Day 5.

David Pupovac

So far…

- ***In last two classes we covered:***
  - AR, MA, ARIMA
  - Seasonality and SARIMA
  - ARMAX
- ***Now, we continue with independent variables***

Vector autoregression - VAR

- Univariate time series enable forecasting on the basis of its **past values.** Optionally, we could have included **past and present values of independent <u>exogenous</u>** variables.

- However, in many cases we are interested in the interaction of several **endogenous** time series. In this context, by endogeneity we assume that a time series are being influenced from within a system. We are assuming a loop of causality between the independent and dependent variables of a model.

- We saw yesterday that we can reformulate the structural form, e.g.:

$$y_{1,t} = \alpha_1 y_{2,t} + \varphi_{11}\, y_{1,t-1} + \varphi_{12} y_{2,t-1} + \varepsilon_{1,t}$$
$$y_{2,t} = \alpha_2 y_{1,t} + \varphi_{21} y_{1,t-1} + \varphi_{22} y_{2,t-1} + \varepsilon_{2,t}$$

into the reduced form:

$$y_{1,t} = \Pi_{11} y_{1,t-1} + \Pi_{12} y_{2,t-1} + \varepsilon_{1,t}$$
$$y_{2,t} = \Pi_{21} y_{1,t-1} + \Pi_{22} y_{2,t-1} + \varepsilon_{2,t}$$

- This is the basis of vector autoregression models

Structure of VAR

- ***VAR models generalize the univariate autoregressive model (AR model) by allowing for more than one evolving variable.***

- A VAR model describes the evolution of a set of k endogenous variables over the same sample period ($t = 1, \ldots, T$) as a linear function of their past values.

- Each of the n endogenous time series $y_{i,t}$ depends on its own lagged values and lagged values of all other endogenous time series. Thus, VAR captures information about the interactions of jointly endogenous variables

- However, as a consequence of the dependence on the lagged values of independent variables, $y_{i,t}$ is influenced not only by the lagged values of its random disturbances, $w_{i,t}$ but also by the lagged values of all the other $w_{i,t}$

- Exogenous variables can be included in the model and as well as their lagged values.

Basic structure

- A p order vector autoregressive process generalizes a one variable AR(p) process to n variables:
- If Y_t is a vector of n jointly endogenous variables:

$$\boldsymbol{Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + w_t}$$

$$Y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ \vdots \\ y_{n,t} \end{bmatrix} \quad \Phi_0 = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \mu_p \end{bmatrix} \quad \Phi_i = \begin{bmatrix} \varphi_{i,11} & \cdots & \varphi_{i,1n} \\ \vdots & \ddots & \vdots \\ \varphi_{i,n1} & \cdots & \varphi_{i,nn} \end{bmatrix} \quad w_t = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ \vdots \\ w_p \end{bmatrix}$$

where w_t is multivariate white noise and can be contemporaneously correlated

- Equation is a vector expression for an AR(p) process, hence the name: **vector autoregressive model**

Multivariate white noise

- Two series $w_{x,t}$, $w_{y,t}$ are **bivariate white noise**, if they are stationary and their cross-covariance $\gamma_{xy}(k) = Cov(w_{x,t}, w_{y,t+k})$:

$$\gamma_{xx}(k) = \gamma_{yy}(k) = \gamma_{xy}(k) = 0 \text{ for all } k \neq 0$$

- Therefore, series may be cross-correlated at lag 0 but they cannot be correlated at lags $k \neq 0$

- Stationary series $w_{1,t}$, $w_{2,t}$ ..., $w_{n,t}$ are **multivariate white noise** if each individual series is white noise and, for each pair of series ($i \neq j$), $\gamma_{ik}(k) = 0$ for all $k \neq 0$.

