8.5 Dynamic factor analysis, common trends and co-integration

The essence of a STM is that it is set up in terms of components with distinct dynamic properties. A natural generalisation of the SUTSE class is therefore to allow the series to have certain components in common. just as prior considerations may help in formulating components, so they may help in deciding which components are candidates for common factors. The fact that some of the components may be non-stationary does not pose any difficulties. Rather it is an asset, because the more different are the properties of components, the easier they are to distinguish. Factor analysis of stationary time series, as in Geweke and Singleton (1981) and Velu et al (1986), can be less straightforward.

A common factor model contains components with fewer than N elements. Recognition of common factors yields models which may not only have an interesting interpretation, but may also provide more efficient inferences and forecasts. In terms of a SUTSE model, the presence of common factors means that the covariance matrices of the relevant disturbances are less than full rank.

Common factors in the trend imply co-integration. At its simplest cointegration means that there exists a linear combination of two I(1) series which is stationary. A more general definition is given later in sub-section 8.5.6.

The basic ideas can be illustrated with a bivariate local level model

$$\begin{aligned} y_{1t} &= \mu_{1t} + \varepsilon_{1t}, \qquad \mu_{1t} = \mu_{1t-1} + \eta_{1t}, \qquad t = 1, ..., T, \\ y_{2t} &= \mu_{2t} + \varepsilon_{2t}, \qquad \mu_{2t} = \mu_{2t-1} + \eta_{2t} \end{aligned} \tag{8.5.1}$$

The covariance matrix of $(\eta_{1t}, \eta_{2t})'$ may be written

$$\mathbf{\Sigma}_\eta = \begin{bmatrix} \sigma^2_{1\eta} & \rho_\eta \sigma_{1\eta}\sigma_{2\eta} \\ \rho_\eta \sigma_{1\eta}\sigma_{2\eta} & \sigma^2_{2\eta} \end{bmatrix} \tag{8.5.2}$$

where ρ_η is the correlation. The model can be transformed as follows:

$$\begin{aligned} y_{1t} &= \mu_{1t} + \varepsilon_{1t}, \\ y_{2t} &= \pi\mu_{1t} + \overline{\mu}_t + \varepsilon_{2t}, \end{aligned} \tag{8.5.3}$$

where $\pi = \rho_\eta \sigma_{2\eta}/\sigma_{1\eta}$ and the covariance matrix of the disturbances driving the new multivariate random walk $(\mu_{1t}, \overline{\mu}_t)'$ is

$$Var \begin{bmatrix} \eta_{1t} \\ \overline{\eta}_t \end{bmatrix} = \begin{bmatrix} \sigma^2_{1\eta} & 0 \\ 0 & \sigma^2_{2\eta} - \rho^2_\eta \sigma^2_{2\eta} \end{bmatrix}.$$

In other words by setting $\eta_{2t} = \pi\eta_{1t} + \overline{\eta}_t$, two uncorrelated levels, μ_{1t} and $\overline{\mu}_t$, based respectively on η_{1t} and $\overline{\eta}_t$, are obtained. (In a Gaussian model, these expressions can be thought of as coming from the expression for the mean and variance of η_{2t}, conditional on η_{1t}; see the appendix of ch3).

If $\rho_\eta = \pm 1$, then $\overline{\mu}_t$ is constant and there is only one common trend which will now be denoted as $\mu_t^\dagger$ rather than μ_{1t}. Hence

$$\begin{aligned} y_{1t} &= \mu_t^\dagger + \varepsilon_{1t}, \quad t = 1, ..., T, \\ y_{2t} &= \pi\mu_t^\dagger + \overline{\mu} + \varepsilon_{2t} \end{aligned} \tag{1}$$

If $\pi = 1$, the trend in the second series is always at a constant distance, $\overline{\mu}$, from the trend in the first series, that is $\mu_{2t} = \overline{\mu} + \mu_{1t}$. This is known as *balanced growth.* When the series are in logarithms, multiplying the first trend by $\exp(\overline{\mu})$ gives the second. If $\overline{\mu} = 0$ the two trends are said to be *identical.*

Pre-multiplying the observation vector in (1) by the vector $(-\pi, 1)$ gives

$$y_{2t} = \pi y_{1t} + \overline{\mu} + \varepsilon_t \tag{2}$$

where $\varepsilon_t = \varepsilon_{2t} - \pi\varepsilon_{1t}$ and $\pi = sgn(\rho)\sigma_2/\sigma_1$. Since the linear combination $y_{2t} - \pi y_{1t}$ is stationary, the series are co-integrated. The full model can be written as

$$\begin{aligned} y_{1t} &= \mu_t^\dagger + \varepsilon_{1t}, \\ y_{2t} &= \pi y_{1t} + \overline{\mu} + \varepsilon_t \end{aligned} \tag{3}$$

This is equivalent to the common trends model, (1), and it can be estimated directly. It ties in closely with the econometrics literature. Writing the model in this form enables us to determine whether the co-integrating equation may be treated in isolation so that π can be sensibly estimated simply by regressing y_{2t} on y_{1t}. It also enables us to formulate tests of the hypothesis that the two series are, in fact, co-integrated.

Interest rates Figure 8.5.1 shows UK quarterly long and short term interest rates, as represented by the yield on 20-year gilts and the 91-day Treasury bill rate. The 'spread' between the two rates is an important variable in assessing theories of the term structure of interest rates; see Mills (1993, p.26-8, p.225). After 1970 the series become considerably more volatile as the UK entered a period of relatively high inflation. We therefore first estimate the multivariate local level plus cycle model using data up to the end of 1970 only. The results show that the trends are perfectly correlated, while the cycle is present only in the short rate series, where it reduces to an AR(1) with coefficient .68. Thus the trend in the short series is just a linear function of the actual observations on the long series. The factor loading matrix shows that the common trend loads .27 on the long rate and .29 on the short (the same as the respective standard deviations of the level disturbance in the two series). If these loadings are taken to be the same, the multivariate model implies that the spread variable, ie $y_{1t} - y_{2t}$, is a stationary AR(1).

Constructing a completely satisfactory model for the whole period is difficult. Given the nature of the data, it is perhaps not surprising that the model fitted to the earlier period fails the normality and heteroscedasticity tests. Nevertheless it still provides an informative description of the data. The random walk trends are no longer perfectly correlated and a plot shows how in the 1980s they move so as to be much closer together. The cross-plot is also interesting in the way it shows a shift in the relationship.

The first three sub-sections below generalize the ideas of common trends and co-integration in the local level model. Tests of co-integration and common trends are described and the models are then extended to handle stochastic slopes, cycles and seasonals.

8.5.1 Local Level

Consider the Gaussian multivariate local level model of (8.2.1). Suppose, initially, that $\mathbf{\Sigma}_\eta$ is of full rank and that it can be partitioned as

$$\mathbf{\Sigma}_\eta = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}$$

where $\mathbf{\Sigma}_{11}$ is $K \times K$ and $\mathbf{\Sigma}_{21} = \mathbf{\Sigma}_{12}'$ is $r \times K$. Correspondingly, $\boldsymbol{\eta}_t = (\boldsymbol{\eta}_{1t}', \boldsymbol{\eta}_{2t}')'$ where $\boldsymbol{\eta}_{1t}$ is $K \times 1$ and $\boldsymbol{\eta}_{2t}$. It is convenient to re-write the model in the following way:

