Vector Autoregressions

Dr. Nils Herger

Study Center Gerzensee

Overview of the Program

1. Introduction and Definition
2. Estimation and Econometric Issues
    1. Estimation
    2. Determining the Lag Length
    3. Stationarity and Cointegration
3. Identification
    1. Short-Term Restrictions
    2. Cholesky Decomposition
4. Monetary Policy Analysis with VARs
    1. Forecasting
    2. Impulse Response Function
5. Cointegration and Vector Error Correction Models (VECM)*
6. Summary and Conclusion

Introduction

- Policy instruments of central banks (base interest rate, money supply, exchange rate) affect and are re-affected by the macroeconomy. The direction of causality between macroeconomic variables is unclear.
- Monetary policy does not lend itself to experiments with randomisation. Monetary policy analysis relies on non-experimental methods.

⇒ Against this background, Sims (1980) proposed the usage of **vector autoregression (VAR)** models to trace the multivariate behaviour of dynamically interrelated time-series. Though lacking the theoretical rigor of structural models, VARs can be a more powerful tool to investigate or forecast economic variables.

Definition: Vector Autoregression (VAR)

A vector autoregression (VAR) is a multivariate time series model representing the dynamic evolution of an array of variables from their common history. Thereby, a VAR entertains the idea that virtually all variables are endogenous.

Example 1: Bivariate VAR(1)

Consider a simple model analysing the dynamic interaction between a non-policy variable v and a policy instrument x. With one lag, the VAR (structural form) is given by:

$$x_t + \phi_{12}^t v_t = \delta_1^t + \phi_{11}^{t-1} x_{t-1} + \phi_{12}^{t-1} v_{t-1} + \epsilon_1^t$$
$$\phi_{21}^t x_t + v_t = \delta_2^t + \phi_{21}^{t-1} x_{t-1} + \phi_{22}^{t-1} v_{t-1} + \epsilon_2^t$$

or in matrix form

$$\mathbf{\Phi_0 y_t} = \mathbf{d_t} + \mathbf{\Phi_1 y_{t-1}} + \boldsymbol{\varepsilon_t}$$

with

$$\mathbf{\Phi_0} = \begin{pmatrix} 1 & \phi_{12}^t \\ \phi_{21}^t & 1 \end{pmatrix} \mathbf{\Phi_1} = \begin{pmatrix} \phi_{11}^{t-1} & \phi_{12}^{t-1} \\ \phi_{21}^{t-1} & \phi_{22}^{t-1} \end{pmatrix},$$

$$\mathbf{y_t} = \begin{pmatrix} x_t \\ v_t \end{pmatrix}, \mathbf{y_{t-1}} = \begin{pmatrix} x_{t-1} \\ v_{t-1} \end{pmatrix}, \mathbf{d_t} = \begin{pmatrix} \delta_1^t \\ \delta_2^t \end{pmatrix}, \boldsymbol{\varepsilon_t} = \begin{pmatrix} \epsilon_1^t \\ \epsilon_2^t \end{pmatrix}.$$

Definition: Structural (or Primitive) Form of a VAR

$$\mathbf{\Phi_0 y_t} = \mathbf{d_t} + \sum_{l=1}^{p} \mathbf{\Phi_l y_{t-l}} + \boldsymbol{\varepsilon}_t$$

where

$$\mathbf{y_{t-l}} = \begin{pmatrix} y_1^{t-l} \\ \vdots \\ y_n^{t-l} \end{pmatrix}, \quad \mathbf{\Phi_l} = \begin{pmatrix} \phi_{11}^{t-l} & \cdots & \phi_{1n}^{t-l} \\ \vdots & \ddots & \vdots \\ \phi_{n1}^{t-l} & \cdots & \phi_{nn}^{t-l} \end{pmatrix}$$

for lags $l = 0, 1, \ldots, p$, and

$$\mathbf{d_t} = \begin{pmatrix} \delta_1^t \\ \vdots \\ \delta_n^t \end{pmatrix}, \qquad \boldsymbol{\varepsilon_t} = \begin{pmatrix} \epsilon_1^t \\ \vdots \\ \epsilon_n^t \end{pmatrix}.$$