Types of VAR

- Three types of VAR will be discussed: 1) reduced form VARs 2) recursive VARs 3) VECM

- VARs are typically used for macroeconomic forecasting or structural (policy) analysis

  – Reduced form VAR are sufficient for forecasting. They place few restrictions on estimated relationship among the elements of Y_t (essentially only p order of VAR and linearity of relationship)

  – Recursive VAR specify priority of elements of Y_t. The equations are ordered so that the error term is uncorrelated with preceding equations

  – VECM enable us to address cointegrating relations between variables

Reduced form VAR

- Reduced form VAR are typically used for forecasting

- A simple (the simplest possible) bivariate VAR with one autoregressive term would be:

$$y_{1,t} = \varphi_{11} y_{1,t-1} + \varphi_{12} y_{2,t-1} + w_{1,t}$$
$$y_{2,t} = \varphi_{21} y_{1,t-1} + \varphi_{22} y_{2,t-1} + w_{2,t}$$

- Each equation can be estimated with OLS

Identification of VAR

- ***Number of lags*** - the tentative nature of identification is still present. To identify appropriate lag length, we estimate multiple VAR of varying lag lengths and compute the information criteria such as those of Akaike [1981], Hannan and Quinn [1979], Quinn [1980], or Schwarz [1978], or the final prediction error.
  - These will likely indicate different number of lags. (According to the Lütkepohl (2005) BIC and HQ provide consistent estimates of the true lag order)
  - Some of the relations suggested by Pfaff(2008) are:

$$\hat{p}(SC) \leq \hat{p}(AIC) \;\; if \; t \geq 8$$
$$\hat{p}(SC) \leq \hat{p}(HQ) \;\; for \; all \; t$$
$$\hat{p}(HQ) \leq \hat{p}(AIC) \;\; if \; t \geq 16$$

- Furthermore, you can use likelihood ratio test
- Naturally, check the significance of lags

Stationarity/stability

- ***Stationarity*** - the stationarity of $\boldsymbol{Y_t}$ is determined from $\boldsymbol{\Phi}$, not the individual elements $\boldsymbol{\Phi}$ – does this mean that individual processes may be non-stationary??? - this point is a bit contentious)

- ***Stability*** - a stable VAR process is stationary, which means that VAR generates time series with time-invariant means, variances, and covariance structure. As with ARIMA, the concept of stability is based on representing VAR as infinite MA process

Stationarity and characteristic equation

- Testing of stationarity can be based on evaluation of characteristic equation

- For instance imagine following VAR(1) process

$$y_{1,t} = \varphi_{11} y_{1,t-1} + \varphi_{12} y_{2,t-1} + w_{1,t}$$
$$y_{2,t} = \varphi_{21} y_{2,t-1} + \varphi_{22} y_{1,t-1} + w_{2,t}$$

$$\Phi = \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{bmatrix}$$

- For a VAR(1), the characteristic equation is given by the determinant of the matrix.

$$\boldsymbol{w_t = \Phi(L) Y_t = (I - \Phi L) Y_t}$$

- VAR(p) model is stationary if the roots of the determinant $|\varphi(L)|$ all exceed unity in absolute value. For the example above, the determinant is given by:

$$\begin{bmatrix} 1 - \varphi_{11} L & -\varphi_{12} L \\ -\varphi_{21} L & 1 - \varphi_{22} L \end{bmatrix} = (1 - \varphi_{11} L)(1 - \varphi_{22} L) - \varphi_{12}\, \varphi_{21} L^2$$

Stationarity and companion matrix

- In practice, the stability of an empirical VAR(p)-process is analyzed by means of the **companion matrix**
- Namely, the *p*th order VAR can always be rewritten as a first order VAR
- The matrix of lag coefficients in the first order representation is called companion matrix

Companion matrix

- We begin the transformation to a first-order VAR by subtracting the vector of constants, μ, from Y_t

$$\tilde{Y}_t = \begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix}$$

- Then we build **companion matrix**

$$\widetilde{\Phi} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix}$$

- And stacking the vector of random disturbances

$$\tilde{\varepsilon}_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$$

- And we get the first-order VAR

$$\tilde{Y}_t = \widetilde{\Phi}\tilde{Y}_{t-1} + \tilde{\varepsilon}_t$$

Characteristic equation vs. companion matrix

- For stationarity all eigenvalues of companion matrix $\tilde{\boldsymbol{\Phi}}$ must lie within unit circle. Thus, if the moduli of the eigenvalues of matrix $\tilde{\boldsymbol{\Phi}}$ are less than one, then the VAR(p)-process is stable.