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_K & \mathbf{0} \\ \mathbf{\Pi} & \mathbf{I}_r \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_t^\dagger \\ \overline{\boldsymbol{\mu}}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{bmatrix}, \quad t = 1, ..., T \tag{4}$$

where $\mathbf{y}_t$ is partitioned into a $K \times 1$ vector $\mathbf{y}_{1t}$ and an $R \times 1$ vector $\mathbf{y}_{2t}$, with $R = N - K$, $\boldsymbol{\varepsilon}_\mathbf{t}$ is similarly partitioned, $\mathbf{\Pi}$ is an $r \times K$ matrix of coefficients and the $K \times 1$ vector $\boldsymbol{\mu}_t^\dagger$ follows a multivariate random walk

$$\boldsymbol{\mu}_t^\dagger = \boldsymbol{\mu}_{t-1}^\dagger + \boldsymbol{\eta}_t^\dagger, \quad \boldsymbol{\eta}_t^\dagger \sim NID(\mathbf{0}, \mathbf{\Sigma}_\eta^\dagger), \tag{5}$$

where $\boldsymbol{\eta}_t^\dagger$ is a $K \times 1$ vector and $\mathbf{\Sigma}_\eta^\dagger$ is a $K \times K$ positive definite matrix. The $r \times 1$ vector $\overline{\boldsymbol{\mu}}_t$ also follows a multivariate random walk with a disturbance vector $\overline{\boldsymbol{\eta}}_t$ which has a p.d covariance matrix $\overline{\boldsymbol{\Sigma}}_\eta$ and is uncorrelated with $\boldsymbol{\eta}_t^\dagger$. This way of writing the model comes about by defining

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_K & \mathbf{0} \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}_r \end{bmatrix}$$

and then forming

$$\mathbf{y}_t = \mathbf{L}^{-1}(\mathbf{L}\boldsymbol{\mu}_t) + \boldsymbol{\varepsilon}_t$$

where the disturbance driving the transformed state, $\mathbf{L}\boldsymbol{\mu}_t$, has a covariance matrix

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{bmatrix}$$

The model is now of the form (4) with $\boldsymbol{\Pi} = \boldsymbol{\Sigma}_{\mathbf{21}}\boldsymbol{\Sigma}_{\mathbf{11}}^{\mathbf{-1}}$ and $\overline{\boldsymbol{\Sigma}}_\eta = \boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. Note that applying the $\mathbf{L}$ transformation to $(\boldsymbol{\eta}_{1t}', \boldsymbol{\eta}_{2t}')'$ yields $(\boldsymbol{\eta}_{1t}', \boldsymbol{\eta}_{2t}'|\boldsymbol{\eta}_{1t})' = (\boldsymbol{\eta}_t^{\dagger'}, \overline{\boldsymbol{\eta}}_t')'$; see chapter3, appendix.

If the rank of the covariance matrix $\boldsymbol{\Sigma}_\eta$ is $K < N$, an appropriate ordering of the series enables the model to be written in the form (4) with $\boldsymbol{\mu}_t^\dagger$ containing K common levels, or trends, and $\boldsymbol{\Sigma}_\eta^\dagger$ continuing to be positive definite while $\overline{\boldsymbol{\Sigma}}_\eta = \mathbf{0}$. Thus (4) becomes

$$\begin{aligned} \mathbf{y}_{1t} &= \boldsymbol{\mu}_t^\dagger + \boldsymbol{\varepsilon}_{1t} \\ \mathbf{y}_{2t} &= \boldsymbol{\Pi}\boldsymbol{\mu}_t^\dagger + \overline{\boldsymbol{\mu}} + \boldsymbol{\varepsilon}_{2t} \end{aligned} \tag{6}$$

The model consisting of (6) and (5) is therefore the common trends model.

Unless there are constraints on $\overline{\boldsymbol{\mu}}$, estimation can be based on the original model simply by setting $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Pi}^\dagger\boldsymbol{\Sigma}_\eta^\dagger\boldsymbol{\Pi}^{\dagger\prime}$, where $\boldsymbol{\Pi}^\dagger = (\mathbf{I}, \boldsymbol{\Pi}')'$. Given that $\boldsymbol{\Sigma}_\eta^\dagger$ is positive definite, a necessary and sufficient condition for all the diagonal elements in $\boldsymbol{\Sigma}_\eta$ to be positive is that $\boldsymbol{\Pi}$ contains no null rows. Time domain estimation is still valid if this condition is violated, but it means that some of the series are stationary so there are implications for co-integration.

Estimation in the frequency domain The frequency domain likelihood function for the multivariate local level when there are no common trends is (8.2.16) with

$$\mathbf{G}_j = \boldsymbol{\Sigma}_\eta + 2\left(1 - \cos\lambda_j\right)\boldsymbol{\Sigma}_\varepsilon, \quad j = 0, ..., T-1 \tag{8.5.10}$$

In the common trends model, (6), $\boldsymbol{\Sigma}_\eta$ in the above expression is replaced by $\boldsymbol{\Pi}^\dagger\boldsymbol{\Sigma}_\eta^\dagger\boldsymbol{\Pi}^{\dagger\prime}$. This creates a difficulty at $j = 0$, since $\mathbf{G}_0 = \boldsymbol{\Pi}^\dagger\boldsymbol{\Sigma}_\eta^\dagger\boldsymbol{\Pi}^{\dagger\prime}$ and this is of rank $N < K$. However, as shown in Fernandez (1990), the term involving $\mathbf{G}_0$ in (8.2.16), that is

$$-\frac{1}{2}\log|\mathbf{G}_0| - \pi tr\left[\mathbf{G}_0^{-1}\mathbf{I}\left(\lambda_0\right)\right]$$

may be replaced by

$$-\frac{1}{2}\log\left|\mathbf{\Sigma}_{\eta}^{\dagger}\right| - tr\left[\mathbf{\Sigma}_{\eta}^{\dagger-1}\mathbf{\Pi}^{\ddagger}\mathbf{I}\left(\lambda_{0}\right)\mathbf{\Pi}^{\ddagger\prime}\right] \tag{8.5.11}$$

where

$$\mathbf{\Pi}^{\ddagger} = \left(\mathbf{\Pi}^{\dagger\prime}\mathbf{\Pi}^{\dagger}\right)^{-1}\mathbf{\Pi}^{\dagger\prime} \tag{8.5.12}$$

If the model has an $K \times 1$ vector of drifts, β, included in the random walks, that is (5) becomes

$$\boldsymbol{\mu}_{t}^{\dagger}= \boldsymbol{\mu}_{t-1}^{\dagger}+\boldsymbol{\beta}+\boldsymbol{\eta}_{t}^{\dagger},$$

then, for a given $\mathbf{\Pi}^{\dagger}$, the frequency domain estimator of β is

$$\widetilde{\beta} = \mathbf{\Pi}^{\ddagger}\overline{\Delta\mathbf{y}} = \left(\mathbf{\Pi}^{\dagger\prime}\mathbf{\Pi}^{\dagger}\right)^{-1}\mathbf{\Pi}^{\dagger\prime}\overline{\Delta\mathbf{y}}, \tag{7}$$

where $\overline{\Delta\mathbf{y}} = (\mathbf{y}_t - \mathbf{y}_1)/(T-1)$; see exercise ?. Thus $\boldsymbol{\beta}$ can easily be concentrated out of the likelihood function.

8.5.2 Factor loadings and rotation

When there is more than one common trend, issues of interpretation arise. This leads on to factor rotations. The first step is to formulate the common level model in such a way that the common factors are uncorrelated with each other and have unit variance, that is

$$\mathbf{y}_t = \mathbf{\Theta}\boldsymbol{\mu}_t^{*}+\overline{\boldsymbol{\mu}}_{\theta} + \boldsymbol{\varepsilon}_{\mathbf{t}}, \quad \boldsymbol{\varepsilon}_{\mathbf{t}} \sim NID(\mathbf{0}, \mathbf{\Sigma}_{\epsilon}), \tag{8}$$

$$\boldsymbol{\mu}_t^{*}= \boldsymbol{\mu}_{t-1}^{*}+\boldsymbol{\eta}_t^{*}, \quad \boldsymbol{\eta}_t^{*} \sim NID(\mathbf{0}, \mathbf{I}_K),$$

where $\mathbf{\Theta}$ is an $N \times K$ matrix of *factor loadings* of the form $\mathbf{\Theta} = \mathbf{\Pi}^{\dagger}(\mathbf{\Sigma}_{\eta}^{\dagger})^{-1/2}$, $\boldsymbol{\mu}_{\theta}$ is an $N \times 1$ vector in which the first K elements are zeroes and the last $N-K$ elements are contained in the vector $\overline{\boldsymbol{\mu}}$, and $\boldsymbol{\mu}_t^{*} = (\mathbf{\Sigma}_{\eta}^{\dagger})^{1/2}\boldsymbol{\mu}_t^{\dagger}$. If $\mathbf{\Sigma}_{\eta}^{\dagger\,1/2}$ is lower triangular, then $\mathbf{\Theta}$ is such that its elements, θ_{ij}, are constrained to be zero for $j > i, i = 1, ..., K$. Note that if the first K rows of $\mathbf{\Theta}$ were not constrained the model would not be identifiable.