- The vector $\mathbf{y_t}$ contains an array of time-series (variables) whose values depend on the past. Usually, $\mathbf{y_t}$ includes a **policy instrument** and several, possibly conflicting, **policy goals**. The choice of variables depends on the question under investigation, the policy regime, and the availability and quality of the data. Keep the model parsimonious!
- The vector $\boldsymbol{\varepsilon_t}$ is an **innovation process**. The part of $\boldsymbol{\varepsilon_t}$ pertaining to policy instruments can be interpreted as unanticipated "policy shock" (such as interventions by the central bank) that is thought to affect the current and future values of the other variables of $\mathbf{y_t}$.[1]
- The vector $\mathbf{d_t}$ collects the constants (intercepts) that make sure that innovations have zero expectations ($E[\varepsilon_\mathbf{t}] = 0$).

[1] Innovations should not be confused with residuals, which are statistical artefact from estimation (see below).

- The structural parameters connecting the time series in $\mathbf{y_t}$ appear in matrices $\mathbf{\Phi_l}$. An element ϕ_{ij}^{t-l} of $\mathbf{\Phi_l}$ reports the marginal impact of variable j observed l periods ago, on the current state of variable i.
- The instantaneous impacts appear in $\mathbf{\Phi_0}$ whose diagonal equals 1 ($\phi_{ii}^{t} = 1$). This is merely a normalisation reflecting that the current value of an element in $\mathbf{y_t}$ maps one-to-one on itself. Non-zero off-diagonal elements $\phi_{ij}^{t} \neq 0$ indicate that the time series of $\mathbf{y_t}$ are contemporaneously correlated. Such feedbacks give rise to an **endogeneity problem** creating statistical pitfalls when we want to unfold causal effects from the data.

- Similar to the univariate case, when $(\mathbf{\Phi_0} - \mathbf{d_t} - \sum_{l=1}^{p} \mathbf{\Phi_l y_{t-l}})$ is **invertible**, every VAR can be written as **vector moving-average** (VMA), linking $\mathbf{y_t}$ with a sequence of past innovations ε_t.

Vector Moving Average (VMA) Model

$$\mathbf{y_t} = \mathbf{\Psi}_0 \varepsilon_t + \mathbf{\Psi}_1 \varepsilon_{t-1} + \mathbf{\Psi}_2 \varepsilon_{t-2} + \dots$$

- The transformation from a VAR to a VMA rests on the following stability conditions:[2]
  1. The AR part must be of finite order ($\sum_{l=1}^{p} \mathbf{\Phi_l y_{t-l}} < \infty$).
  2. The time series entering the VAR must be **stationary**.
- Under these assumptions, the VMA is **convergent** in the sense that $\sum_{i=0}^{\infty} |\mathbf{\Psi_i}| < \infty$. This implies that shocks from e.g. a policy intervention have no permanent effect, but die away.
- The choice between the VAR and VMA representation is a matter of convenience. In particular, the VAR and VMA representation is preferred when, respectively, it comes to forecasting and to unfolding the reactions of endogenous variables to policy interventions.

[2] Textbooks such as Lütkepohl (2007) provide more detail on this.

- The structural form of a VAR does not lend itself to statistical analysis. Thereto, each equation should only contain one dependent variable that is related to a set of (here lagged) independent variables.
- However, pre-multiplication with $\mathbf{\Phi_0}^{-1}$ yields such equations that are termed **reduced form** of a VAR.

The Reduced Form of a VAR

$$\mathbf{y_t} = \mathbf{B_0} + \sum_{l=1}^{p} \mathbf{B_l} \mathbf{y_{t-l}} + \mathbf{e_t}$$

↷ The reduced form coefficients are connected with the structural parameters through the "endogeneity matrix" $\mathbf{\Phi_0}$.