- However, notice the difference:

  1. VAR is stationary if roots of the determinant $\boldsymbol{\Phi}(\boldsymbol{L})$ lie <u>outside unit cycle</u>

  2. Equivalently, VAR is stationary if all the eigenvalues of the companion matrix $\tilde{\boldsymbol{\Phi}}$ lie <u>inside unit circle</u>

Postestimation tests

- Finally, you have to conduct a set of diagnostics test
- While it is fine if you also check individual residuals, the assumption of VAR is multivariate white noise. The tests are:
- **For serial correlation**
  - a Portmanteau test and the Lagrange multiplier Breusch and Godfrey test are most commonly applied.
- **For heteroscedasticity**
  - multivariate ARCH-Lagrange multiplier test
- **For normally distributed residuals**
  - multivariate Jarque-Bera test, skewness, kurtosis
- **For structural stability**
  - CUSUM (or some variation of the plot). CUSUM chart is a time-weighted chart that displays the cumulative sums of deviations of sample value from the target value If a trend develops upward or downward, it indicates that the process mean has shifted

Recursive VAR

- We typically want to understand the causal order of the factors.

- The ordering of events requires strong identifying assumptions. However, recursive VAR enables us to establish order of events without making some of identifying assumptions of other statistical models

- There are $n!$ orderings for n variables. However, the data may be supportive of multiple orderings.

Cross correlations

- The simplest way to establish temporal order is to evaluate cross-correlations present between different leads and lags of the variable x and the current value of the variable y

$$\gamma_{yx}(k) = (y_t - \bar{y})(x_{t+k} - \bar{x})$$

- Correlation function:

$$\rho_{xy}(k) = \frac{\gamma_{yx}\ (k)}{\sigma_y \sigma_x}$$

Granger causality

- Granger causality test fall somewhere between the reduced and recursive approaches to VARs. It makes statements about the order of events but do not require identifying assumptions

- **Granger test checks a VAR for evidence of temporal ordering by testing wheatear lagged values of one variable ($y_{i,t}$) improve the forecasts of another variable ($y_{j,t}$) after its lagged values are taken into account.**

- Cross-correlation functions measure simple correlations, while Granger causality test reports on the incremental predictivness after the influence of prior values of other variables are taken into account.

Granger causality test

- A VAR model

$$y_{1,t} = \varphi_{11} y_{1,t-1} + \varphi_{12} y_{2,t-1} + w_{1,t}$$
$$y_{2,t} = \varphi_{21} y_{2,t-1} + \varphi_{22} y_{1,t-1} + w_{2,t}$$

- written in matrix form is defined as:

$$\boldsymbol{\Phi(L) Y_t = w_t}$$

- Therefore, ignoring vector of constants, a system of two variables, can be partitioned in to two parts:

$$\begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

- The statement that y_{2t} does not Granger-cause y_{1t} is equivalent to the statement $\Phi_{12}(L) = 0$:

$$\begin{bmatrix} \Phi_{11}(L) & 0 \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

- Thus, the null hypothesis of y_{2t} does not Granger-cause y_{1t} is:

$$H_0 \colon \varphi_{12,t-1} = \varphi_{12,t-2} = \cdots = \varphi_{12,t-p} = 0$$

Impulse response function

- Causality tests falls short of quantifying the impact of the impulse variable on the response variable over time. The impulse response analysis is used to investigate these kinds of dynamic interactions

- We have already mentioned the IRFs in discussing dynamic characteristics of AR and MA. In this context we would take, for instance an AR process:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + w_t$$

and write it as a infinite MA:

$$y_t = \nu + \sum_{i=0}^{\infty} \psi_i w_{t-i}$$

- We set all w_t to 0 except w_{t-k}, which is set to 1. The ψ_i trace the impact over time of this random impulse. Sequence of ψ_i is the IRF. Thus, IRFs quantify the persistence of disturbances.