When there is more than one common factor, they are not unique and a *factor rotation* may give components with a more interesting interpretation. Let $\mathbf{H}$ be a $K \times K$ orthogonal matrix. The matrix of factor loadings and the vector of common trends can then be redefined as $\mathbf{\Theta}^{\ddagger} = \mathbf{\Theta}\mathbf{H}'$ and $\boldsymbol{\mu}_t^{\ddagger} = \mathbf{H}\boldsymbol{\mu}_t^{\dagger}$ yielding the model

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{\Theta}^{\ddagger}\boldsymbol{\mu}_t^{\ddagger}+\boldsymbol{\mu}_{\theta} + \boldsymbol{\varepsilon}_{\mathbf{t}}, \quad \boldsymbol{\varepsilon}_{\mathbf{t}} \sim NID(\mathbf{0}, \mathbf{\Sigma}_{\epsilon}) \\
\boldsymbol{\mu}_t^{\ddagger} &= \boldsymbol{\mu}_{t-1}^{\ddagger}+\boldsymbol{\eta}_t^{\ddagger}, \quad \boldsymbol{\eta}_t^{\ddagger} \sim NID(\mathbf{0}, \mathbf{I}_K)
\end{aligned}$$

The distribution of the observations remains unchanged and the covariance matrix of the disturbance vector is still scalar, so that the common trends are still mutually uncorrelated

A number of methods for carry out rotations have been developed in the classical factor analysis literature. These may be employed here. For example with two factors a commonly used rotating matrix is

$$\mathbf{H} = \begin{bmatrix} \cos\lambda & -\sin\lambda \\ \sin\lambda & \cos\lambda \end{bmatrix}$$

with the angle, λ, being set by a graphical method. The aim is often to give a factor significant loadings, perhaps all positive, on some variables while the other variables get loadings near zero.

Stochastic volatility in exchange rates An application concerned with modelling the volatility of four exchange rates for the US dollar is described later in sub-section 10.7.??; see also Harvey, Ruiz and Shephard (1994). After a suitable transformation of the observations, a multivariate local level with two common trends was fitted. ML estimation gave factor loadings as shown in the first two columns of table 8.5.1. For clockwise rotation, setting λ to 16.23^0 gives a loading of zero for the first factor on the third series, the Japanese Yen. Setting the angle to 352.48^0 gives a loading of zero for the second factor on the German mark and very small loadings on the other two European currencies. This can be seen clearly in figure 8.5.2. It can be seen from the sizes of the loadings that in the rotation of 16.23^0 the first common trend is the dominant factor for the European currencies, with the second common trend representing a world factor which affects all currencies including the Yen. In the second rotation, the second common trend is primarily associated with the Yen, while the first is mainly European. The message in the two rotations is essentially the same, but it is perhaps slightly clearer in the second.

.

Table 8.5.1 Factor loadings in stochastic volatility model before and after rotation

	Before rotation		$\lambda = 16 \cdot 23^0$		$\lambda = 352 \cdot 48^0$	
Pound/Dollar	0.108	0	0.103	0.030	0.107	-0.014
DM/Dollar	0.102	0.014	0.095	0.042	0.103	0
Yen/Dollar	0.016	0.054	0	0.056	0.023	0.051
SF/Dollar	0.095	0.023	0.085	0.048	0.097	0.010

8.5.3 Co-integration

The presence of common trends implies co-integration. In the local level model, (6), there exist $r = N - K$ co-integrating vectors. To see that this is the case let $\mathbf{A}$ be an $r \times N$ matrix partitioned as $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$. Then pre-multiplying (6) by $\mathbf{A}$ gives

$$\mathbf{A}\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{1t} + \mathbf{A}_2\mathbf{y}_{2t} = (\mathbf{A}_1 + \mathbf{A}_2\mathbf{\Pi})\boldsymbol{\mu}_t^{\dagger} + \mathbf{A}_2\overline{\boldsymbol{\mu}} + \mathbf{A}_1\boldsymbol{\varepsilon}_{1t} + \mathbf{A}_2\boldsymbol{\varepsilon}_{2t}, \quad t = 1, ..., T. \tag{9}$$

The r series in $\mathbf{A}\mathbf{y}_t$ are stationary, and hence $\mathbf{A}$ consists of co-integrating vectors, if $\mathbf{A_1} + \mathbf{A_2\Pi} = \mathbf{0}$. Note that this condition is the same as requiring that $\mathbf{A\Sigma}_\eta = \mathbf{0}$ in (8.2.1).

Balanced growth in macroeconometrics Suppose there is a single common trend in the logarithms of income, consumption and investment, so that $N = 3$, $K = 2$. If there is balanced growth then $\mathbf{\Pi} = (1, 1)'$. The common trend is associated with income and the two co-integrating equations correspond to the 'great ratios' of consumption and investment to income; see King *et.al.* (1991). In other words

$$\mathbf{A} = [-\mathbf{\Pi} \quad \mathbf{I}_2] = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

As in (8.5.5), the common trend system in (6) can be transformed to an equivalent co-integrating system. Pre-multiplying by an $N \times N$ matrix

$$\begin{bmatrix} \mathbf{I}_K & \mathbf{0} \\ \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \tag{10}$$

leaves the first K equations the same while, following (9), the last r become

$$\mathbf{A}_1\mathbf{y}_{1t} + \mathbf{A}_2\mathbf{y}_{2t} = \mathbf{A}_2\overline{\boldsymbol{\mu}} + \mathbf{A}_1\boldsymbol{\varepsilon}_{1t} + \mathbf{A}_2\boldsymbol{\varepsilon}_{2t}, \quad t = 1, ..., T. \tag{11}$$

If $\mathbf{A} = (-\mathbf{\Pi}, \mathbf{I})$ the system becomes

$$\mathbf{y}_{1t} = \boldsymbol{\mu}_t^{\dagger} + \boldsymbol{\varepsilon}_{1t},$$

$$\mathbf{y}_{2t} = \mathbf{\Pi}\mathbf{y}_{1t} + \overline{\boldsymbol{\mu}} + \boldsymbol{\varepsilon}_t, \tag{12}$$

where $\boldsymbol{\varepsilon}_\mathbf{t} = \boldsymbol{\varepsilon}_{2t} - \mathbf{\Pi}\boldsymbol{\varepsilon}_{1t}$.Thus the second set of equations consists of co-integrating relationships, while the first set contains the common trends. This is a special case of the *triangular representation* of a co-integration model analysed in Phillips (1991, 1994). Estimation of the common trends model (6) and the triangular co-integration model should, in principle, give the same results. However, the triangular system may be easier to estimate as the elements of $\mathbf{\Pi}$ can be treated as the coefficients of explanatory variables. Note that the the likelihood of the transformed model is the same as the original in this case since the determinant of the transforming matrix, (10), is $|\mathbf{A}_2| = |\mathbf{I}|$, which is one.

If there are other components such as (similar) cycles and seasonals, a transformation similar to (12) can be applied. Unless there are restrictions on these components in the original formulation, there are no restrictions in the triangular form. In other words the seasonals and cycles can be specified as before and estimated with the transformed parameters. This result does not hold if $\boldsymbol{\psi}_t$ is a VAR rather than a set of similar cycles.

