsymbol{\Omega}$ non-zero. If there exists a $k \times l$ matrix $\boldsymbol{\vartheta}$ such that $\boldsymbol{\vartheta}'\boldsymbol{\Omega} = \mathbf{0}$ then these linear combinations of the series have no deterministic seasonality and this seasonality is common. In this case, the rank of $\boldsymbol{\Omega}$ can be at most equal to $q = \min(k - l, m)$. The $k \times m$ matrix $\boldsymbol{\Omega}$ may then be written as the product of two matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$ both of rank q and of order $k \times q$ and $q \times m$ respectively. The test of the (reduced) rank of $\boldsymbol{\Omega} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ can then be performed as a test of the number of zero canonical correlation coefficients between $\Delta \boldsymbol{y}_t$ and $\boldsymbol{v}_t$ conditional on the lagged values of $\Delta \boldsymbol{y}_t$ and on $\boldsymbol{y}_{t-p}$. Such an approach is also familiar from the cointegration tests of Johansen in section 3.1. Preserve the notation such that only (3.11) changes into

$$\begin{aligned} \boldsymbol{z}_{0t} &= \Delta \boldsymbol{y}_t, \\ \boldsymbol{z}_{1t} &= (\Delta \boldsymbol{y}'_{t-1}, \ldots \Delta \boldsymbol{y}'_{t-p+1}, \boldsymbol{y}_{t-p})', \\ \boldsymbol{z}_{2t} &= \boldsymbol{v}_t, \end{aligned} \tag{3.61}$$

then the canonical correlations[3] can be found as the square roots of the eigenvalues of the $k \times k$ matrix $\boldsymbol{A} = \boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{02}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{20}$ where the $\boldsymbol{S}_{i,j}$ are the moment matrices between $\boldsymbol{z}_{it}$ and $\boldsymbol{z}_{jt}$, $i, j = 0, 2$. Denote the ordered eigenvalues of $\boldsymbol{A}$ as $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k < 1$ and the corresponding eigenvectors $\boldsymbol{w}_i$, $i = 1, 2 \ldots k$ normalised such that $\boldsymbol{w}'_i \boldsymbol{S}_{00} \boldsymbol{w}_i = 1$. Then the likelihood ratio test statistic

$$LR = -(n - p) \sum_{i=1}^{r} \ln(1 - \lambda_i) \tag{3.62}$$

has an asymptotic $\chi^2(r[k[m-1] + r])$ distribution under the null that the smallest r canonical correlations are zero. Under the alternative that the r-th canonical correlation is nonzero the statistic will be larger so that a one tailed test is appropriate. When there are exactly r zero canonical correlations, the linear combinations $\boldsymbol{w}'_i \Delta \boldsymbol{y}_t$, $i = 1, 2, \ldots r$, are uncorrelated with any linear combinations of the elements of the seasonal variables in $\boldsymbol{v}_t$, whereas the linear combinations $\boldsymbol{w}'_i \Delta \boldsymbol{y}_t$, $i = r + 1, \ldots k$, are maximally correlated with the seasonals, carry all the seasonal information and can be thought of as the seasonal common feature which determines the seasonal pattern in all of the $(k - 1)$ elements of $\Delta \boldsymbol{y}_t$ in a way which makes $\boldsymbol{v}_t$ superfluous.

[3]Imagine for few seconds that no lagged values of the left hand side variables are needed to render the errors white noise, i.e. $\boldsymbol{\Gamma}_i = \mathbf{0}$ for all i and $\boldsymbol{\Pi} = \mathbf{0}$ as well. Then the canonical correlations coefficients measure the correlations between linear combinations of the elements of $\Delta \boldsymbol{y}_t$ and linear combinations of the elements of $\boldsymbol{v}_t$.

As [14] continues, three potentially serious, and closely related problems arise. First, the variables in $\boldsymbol{v}_t$ must encompass the kind of seasonality which is common among the elements of $\boldsymbol{y}_t$. Second, the rank and parameters of any potential cointegration (corresponding to term $\boldsymbol{\Pi}\boldsymbol{y}_{t-p}$) must be determined. And third, the augmentation by lagged values of the dependent variables must render the errors white noise. In case the first condition is not met, the test is deficient and will lack power to detect the common seasonal features. If the second or third condition is violated, the size of the test may be higher than the selected level of significance. However, the augmentation must not be overparametrised either as this may reduce the power of the test.

Additional problem arise for data with sampling frequency much smaller than are the seasonal (annual, daily or other natural) frequencies. To reduce the potentially large number of parameters (because of large S) in multivariate models, several plausible strategies can be used, see for example [16]. One way is to impose restrictions on parameters by a certain smooth function, of which the most familiar example being trigonometric polynomials (as in (1.44) for instance).

The specification of the proper augmentation is done by estimating a general VAR (3.60) and then testing down in order to obtain the most parsimonious lag structure, rendering the errors multivariate white noise. Then cointegration is tested and a vector error correction model specified. Finally, restrictions on seasonals are tested.