Impulse response analysis and VAR

- This idea is applied on VARs.

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + w_t$$

Assuming stability we can rewrite the model in moving average form:

$$Y_t = \nu + \sum_{i=0}^{\infty} \Psi_i w_{t-i}$$

- The IRF gives the *j*th-period response when the system is shocked by a one-standard-deviation shock.
- **Unlike with ARIMA, in VAR we have a system of equations, and we can trace response of each variable to a shock to each equations**
- The effects can be cumulated through time $t = 1, 2, \ldots,$ and hence one would obtain the cumulated impact of a unit change

Orthogonalized impulse response function

- OIRF is alternative to IRF. This is the case if the underlying shocks are less likely to occur in isolation but rather contemporaneous correlation between the components of the error process w_t exists; i.e., the off-diagonal elements of correlation matrix $\boldsymbol{\Sigma}$ are non-zero.

  (remember that w_t is multivariate white noise, which may be may be cross-correlated at lag 0)

- This would imply that we cannot impose causal interpretation on the impulses.

- The point is in decomposing the correlation matrix $\boldsymbol{\Sigma}$ in product of a lower triangular matrix and its transpose so that there is no need to consider correlation with other elements.

- However, due to this transformation OIRF presents definite break between reduced and recursive form VAR, because we need to impose the recursive order

OIRF

- The orthogonal impulse responses are derived from a Choleski decomposition

$$\boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}'$$

- Because the matrix $\mathbf{P}$ is lower triangular it follows that only a shock in the first variable of a VAR(p)-process exerts an influence on all the remaining ones and that the second and following variables cannot have a direct impact

$$\mathbf{P}^{-1}\boldsymbol{\Sigma}\mathbf{P}'^{-1} = \mathbf{I}_n$$

- We can use $\boldsymbol{P}^{-1}$ to convert w_t to a vector of uncorrelated random disturbances. So we can rewrite our moving average representation as:

$$\boldsymbol{y_t} = \boldsymbol{\gamma} + \sum_{\boldsymbol{i=0}}^{\infty} \boldsymbol{\Psi_i}\mathbf{P}\mathbf{P}^{-1}\boldsymbol{w_{t-i}} = \boldsymbol{\gamma} + \sum_{\boldsymbol{i=0}}^{\infty} \boldsymbol{\Xi_i}\boldsymbol{\nu_{t-i}}$$

$$\boldsymbol{\Xi_i} = \boldsymbol{\Psi_i}\mathbf{P}$$

$$\boldsymbol{\nu_t} = \mathbf{P}^{-1}\boldsymbol{w_{t-i}}$$

- By construction the elements of $\boldsymbol{\nu_t}$ are mutually orthogonal $\boldsymbol{\nu_t}\boldsymbol{\nu_t'} = \mathbf{I}$, so we can make causal statements about shocks to one element without considering correlations with other elements. The $\boldsymbol{\Xi_i}$ are called OIRF.

- Additionally, you can use cumulative OIRF.

FEVD - forecast-error variance decompositions

- The forecast error variance decomposition (FEVD) is based upon the orthogonal impulse response
- It enables us to measure the fraction of the total forecast-error variance that is attributable to each orthogonalized shock.
- However, keep in mind that ordering is also important here.