Since $\Delta \mathbf{y}_{1t}$ can, in principle, be modelled by a VAR, many of the results in Phillips (1991, 1994) apply when (12) is estimated as a system. In particular, the $t-$statistics on elements of $\mathbf{\Pi}$ can be treated as asymptotically normal. It is interesting to note that estimating the co-integrating equations in isolation will yield superconsistent estimators of the parameters in $\mathbf{\Pi}$; see Stock (1987). Unfortunately, these estimates are often badly biased in small samples and unless $\boldsymbol{\epsilon}_{1t}$ and $\boldsymbol{\varepsilon}_{\mathbf{t}}$ happen to be independent they will be inefficient. The bias can be corrected by adding leads and lags of $\Delta \mathbf{y}_{1t}$ to the right hand side of the equation as suggested by Saikkonen (1991) and Stock and Watson (1993). This effectively leads to weak exogeneity and so, from the point of view of estimating $\mathbf{\Pi}$, it is similar to working with the full system. However, for the present model a large number of leads and lags may be needed to effectively deal with the bias. Furthermore, this approach does not easily generalise to situations where there are other components in the equations. Other methods, such as the fully modified ordinary least squares estimator of Phillips and Hansen (1990), are possible but the safest course of action is probably to work with the full system in order to ensure efficient inference.

The co-integrating equations in (11) may have a specific interpretion. As a result we may wish to estimate them rather than work with the system in (12). This raises issues concerned with identifiability and there is a close link with the classical theory of simultaneous equation systems in econometrics. This is explored in section 8.8. If overidentifying restrictions can be placed on $\mathbf{A}$ there is the possibility of gains in efficiency.

8.5.4 Testing the null hypothesis of co-integration

Suppose a system contains a single co-integrating relationship in which the coefficients are given. An example would be balanced growth in the bivariate model (1) with $\pi = 1$. Multiplying the observations in (1) by the proposed co-integrating coefficients gives

$$y_{2t} = \pi y_{1t} + \overline{\mu}_t + \varepsilon_{2t} \tag{kk}$$

This suggests that a test of the validity of the co-integrating relationship be carried out by testing against the presence of a random walk component in $y_{2t} - \pi y_{1t}$ using the stationarity test of (5.4.1).

GNP and Investment- The η test of 5.4.?? can also be used to determine whether GNP and Investment exhibit balanced growth, that is they are co-integrated with their difference, $i - y$, being a stationary process. Fitting a random walk (without drift) plus a stochastic cycle gives the parameter estimates

$$\widetilde{P} = 19.36, \widetilde{\phi} = 0.88, \widetilde{\sigma}_\kappa = 0.0411, \widetilde{\sigma}_\eta = 0.0027$$

A plot of $i - y$ with the smoothed level shows that the change in the level over time appears rather small and this is confirmed by the 0.277 value for the test statistic which is clearly well below the 5% critical value of 0.461.

More generally, if $\mathbf{A}$ in (9) satisfies $\mathbf{A}_1 + \mathbf{A}_2\mathbf{\Pi} = \mathbf{0}$, then $\mathbf{Ay}_t$ is stationary and the rows of $\mathbf{A}$ constitute a set of r co-integrating vectors. It may be the case that economic theory specifies the $\mathbf{A}$ matrix directly, because of knowledge of the co-integrating relationships or common trends, as in the balanced growth example of the previous sub-section. To test the null hypothesis that the specified co-integrating relationships are valid, the multivariate random walk test of sub-section 8.4.2 is applied to the $R \times 1$ vector $\mathbf{Ay}_t$. Its limiting distribution under the null hypothesis is, of course, $CvM(R)$. The test can be modified to allow for serial correlation, as described towards the end of sub-section 8.4.2.

Convergence of economies Hobijn and Franses (2000) use the nonparametric test to determine whether per capita productivity levels have converged in a group of economies. Under the relative convergence hypothesis, a mean is fitted and the asymptotic distribution of the test statistic applied to a set of N pairwise comparisons is $CvM(N)$. Under absolute, or perfect, convergence, the mean is taken to be zero and the asymptotic distribution is $CvM_0(N)$.

Suppose now that the coefficients of the variables in the hypothesized co-integrating relationships are not pre-specified. The test described above will not, in general, have the same asymptotic distribution if it is formed from the residuals obtained by regressing $\mathbf{y}_{2t}$ on $\mathbf{y}_{1t}$. As noted in the previous subsection, straightforward OLS needs to be modified to produce an estimator of the co-integrating vector with good small sample properties. With a single co-integrating vector, Shin(1994) shows how the CvM critical values need to be changed when the stationarity test statistic is constructed from the residuals which result when leads and lags of $\Delta\mathbf{y}_{2t}$ are included as regressors. Harris and Inder (1993) produce corresponding tables based on an approach derived from the fully modified ordinary least squares estimator of Phillips and Hansen (1990). See also Choi and Ahn (1995).

Note that in the bivariate model, a test of co-integration is a test of whether the correlation, ρ_η, is ± 1.

8.5.5 Testing for a specified number of common trends

Nyblom and Harvey (2000) give a test for a specified number of common trends in the multivariate local level model, that is

$$H_0 : rank(\mathbf{\Sigma}_\eta) = K \qquad \text{against} \quad H_1 : rank(\mathbf{\Sigma}_\eta) > K$$

for $K < N$. In a bivariate model a null of one common trend amounts to testing $\left|\rho_\eta\right| = 1$ against $\left|\rho_\eta\right| < 1$.

The test is based on the matrices $\mathbf{S}$ and $\mathbf{C}$ defined in (8.4.14). Let $\ell_1 \geq \cdots \geq \ell_N$ be the ordered eigenvalues of $\mathbf{S}^{-1}\mathbf{C}$, given by

$$|\mathbf{C} - \ell_j\widehat{\mathbf{\Sigma}}| = 0, \quad j = 1, \ldots, N. \tag{13}$$

The $\eta(N)$ test statistic is the sum of these eigenvalues, but it can be shown that when the rank of $\mathbf{\Sigma}_\eta$ is $K^\dagger$, the limiting distribution of $T^{-1}\eta(N)$ is the limiting distribution of T^{-1} times the sum of the $K^\dagger$ largest eigenvalues. This suggests basing a test of the hypothesis that $rank(\mathbf{\Sigma}_\eta) = K$ on the sum of the $N - K$ smallest eigenvalues, that is

$$\zeta(K, N) = \ell_{K+1} + \cdots + \ell_N, \quad K = 1, ..., N-1. \tag{14}$$

Then if $K^\dagger > K$ the relatively large values taken by the first $K^\dagger - K$ of these eigenvalues will tend to lead to the null hypothesis being rejected. This is the *common trends* test. If K is allowed to be zero in (14), then $\zeta(0, N) = \eta(N)$.

A rationalisation for the test can also be given in terms of co-integration. As was demonstrated earlier, if the rank of $\mathbf{\Sigma}_\eta$ is K then there are K common trends and $R = N - K$ linear combinations of $\mathbf{y}_t$, denoted $\mathbf{Ay}_t$, that are white noise. Thus under H_0 we have $\mathbf{A\Sigma}_\eta\mathbf{A}' = \mathbf{0}$ and if $\mathbf{A}$ can be specified, the multivariate random walk test is applied to $\mathbf{Ay}_t$ as in the previous subsection. The test statistic may be written as

$$\eta(\mathbf{A}) = tr(\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}')^{-1}\mathbf{ACA}' \tag{15}$$

If $\mathbf{A}$ cannot be specified, a natural test statistic for the null hypothesis that there are K common trends, against the alternative that there are more, arises by minimizing (15) over $\mathbf{A}$, with $\mathbf{A}$ being $R \times N$ and of rank r. Nyblom and Harvey (2000) show that this leads to $\zeta(K, N)$.