Models of nonstationary time series

Limits of models so far:

- The stationary univariate models provide useful information predominantly for only a few periods ahead. By definition, disturbances in stationary time series dissipate quickly.
- The influence of present time disturbances can persist a bit longer in VARs.
- Forecasting beyond a short horizon depends primarily on trends. So far trend was treated as nuisance, a component to eliminate prior to modeling.
- We can model the trends, nevertheless, modeling trends presents challenges.

Spurious regression

- For time series variables we have to be particularly careful before ascribing any causal relationship since an apparent relationship could exist due to underlying trend

- If only deterministic trends are present in the data-generating processes of the variables in question then these can be removed before estimation or can be included in the regression. The inference on the coefficients is the same regardless of which method is employed.

- However, matters are different in the case of difference-stationary data. If y_t and x_t are **independent** unit-root processes (they are **not cointegrated** and $\beta = 0$), the classical least square will almost always reject true hypothesis of $\beta = 0$ – this is **spurious regression.**

- **Granger and Newbold (1974) provided a test for spurious regression - if your R^2 is larger than your Durbin Watson statistics you have a spurious regression**

Differencing and cointegration

- The spurious regression problem can be circumvented by taking first differences of the I(1)-variables and using these instead. However, we can not use differencing with cointegrated series.

- Why not employ differencing with cointegrated series?

1. Differencing greatly attenuates large positive residual autocorrelation; hence, false inferences upon the coefficients in the regression equation could be drawn and the equation is misspecified

2. Differencing creates great obstacle in the testing the long-run relationships between variables

Concept of cointegration

The central intuition of cointegration:

- ***The nonstationary variables may obey a long-run relationship with each other whose residual is stationary***

- A conintergrating relationship is a stationary linear combination of two or more non-stationary variables.

- Assume regression $y_t = \alpha + \beta x_t + \nu_t$. If y_t is nonstationary either x_t or ν_t, or both must be nonstationary. **If $\boldsymbol{\nu_t}$ is stationary, we call $\boldsymbol{y_t - \alpha - \beta x_t}$ the cointegrating relationship between $\boldsymbol{y_t}$ and $\boldsymbol{x_t}$.**

- There can be at most $k - 1$ cointegrating relationships (where k is number of variables)

Error correction model

- Take the equation (with constant μ in the equation):

$$y_t = \mu + \alpha y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + \varepsilon_t$$

- After differencing y_t and simultaneously adding and subtracting $\gamma_0 x_{t-1}$ we get:

$$\Delta \boldsymbol{y_t} = \mu' - (\boldsymbol{1-\alpha})\boldsymbol{y_{t-1}} + \boldsymbol{\gamma_0}\Delta \boldsymbol{x_t} + \gamma_0\, x_{t-1} + \gamma_1\, x_{t-1} + \varepsilon_t$$
$$\Delta y_t = \mu' - (1-\alpha) y_{t-1} + \gamma_0 \Delta x_t + (\gamma_0 + \gamma_1) x_{t-1} + \varepsilon_t$$

- If $\lambda \equiv (1-\alpha)$ and $\beta = \frac{\gamma_0+\gamma_1}{1-\alpha}$ we get the following formula:

$$\Delta y_t = \mu' + \gamma_0 \Delta x_t - \lambda(y_{t-1} - \alpha - \beta x_{t-1}) + \varepsilon_t$$

This equation is error-correction model (ECM)

- If $\boldsymbol{\nu_t = y_t - \alpha - \beta x_t}$, the basic form of error correction model is defined as:

$$\Delta y_t = \gamma_0\, \Delta x_t + \boldsymbol{\lambda \nu_{t-1}} + \varepsilon_t$$

Error correction model

- The model:

$$\Delta y_t = \gamma_0 \Delta x_t + \boldsymbol{\lambda \nu_{t-1}} + \varepsilon_t$$

is generalizable to models with additional y_t and x_t lags:

$$\Delta y_t = \sum_{\boldsymbol{i=0}}^{\boldsymbol{p}} \boldsymbol{\gamma_i \Delta x_{t-i}} + \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \boldsymbol{\psi_j \Delta y_{t-j}} + \lambda \nu_{t-1} + \varepsilon_t$$

notice the indexes below summation operators

  - The cointegrating relationship $\boldsymbol{\lambda \nu_{t-1}}$ describes the **long term relationship that links the levels of two (or more) nonstationary variables**. In the parlance of economics, deviations from a long-run equilibrium path are possible, but these errors are characterized by a mean reversion to its stable long-run equilibrium.