The limiting distribution of the common trends test statistic under the null hypothesis depends on functionals of Brownian motion and, although the test is derived for Gaussian models, this result only requires that the disturbances be IID. The series expansion for the limiting distribution is

$$\zeta(K,N) \xrightarrow{d} \sum_{k=1}^{\infty} (\pi k)^{-2} \mathbf{u}_k' \mathbf{u}_k - \operatorname{tr} \left(\sum_{k=1}^{\infty} (\pi k)^{-3} \mathbf{u}_k \boldsymbol{v}_k' \right) \left(\sum_{k=1}^{\infty} (\pi k)^{-4} \boldsymbol{v}_k \boldsymbol{v}_k' \right)^{-1} \left(\sum_{k=1}^{\infty} (\pi k)^{-3} \boldsymbol{v}_k \mathbf{u}_k' \right) \tag{16}$$

where $\boldsymbol{v}_k$ and $\mathbf{u}_k$ are, respectively, $K \times 1$ and $r \times 1$ vectors which are mutually independent $\mathrm{NID}(\mathbf{0}, \mathbf{I})$. It is perhaps surprising that such a limiting distribution is obtained in view of the fact that $\mathbf{S}$ is not a consistent estimator of $\boldsymbol{\Sigma}_\varepsilon$ when $K > 0$. The 10%, 5% and 1% significance points of the limiting distribution of $\zeta(K, N)$ depend on both K and N and are shown in Table 8.5.2a. The values for $K = 0$ correspond to those given in table???

.

Table 8.5.2a Critical values for the common trends test, $\zeta(K, N)$

N	Significance level	K: 0	1	2	3
	10%	.607	.162		
2	5%	.748	.218		
	1%	1.074	.383		
	10%	.841	.297	.094	
3	5%	1.000	.382	.120	
	1%	1.359	.614	.200	
	10%	1.063	.427	.170	.063
4	5%	1.237	.538	.208	.078
	1%	1.623	.830	.317	.121

Table 8.5.2b Critical values for the common trends test with time trends, $\zeta_2(K, N)$

N	Significance level	K 0	1	2	3
	10%	.211	.085		
2	5%	.247	.105		
	1%	.329	.160		
	10%	.296	.151	.061	
3	5%	.332	.180	.075	
	1%	.428	.245	.113	
	10%	.377	.215	.110	.046
4	5%	.423	.246	.128	.055
	1%	.521	.321	.176	.081

Under the alternative hypothesis, $T^{-1}\zeta(K, N)$ has a limiting distribution which depends only on $K^{\dagger} - K$. Hence the test based on $\zeta(K, N)$ is consistent. Specifically, if the rank of Σ_η is $K^{\dagger} > K$, the limiting distribution of $T^{-1}\zeta_{K,N}$ is the same as the limiting distribution of $T^{-1}(\ell_{K+1} + ... + \ell_{K^{\dagger}})$.

When time trends are fitted, as in (8.4.16??), the critical values for the test statistic, denoted $\zeta_2(K, N)$, are as in Table 8.5.2b. The test can make no allowance for the fact that there may be common slopes corresponding to the common levels. With known co-integrating vectors this possibility can be accommodated quite naturally since it implies that $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$, and so $\mathbf{Ay}_t$ includes no time trend.

A nonparametric adjustment for serial correlation may be made in the common trends test in much the same way as was suggested for the ξ_N test. All that needs to be done is to replace the eigenvalues in (14) by those of $\mathbf{S}(\ell)^{-1}\mathbf{C}$, where $\mathbf{S}(\ell)$ is formed as in (8.4.17), leading to test statistics which may be denoted as $\zeta(K, N; \ell)$ or, when there is a time trend, $\zeta_2(K, N; \ell)$. Under the null hypothesis, the asymptotic results stated above continue to hold.

Following the discussion in sub-sections 5.4 and 8.4.2, a parametric test may be thought more attractive, but one may not, at least initially, wish to fit a multivariate model. A reasonable compromise is the use the innovations from univariate models.

Busetti (1999) extends the test to deal with structural breaks at known points, thereby generalising the test described in sub-section 7.6.2??

The papers by Harris (1997) and Snell (1998) propose tests of a similar hypothesis to the one addressed by the common trends test. These tests are different in that they they are based on estimating co-integrating relationships by principal components. In the common trends test no explicit estimators are needed.

[Probably drop ??]The common trends test is testing the null hypothesis that there are K common trends against the alternative that there are more. Equivalently, it is testing the null hypothesis that there are $N-K$ co-integrating vectors against the alternative that there are less. Thus it is similar to the tests described at the end of sub-section ???. However, when $K>1$, the common trends test is testing a slightly different hypothesis to the one maintained there because the latter are based on knowing that all the common trends can go in $\mathbf{y}_{1t}$, that is the series in $\mathbf{y}_{1t}$ are not co-integrated. In terms of (4) and (5) this means that $\mathbf{\Sigma}_{\eta}^{\dagger}$ is positive definite. Thus suppose that $N=3$ and that y_1 and y_2 depend on a single common trend driven by a disturbance which is uncorrelated with the disturbance of the trend in y_3. Then $K=2$ and $r=1$, but regressing y_3 on y_1 and y_2 will not provide a test of the null hypothesis that $r=1$ agianst the alternative that $r=0$. (Another way of seeing this problem is that the dependent variable is not in the co-integrating relationship)????.

8.5.6 Local linear trend

The multivariate local linear trend model is

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_{\mathbf{t}}, \quad \boldsymbol{\varepsilon}_{\mathbf{t}} \sim NID(\mathbf{0}, \mathbf{\Sigma}_{\epsilon}), \tag{17}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim NID(\mathbf{0}, \mathbf{\Sigma}_{\eta}),$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t, \quad \boldsymbol{\zeta}_t \sim NID(\mathbf{0}, \mathbf{\Sigma}_{\zeta}),$$

Common factors occur because $\mathbf{\Sigma}_{\eta}$ and/or $\mathbf{\Sigma}_{\zeta}$ is less than full rank. Since this leads to a wide range of possibilities and the analysis can become quite complex, it is as well to start off looking at a bivariate model. It is useful to write the covariance matrices of $\boldsymbol{\eta}_t$ as in (8.5.2) and similarly for $\boldsymbol{\zeta}_t$ so

$$\mathbf{\Sigma}_{\eta} = \begin{bmatrix} \sigma_{1\eta}^2 & \rho_{\eta}\sigma_{1\eta}\sigma_{2\eta} \\ \rho_{\eta}\sigma_{1\eta}\sigma_{2\eta} & \sigma_{2\eta}^2 \end{bmatrix} \qquad and \qquad \mathbf{\Sigma}_{\zeta} = \begin{bmatrix} \sigma_{1\zeta}^2 & \rho_{\zeta}\sigma_{1\zeta}\sigma_{2\zeta} \\ \rho_{\zeta}\sigma_{1\zeta}\sigma_{2\zeta} & \sigma_{2\zeta}^2 \end{bmatrix}, \tag{18}$$

where ρ_{η} and ρ_{ζ} are correlations.

When $\rho_{\zeta} = \pm 1$ there is only one source of stochastic movement in the two slopes. This is the *common slopes* model. We can write

$$\beta_{2t} = \bar{\beta} + \theta\beta_{1t}, \quad t = 1, ..., T \tag{19}$$

where $\theta = sgn(\rho_{\zeta})\sigma_{2\zeta}/\sigma_{1\zeta}$ and $\bar{\beta}$ is a constant. When $\bar{\beta}=0$, the model has *proportional slopes.* If, furthermore, θ is equal to one, that is $\sigma_{2\zeta}=\sigma_{1\zeta}$ and ρ_{ζ} positive, we have *identical slopes.* The restriction that $\bar{\beta}$ is zero is imposed simply by excluding β_{2t} from the state vector and replacing it by $\theta\beta_{1t}$ in the transition equation for μ_{2t}.