  - Lagged differences $\sum_{i=0}^{p} \gamma_i \Delta x_{t-i} + \sum_{j=1}^{k} \psi_j \Delta x_{t-j}$ represent the **short run dynamics**.

Cointegration tests

- **In case of cointegration if we difference variables we misspecify equation**
- **If we do not difference we run the risk of estimating spurious equation.**

- So, first we need to establish if there is cointegration and determine how many cointegrating relations are there

- **Phillips and Ouliaris** [1990] introduced two residual-based tests, namely a variance ratio and a multivariate trace statistic for pairs of variables. The latter has the advantage that it is invariant to which variable is on the left hand side of the equation.

Error-correction model estimation - step 1

- With two exogenous variables is easy. Engle and Granger [1987] proposed a two-step estimation technique.

- To estimate parameters of long run cointegrating relationship $y_{t-1} - \alpha - \beta x_{t-1}$ you can just use OLS:

$$y_t = \alpha + \beta x_t + \nu_t$$

- **If this series is stationary the variables are cointegrated, and this can be tested with the Dickey-Fuller (DF) type of test.**

- **Consequently, we can also use unit root tests as tests of cointegration.**

Error-correction model estimation - step 2

- The estimation of λ is based on ν_t. Based on our assumed ECR we estimate:

$$\Delta y_t = \sum_{i=0}^{p} \gamma_i \Delta x_{t-i} + \sum_{j=1}^{k} \psi_j \Delta y_{t-j} + \lambda \nu_{t-1} + \varepsilon_t$$

- ν_{t-1} is the "error" in the system and equations describe how system adjusts or corrects back to the equilibrium

- Theoretically, λ should range between -1 and 0, which you can interpret as a percentage return to equilibrium or to what percentage disequilibrium corrected in one period.

  - Larger values, tending to 0, indicate that adjustment is slow,
  - Extremely small values, less than -2, indicate an overshooting of economic equilibrium(≤ -2)
  - Positive values would imply that the system diverges from the long-run equilibrium path (this is not applicable to VECM)

Vector error-correction model - VECM

- Simple ECM is restricted to only a single equation with one variable designated as the dependent variable, explained by one or more variables. It also relies on pretesting the time series to find out whether variables are $I(0)$ or $I(1)$.

- However things get more complicated when there are more than two variables. One consequence is that we need to impose some restrictions to obtain identification.

- So, if we write a vector autoregressive model for K element vector $\boldsymbol{Y_t}$ where all $\boldsymbol{Y_t}$ have unit roots, and $\varepsilon_t \sim N(0, \Sigma)$:

  $\boldsymbol{Y_t = \mu + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \varepsilon_t}$ and we subtract $\boldsymbol{Y_{t-1}}$ we get:

  $\boldsymbol{\Delta Y_t = \mu + (\Phi_1 + \Phi_2 - I) Y_{t-1} - \Phi_2 \Delta Y_{t-1} + \varepsilon_t}$