The general definition of co-integration, given by Granger and Engle (1987), applies to a vector $\mathbf{y}_t$, the elements of which are all integrated of order d. Then if there exists a non-null vector ,$\boldsymbol{\alpha}$, such that $\boldsymbol{\alpha}'\mathbf{y}_t$ is $I(d-b)$ with $b > 0$, the series are said to be *co-integrated of order* (d, b), denoted $\mathbf{y}_t\tilde{\ } CI(d,b)$. The series in a common slopes model are therefore co-integrated of order (2,1), that is CI(2,1). Although both y_{1t} and y_{2t} require second differencing to make them stationary, there is a linear combination of first differences which is stationary. If there is also perfect correlation between the level disturbances, that is $\rho_\eta = \pm 1$, and, furthermore, $\sigma_{2\eta}/\sigma_{1\eta} = \sigma_{2\zeta}/\sigma_{1\zeta}$, then the series are CI(2,2), meaning that there is a linear combination of the observations themselves which is stationary. These conditions mean that $\mathbf{\Sigma}_\zeta$ is proportional to $\mathbf{\Sigma}_\eta$. This being the case, the model is said to display *trend homogeneity.*

Smooth trends with common slopes- First consider the case where the covariance matrix of the slope disturbances is of rank K_β, but the covariance matrix of level disturbances is null so that the the trends are integrated random walks which, when estimated, are relatively smooth. The model may be written in an analogous way to the local level model in (6), that is

$$\begin{aligned} \mathbf{y}_{1t} &= \boldsymbol{\mu}_t^{\dagger\dagger} + \boldsymbol{\varepsilon}_{1t}, \quad \boldsymbol{\varepsilon}_\mathbf{t} \sim NID(\mathbf{0}, \mathbf{\Sigma}_\epsilon), \\ \mathbf{y}_{2t} &= \mathbf{\Pi}\boldsymbol{\mu}_t^{\dagger\dagger} + \overline{\boldsymbol{\mu}} + \overline{\boldsymbol{\beta}}t + \boldsymbol{\varepsilon}_{2t} \end{aligned} \tag{20}$$

$$\boldsymbol{\mu}_t^{\dagger\dagger} = \boldsymbol{\mu}_{t-1}^{\dagger\dagger} + \boldsymbol{\beta}_{t-1}^{\dagger}, \tag{21}$$

$$\boldsymbol{\beta}_t^{\dagger} = \boldsymbol{\beta}_{t-1}^{\dagger} + \boldsymbol{\zeta}_t^{\dagger}, \quad \boldsymbol{\zeta}_t^{\dagger} \sim NID(\mathbf{0}, \mathbf{\Sigma}_\zeta^{\dagger}),$$

such that the $K_\beta \times K_\beta$ matrix $\mathbf{\Sigma}_\zeta^{\dagger}$ is positive definite and $\mathbf{y}_{1t}$ and $\boldsymbol{\mu}_t^{\dagger\dagger}$ are $K_\beta \times 1$ vectors. The deterministic components $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\beta}}$ are vectors of length $N - K_\beta$. From the computational point of view it is probably best to include these in a state vector as in ???????

If $K_\beta = 1$ and the series have the same underlying change, $\mathbf{\Pi}$ is equal to a vector of ones and $\overline{\boldsymbol{\beta}} = \mathbf{0}$. This is a model of balanced growth, as in sub-section 8.5.3, and the trend components in the forecast functions are parallel.

GNP and investment - A bivariate model for US GNP and investment was originally estimated with level disturbance terms. However, its covariance matrix was relatively small and was found to be statistically insignificant when LR tests were carried out. The zero level variances result in quite smooth trends, as shown in Figure 8.5.4??. The variance of the irregular component for GNP was estimated as zero. The slope disturbances are perfectly correlated. Figure 8.5.5?? shows the cycles in GNP and investment on the same graph.

Wage distributions in the US Figure ???shows the deciles in the distribution of real US wages from ??to??. Harvey and Bernstein (2000) fit the smooth trend model in order to better display the stylised facts. A common trend would indicate that the underlying distribution is not changing over time. This is another example of balanced growth. The null hypothesis of no change over time could be tested by the multivariate test statistic as outlined in sub-section 8.5.4. (As argued in section 5.5, the random walk test is as effective as the test against a smooth trend).

Common levels and slopes - When the level disturbance covariance matrix is of rank K and the slope covariance matrix is of rank K_β, one way of writing the model is

$$\mathbf{y}_t = \mathbf{\Pi}^{\dagger}\boldsymbol{\mu}_t{}^{\dagger} + \mathbf{\Pi}^{\dagger\dagger}\boldsymbol{\mu}_t^{\dagger\dagger} + \overline{\boldsymbol{\mu}}_N + \overline{\boldsymbol{\beta}}_N t + \boldsymbol{\varepsilon}_\mathbf{t}, \quad \boldsymbol{\varepsilon}_\mathbf{t} \sim NID(\mathbf{0}, \mathbf{\Sigma}_\varepsilon), \tag{22}$$

where $\boldsymbol{\mu}_t^{\dagger}$ is a $K \times 1$ vector of random walks as in (6), $\boldsymbol{\mu}_t^{\dagger\dagger}$ is a $K_\beta \times 1$ vector of integrated random walks as in (20), $\mathbf{\Pi}^{\dagger}$ contains K rows (not necessarily the first K) which together make up an identity matrix, $\mathbf{\Pi}^{\dagger\dagger}$similarly contains K_β rows (not necessarily the same ones) which make up an identity and $\overline{\boldsymbol{\mu}}_N$ and $\overline{\boldsymbol{\beta}}_N$ contain the appropriate number of constants and slopes. The smooth trend model and the random walk plus drift are special cases. If $K = N$, the model can be written as

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \begin{bmatrix} I \\ \Pi \end{bmatrix} \boldsymbol{\mu}_t^{\dagger\dagger} + \begin{bmatrix} \mathbf{0} \\ \overline{\boldsymbol{\beta}} t \end{bmatrix} + \boldsymbol{\varepsilon}_\mathbf{t}$$

or a common slope can be incorporated into the trend equation.

Partially homogeneous common trend model - If $K = K_\beta$ and $\mathbf{\Pi}^{\dagger}$ and $\mathbf{\Pi}^{\dagger\dagger}$, the slope can be incorporated in the level, as in the standard formulation of a local linear trend (17), to give a single set of K stochastic trends. If, in addition, $\mathbf{\Sigma}_\zeta^{\dagger} = q_\zeta \mathbf{\Sigma}_\eta^{\dagger}$, where q_ζ is a signal-to-noise ratio, the K trends can be written with uncorrelated standardized disturbances for both level and slope, thereby generalizing (8) and giving

$$\mathbf{y}_t = \mathbf{\Theta}^{*}\boldsymbol{\mu}_t^{*} + \overline{\boldsymbol{\mu}}_N + \overline{\boldsymbol{\beta}}_N t + \boldsymbol{\varepsilon}_\mathbf{t}, \quad \boldsymbol{\varepsilon}_\mathbf{t} \sim NID(\mathbf{0}, \mathbf{\Sigma}_\epsilon), \tag{23}$$

$$
\begin{aligned}
\boldsymbol{\mu}_t^* &= \boldsymbol{\mu}_{t-1}^* + \boldsymbol{\beta}_{t-1}^* + \boldsymbol{\eta}_t^*, \quad \boldsymbol{\eta}_t^* \sim NID(\mathbf{0}, \mathbf{I}_K), \\
\boldsymbol{\beta}_t^* &= \boldsymbol{\beta}_{t-1}^* + \boldsymbol{\zeta}_t^*, \quad \boldsymbol{\zeta}_t^* \sim NID(\mathbf{0}, q_\zeta \mathbf{I}_\zeta)
\end{aligned}
$$

where $\boldsymbol{\Theta}^* = \boldsymbol{\Pi}^\dagger(\boldsymbol{\Sigma}_\eta^\dagger)^{-1/2}$. The common trends are fully independent in both level and slope for all possible rotations. Since the restrictions mean that $\boldsymbol{\Sigma}_\zeta = q_\zeta \boldsymbol{\Sigma}_\eta$ in (20) , the model is *trend homogeneous*.

Co-integration- The co-integrating interpretation continues to hold if the trends contain stochastic slopes so that the series are I(2). Thus, for example, with smooth trends the system is co-integrated of order (2,2), that is CI(2,2), since there is a combination of the series which is stationary. The triangular representation is as before but with the inclusion of a set of time trends in the co-integrating equations. More generally, with common levels and slopes it should be apparent from (22) that there are two sets of co-integrating vectors, one to annihilate the stochastic slopes, that is with the property $\mathbf{A}\boldsymbol{\Pi}^{\dagger\dagger} = \mathbf{0}$, the other to remove the stochastic levels, so that $\mathbf{A}\boldsymbol{\Pi}^\dagger = \mathbf{0}$. If both are applied the result is a mixture of the two. If the model is trend homogeneous, it has the nice property that the two sets of co-integrating vectors are identical.

[sect 8.8????] *Money demand* Economic theory on money demand suggests a co-integrating relationship between the logarithm of real money balances, the logarithm of real income and the nominal interest rate; that is $N = 3$, $r = 1$. Thus $\boldsymbol{\Pi}$ is a (1×2) row vector containing the coefficients of income and the interest rate, and the two stochastic trends can be associated with each of these variables. Some judgement as to how reasonable the estimates are can be made from economic considerations; at the minimum theory suggests what the signs should be.

If income follows a smooth trend rather than a random walk, the system may be written in the form (22), that is

$$
\begin{pmatrix} y_t \\ r_t \\ m_t \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ \pi_r \end{pmatrix} \mu_t^\dagger + \begin{pmatrix} 1 \\ 0 \\ \pi_y \end{pmatrix} \mu_t^{\dagger\dagger} + \begin{pmatrix} 0 \\ 0 \\ \overline{\mu} + \overline{\beta} t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}
$$

The co-integrating row vector $(-\pi_y, 0, 1)$ removes the stochastic slope, giving a relationship between money and income which has an $I(1)$ stochastic part; hence these variables are $CI(2,1)$. The vector $(0, -\pi_r, 1)$ removes the stochastic level, giving a money demand equation with a stationary disturbance term (plus a deterministic constant and slope). If the trend in income also has a disturbance term in the level, the relationships between the factor loadings and the coefficients in the money demand equation are slightly more complicated.

8.5.7 Model specification

The most difficult aspect of model specification lies in determining the number of common factors. The common trends test of sub-section 8.5.5 offers a way of checking the appropriateness of a specified number of common levels. Furthermore the fact that the distribution of the η_N test depends on the rank of Σ_η suggests that an indication of the rank of Σ_η can be obtained by looking at the eigenvalues of $\mathbf{S}(\ell)^{-1}\mathbf{C}$ to see how many are relatively large. This has the advantage that no models actually need to be estimated. However, it can only be used when the series are assumed to be generated by a multivariate random walk plus a stationary process. For more general models it seems that the only way to proceed is to estimate a general model in which the relevant covariance matrices are unrestricted and then to make judgements about their ranks by examining their eigenvalues. Indeed this might be the best approach even when looking at the eigenvalues of $\mathbf{S}(\ell)^{-1}\mathbf{C}$ is feasible.

In a bivariate model it is not necessary to compute eigenvalues of an estimated covariance matrix; all one has to do is to look at the estimate of the correlation, ρ, between the two series. Perfect correlation implies a single common factor.

Stochastic volatility In the stochastic volatility example referred to in the sub-section on factor rotation, unrestricted estimation gave an estimated Σ_η matrix in which the first eigenvalue accounted for 89% of the total variance, while the second eigenvalue accounted for a further 6.4%. The $\zeta_{2,4}$ common trends statistic for the observations gives a value of 0.17. This is less than the 5% critical value in Table 8.5.2 and its prob. value is 10%. Thus the assumption of two common trends is not rejected. On the other hand, the hypothesis of one common trend is rejected since $\zeta_{1,4} = 1.21$ and the 5% critical value is 0.54.

In principal components analysis, a covariance matrix is decomposed as $\mathbf{EDE}'$, where $\mathbf{E}$ is a matrix of eigenvectors and $\mathbf{D}$ is a diagonal matrix of eigenvalues. With such a decomposition for Σ_η the local level model can be written

$$\mathbf{y}_t = \mathbf{\Theta}^{PC}\boldsymbol{\mu}_t^{PC} + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{\Theta}^{PC} = \mathbf{ED}^{1/2}$ and $\boldsymbol{\mu}_t^{PC} = \mathbf{D}^{-1/2}\mathbf{E}\boldsymbol{\mu}_t$. The principal component trends, $\boldsymbol{\mu}_t^{PC}$, are generated by a disturbance, $\boldsymbol{\eta}_t^{PC}$, with a covariance matrix equal to the identity matrix. The loadings, $\mathbf{\Theta}^{PC}$, can give an indication of possible interpretations of (rotated) common trends; see the discussion in Harvey *et al (*1994). Thus principal components analysis may give useful additional information when the eigenvalues of an estimated covariance matrix are used to make a judgement about the number of common factors.

Once a common factor model has been estimated it may be compared with competing models with different numbers of common factors using AIC or BIC criteria. Likelihood ratio tests will have nonstandard distributions.

Given that the number of common factors has been set and a model has been estimated, inference is straightforward insofar as standard distribution theory applies. Thus in the context of the local level model,(), LR and Wald tests may be carried out on parameters in $\mathbf{\Pi}$ and $\overline{\boldsymbol{\mu}}$ and on off-diagonal elements of Σ_η (provided that the null hypothesis does not imply a reduction in rank). The chi-square distribution may be used to conduct inference in the usual way. Although all inference only has asymptotic validity in general, t and F distributions may also be used. An illustration can be found in the seat belt example of the next section.

8.5.8 Vector autoregressions and co-integration

Stock and Watson (1988) show how common trends, that is random walks, can be extracted from a VAR fitted to $I(1)$ variables. These estimated trends can be regarded as a multivariate generalisation of the Beveridge-Nelson trend for a univariate series. As was pointed out in sub-section 6.1.2, the Beveridge-Nelson trend is computed using a one-sided filter and so it will tend to be more erratic than a trend extracted using a smoother. This is evident from an inspection of the common trends presented for certain US macroeconomic series in King *et al (*1991).

Co-integrating restrictions are normally imposed on a VAR model in two ways. The first is via the *vector error correction mechanism* (VECM) as in Johansen (1988). The second, adopted by Pesaran and Shin (1998), is based on a triangular representation. The second approach ties in more closely with the multivariate structural time series models, where, as we saw in (12), there is also an implied triangular representation. Obtaining the VECM representation from a common trends model is not easy except in special cases. However, it can be computed numerically, using the algorithm of Koopman and Harvey (1999) described in sub-section 4.7.4. An example is given below.

The VECM can be written as

$$\Delta\mathbf{y}_t = \mathbf{\Phi}^*\mathbf{y}_{t-1}+\boldsymbol{\delta} + \sum_{j=1}^{\infty}\mathbf{\Phi}_j^*\Delta\mathbf{y}_{t-j}+\mathbf{v}_t, \tag{24}$$

where the relationship between the $N \times N$ parameter matrices and those in the VAR model of (??) is

$$\Phi^* = \sum_{k=1}^{\infty} \Phi_k - \mathbf{I}, \qquad \Phi_j^* = -\sum_{k=j+1}^{\infty} \Phi_k, \qquad j = 1, 2, \ldots. \tag{25}$$

The rank of Φ^* is R. It is normally expressed as $\Phi^* = \Gamma A$, where A is the $r \times N$ matrix of co-integrating vectors and Γ is a $N \times r$ matix of coefficients. When there is no co-integration, $\Phi^* = 0$.

A homogeneous local level model, (8.2.??) with $\Sigma_\eta = q\Sigma_\varepsilon$, has no co-integration, so $\Phi^* = 0$, and the model is just a VAR in first differences. The parameter matrices are $\Phi_j^* = (-\theta)^j I$, $j = 1, 2, \ldots$, where θ is obtained from (**??**). If q is small, θ is close to minus one and there is a slow decline in the $\Phi_j^* = \$'$s.