Vector error-correction model

- From:

$$\Delta \boldsymbol{Y_t} = \boldsymbol{\mu} + (\boldsymbol{\Phi_1} + \boldsymbol{\Phi_2} - \boldsymbol{I})\boldsymbol{Y_{t-1}} - \boldsymbol{\Phi_2}\,\Delta \boldsymbol{Y_{t-1}} + \boldsymbol{\varepsilon_t}$$

the general VECM model can be defined in the following way:

$$\boldsymbol{\Pi} = \sum_{j=1}^{p} \boldsymbol{\Phi_j} - \boldsymbol{I}$$

$$\boldsymbol{\Gamma_i} = -\sum_{j=i+1}^{p} \boldsymbol{\Phi_j}$$

$$\Delta \boldsymbol{Y_t} = \boldsymbol{\mu} + \boldsymbol{\Pi Y_{t-1}} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma_i}\,\boldsymbol{\Delta Y_{t-1}} + \boldsymbol{\varepsilon_t}$$

Here $\boldsymbol{\Pi}$ is the error-correction term, and it is the focus of our interest

(It should be noted that VECM can have two representations, long-run and transitory, which are equivalent. The upper representation is transitory.)

Rank of Π and cointegration

- If the rank of matrix Π is defined as *rk()*, there are three cases:
  - $rk(\Pi) = K$; in this case all K linearly independent combinations are stationary around deterministic components
  - $rk(\Pi) = 0$; in this case no linear combination exists to make Π - the VECM reduces to a VAR in first differences
  - $0 < rk(\Pi) = r < K$
- The third case it most important in this context

Rank of $\boldsymbol{\Pi}$ and cointegration

- Because the matrix does not have full rank, two $(K \times r)$ matrices α and β^T exist such that $\boldsymbol{\Pi = \alpha\beta^T}$. Thus we can write:

$$\Delta Y_t = \mu + \alpha\beta^T Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \varepsilon_t$$

  - The elements of α determine the speed of adjustment to the long-run equilibrium, and α referred to as the loading or adjustment matrix
  - The r linear independent columns of β are the cointegrating vectors

- Without further restrictions, the parameters of the matrices α and β are undefined because any non-singular matrix $\boldsymbol{\Xi}$ would yield $\boldsymbol{\alpha\Xi(\beta\Xi^{-1})' = \Pi}$. To obtain unique values of α and β^T requires further restrictions on the model.. The solution is to normalize one element of β to one. For instance for bivariate case:

$$\beta^T = (1 \quad -\beta)$$

VECM - cointegrating rank

- Phillips and Ouliaris test is only applicable to a single cointegrating relation.

- Johansen [1995] developed maximum-likelihood estimators of set of cointegration vectors based on canonical correlation to determine the extent to which the multicollinearity in the data allows for a smaller r-dimensional space. This statistic has been named the **trace statistic**.

- Besides the trace statistic, Johansen and Juselius [1990] have suggested the maximal **eigenvalue statistics**. Namely, the rank of the matrix is equal to the number of its eigenvalues (characteristic roots) that are different from zero. Hence, an eigenvalue significantly different from zero indicates a significant cointegrating vector

- Once the cointegration rank r (i.e. number of linearly independent cointegrated vectors) has been determined, the cointegrating vectors can be estimated.

VECM – lag length

- For the VECM model, select the lag using:
  - VAR with variables at level (differenced variables),
  - statistical criterion of AIC,
  - no evidence of autocorrelation or heteroscedasticity in the residuals of VAR

Identifying models

- For interpretations, it is often convenient to normalize or identify the cointegrating vectors by choosing a specific coordinate system in which to express the variables.

- Johansen proposed a default identification scheme that has become the conventional method of identifying models in the absence of theoretically justified restrictions.

- Johansen [1995] proposed restricting β' in the form where the first part is an identity matrix. In practice, the estimation of the parameters of a VECM requires at least r^2 identification restrictions

Diagnostics and forecasting

- Diagnostics test follow the procedure analogous to the VAR. Thus we want to assure that the resulting residual is white noise.
- Technically you could forecast directly from the VECM
- However, it is computationally more convenient to estimate a VECM, transforming it into a VAR model and forecasting using the VAR model
- VECM and the corresponding VAR model are two equivalent representations of the same model. The model outcomes do not change due to algebraic manipulation of the equations.