(Cristiano, a ecuação última parece estar errada, confere-a)

Given a common trends model, (6), the weighting algorithm computes the coefficient matrices in the corresponding VAR. The VECM matrices, Φ^* and the $\Phi_j^{*\prime} s$ are then obtained from (25). The matrix $\mathbf{A}$ is not unique, but it can be set to $(-\Pi,\mathbf{I})$. The Γ matrix is then computed so as to satisfy $\Phi^* = \Gamma A$, while $\delta = -\Phi^*(0, \bar{\mu}')'$. If Σ_η is dominated by Σ_ε, a slow decline in the Φ_j^*'s can be expected. This has implications for using a VECM to model a system for which a common trends model is appropriate.

Example In the bivariate local level model (6) with $\Sigma_\varepsilon = I_2$ and $\Sigma_\eta = 0.1[2, 1][2, 1]'$. The implied co-integrating vector is equal to $[-.5, 1]$ and we find

$$\Phi^* = \begin{bmatrix} -.2 & .4 \\ .4 & -.8 \end{bmatrix} = \begin{bmatrix} .4 \\ -.8 \end{bmatrix} \begin{bmatrix} -.5 & 1 \end{bmatrix}.$$

Figure ?? in KH shows the weights attached to $\Delta y_{1,t-j}$ and $\Delta y_{2,t-j}$, for $j = 1, 2, \ldots$, decline exponentially as in the simple homogeneous case analysed earlier.?? The *Ox* program `vecm.ox` is used to generate these results ??

For a single series, the VECM model reduces to the AR formulation used in the ADF test. The null hypothesis in the ADF test is that the series contains a unit root, while under the alternative it is stationary. For multivariate series, the hypothesis that there are no co-integrating vectors means that there are N unit roots, so the model can be estimated in first differences, while with N co-integrating vectors the model is stationary. The LR tests proposed in Johansen (1988) are made up of a sequence in which the null hypothesis of $r-1$ co-integrating vectors is tested against the alternative of r co-integrating vectors for $r = 1, ..., N$. This means that the tests are in the direction of fewer unit roots or, what is more to the point, fewer common trends.

The common trends test in Stock and Watson (1988) is based on the the unrestricted estimation of a VAR(1) model with allowance made for further serial correlation using a nonparametric approach similar to that of Phillips and Perron (1988). The sequence of tests, like those in Johansen, is in the direction of fewer common trends.

The common trends test of sub-section 8.5.5 differs from the Johansen and Stock-Watson tests just as the NM and KPSS tests differ from augmented Dickey-Fuller. Thus the alternatives for the $\zeta_{K,N}$ test are in the direction of *more* common trends rather than less. Furthermore no models are being estimated, either under the null or the alternative, and we are not testing against the alternative that there should specifically be one more common trend (or one less co-integrating vector).

8.5.9 Common cycles, common seasonals and explanatory variables

As noted in sub-section 8.2.4, common cycles are a special case of similar cycles. A model with common trends and common cycles could be written

$$\mathbf{y}_t = \mathbf{\Theta}_{\mu}\boldsymbol{\mu}_t + \boldsymbol{\mu}_{\theta} + \mathbf{\Theta}_{\psi}\boldsymbol{\psi}_t + \boldsymbol{\epsilon}_t, \tag{26}$$

where $\mathbf{\Theta}_{\psi}$ is $N \times K_{\psi}$ with K_{ψ} being the number of common cycles. There is no vector of constant terms corresponding to $\boldsymbol{\mu}_{\theta}$ for the level since the expectation of a cycle is zero. Thus with two series and one common cycle, one cycle is proportional to the other. The model is a special case of the bivariate similar cycles model in which $\mathbf{\Sigma}_{\kappa}$ is of rank one. Common cycles are like the common feature cycles of Engle and Kozicki (1993).

Income and unemployment Clark (1989) estimated a bivariate model in which a common cycle, specified as an AR(2), appeared in the GNP and unemployment for the US.

In the case of trigonometric seasonals there are two sets of $N \times 1$ vectors for each seasonal frequency such that

$$\begin{array}{rclcl} E(\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}') & = & E(\boldsymbol{\omega}_{it}^{*}\boldsymbol{\omega}_{it}^{*'}) & = & \mathbf{\Sigma}_{\omega}, \\ E(\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}^{*'}) & = & \mathbf{0}, & & i = 1, 2, \ldots, [s/2], \end{array} \tag{27}$$

and all disturbances at different frequencies are independent.

Common factors in seasonality implies a reduction in the number of disturbances driving changes in the seasonal patterns. For trigonometric seasonals it is possible, in principle, to have different disturbance covariance matrices for each of the seasonal frequencies, thereby allowing common factors in some frequencies but not in others. This implies seasonal co-integration at different frequencies; see Hylleberg *et.al.* (1990). With the same covariance matrix, $\mathbf{\Sigma}_\omega$, for each frequency only full seasonal co-integration is possible.

Seasonal co-integration in a bivariate model does not imply that the seasonal patterns are the same unless the deterministic seasonal components, which play an analogous role to the elements of $\overline{\boldsymbol{\mu}}$ in (6), are equal to zero. What it does mean is that the source of any evolving seasonal behaviour in the two series is the same, so that a suitable linear combination of them will have a deterministic seasonal pattern.

8.5.10 Application to income and consumption

There a number of theories about the behaviour of aggregate consumption. for example, Keynesians stress the 'consumption function' in which consumption depends on income, while the REPIH predicts that consumption should follow a random walk A bivariate structural time series model provides a description of the way in which income and consumption move together and enables forecasts to be made. It does not attempt to test the various theories formally, but it can yield useful insights.

An unrestricted model with trend, cycle, seasonal and irregular components was fitted to UK reaL GDP using quarterly data over the period ???.

Now consider the trend. The slopes have a correlation of 0.987, and so it is reasonable to assume that they are co-integrated. We may go further and restrict the growth rates to be the same. This is the identical slopes model in which $\overline{\beta}$ and θ in (8.5.19??) are set to zero and one respectively.

The levels have a correlation of 0.891. A plot of the levels shows they drift apart somewhat, but this feature is also apparent in the original data. Plot of Exp () show the APC. could enforce co-integration ?

The cycle repesents the short-term movements. The period is just over eight years. The correlation is 0.955 and the plot in joint components shows how the cycles move together. The expected value of the cycle in consumption given the cycle in income, that is $\mathrm{E}(\psi_{ct} \mid \psi_{yt}) = 0.557$. Thus the model shows that the APC is around ...., but changing slowly, probably because consumption does not include durables and there may have been a shift in the balance between durables and non-durables over the period in question. See plot of the difference in the trend components. The MPC, on the other hand, is 0.557. These stylised facts are quite consistent with Keynesian theory, but the model makes no judgement as to the correctness of the theory or even the direction of causality. Indeed the trend component could be interpreted as permanent income, with the cycle in consumption being due to credit constrained consumers. (slope in cons - see Scott).

The seasonal disturbances are virtually uncorrelated, with a correlation of -0.193. On the other hand a plot of the seasonals shows that they move faily closely together and have a correlation of 0.774. The interpretation is that while there is a moderate correlation between the seasonal patterns in the two series, which might possibly be explained by seasonal effects in income resulting in corresponding seasonal effects in consumption, the *changes* in the seasonal patterns in the two series are unrelated. This is why a single equation model for the consumption function requires a stochastic seasonal component; see Harvey and Scott (1994).

Noe estimate another model , but this time replace the cycle by a VAR(1) component. The fit is similar, with the VAR estimated as:

???

The VAR indicates an interaction between y and c, the net result of which is to give a cycle with a similar period to that obtained before ?? A plot of the VAR components in the two series is similar to the plot obtained with the cycle model, with the cycles moving very closely together. Thus the VAR is picking up the main dynamic features of the series , but in a different way. It would be very dangerous to interpret the coefficients of the VAR as having some kind of behavioural meaning, and indeed their values are not such as to make such an interpretation particularly appealing.

Finally fit a VECM to the model. Series I(1). Seasonally adjust.

Exercise

In the local level model with drifts, derive the estimator for β in (7) by first writing down the spectral likelihood for $\mathbf{\Delta y}_t - \mathbf{\Pi}^{\dagger}\boldsymbol{\beta}$.