Reduced Form Coefficients and Structural Form Parameters

$$\mathbf{B_0} = \mathbf{\Phi_0^{-1}} \mathbf{d_t} \quad \text{and} \quad \mathbf{B_l} = \mathbf{\Phi_0^{-1}} \mathbf{\Phi_l} \quad \text{for} \quad l = 1, \ldots, p.$$

- The connection between structural innovations and reduced form disturbances is given by:

Innovations and Residuals

$$\mathbf{A}\mathbf{e_t} = \boldsymbol{\Phi}_\mathbf{0}^{-1}\boldsymbol{\varepsilon}_t$$

- Innovations $\boldsymbol{\varepsilon}_t$ are usually thought to be unrelated, one-unit shocks. Their variance-covariance matrix is often the identity matrix $\boldsymbol{\Sigma}_\varepsilon = \mathbf{I}$.
- Rather than a (policy) innovation, $\mathbf{e_t}$ are residuals balancing each regression. Unless $\boldsymbol{\Phi}_\mathbf{0}$ is a diagonal matrix, the reduced form equations are interrelated through the structural form residuals meaning that policy shocks transmit across equations.
- For estimation purposes, $\mathbf{e_t}$ is typically assumed to follow a multivariate **white noise process**, that is $\mathbf{e_t} \sim WN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a Gaussian distributed $n \times n$ variance-covariance matrix.

Example 1:(*continued*) Solving the equation system for current values of the policy variable x_t and the outcome v_t yields the reduced form:

$$\begin{aligned}
x_t &= \beta_{10}^t - \beta_{11}^{t-1} x_{t-1} + \beta_{12}^{t-1} v_{t-1} + e_1^t \\
&= \frac{\delta_1^t}{1-\phi_{12}^t\phi_{21}^t} - \frac{\phi_{12}^t\delta_2^t}{1-\phi_{12}^t\phi_{21}^t} + \frac{\phi_{11}^{t-1} - \phi_{12}^t\phi_{21}^{t-1}}{1-\phi_{12}^t\phi_{21}^t} x_{t-1} \\
&\quad + \frac{\phi_{12}^{t-1} - \phi_{22}^t\phi_{12}^{t-1}}{1-\phi_{12}^t\phi_{21}^t} v_{t-1} + \frac{\epsilon_1^t}{1-\phi_{12}^t\phi_{21}^t} - \frac{\phi_{12}^t\epsilon_2^t}{1-\phi_{12}^t\phi_{21}^t} \\
v_t &= \beta_{20}^t - \beta_{21}^{t-1} x_{t-1} + \beta_{22}^{t-1} v_{t-1} + e_2^t \\
&= -\frac{\phi_{21}^t\delta_1^t}{1-\phi_{12}^t\phi_{21}^t} + \frac{\delta_2^t}{1-\phi_{12}^t\phi_{21}^t} + \frac{\phi_{21}^{t-1} - \phi_{21}^t\phi_{11}^{t-1}}{1-\phi_{12}^t\phi_{21}^t} x_{t-1} \\
&\quad + \frac{\phi_{22}^{t-1} - \phi_{21}^t\phi_{12}^{t-1}}{1-\phi_{12}^t\phi_{21}^t} v_{t-1} - \frac{\phi_{21}^t\epsilon_1^t}{1-\phi_{12}^t\phi_{21}^t} + \frac{\epsilon_2^t}{1-\phi_{12}^t\phi_{21}^t}
\end{aligned}$$

Using matrices saves, again, a great deal of notation:

$$\mathbf{y_t} = \mathbf{B_0} + \mathbf{B_1 y_{t-1}} + \mathbf{e_t}$$

Estimation and Econometric Issues

1. Estimation

- The coefficients of a VAR can be estimated equation by equation with ordinary least squares (OLS). Furthermore, from the residuals, the reduced form variance-covariance matrix $\boldsymbol{\Sigma}$ can be estimated and, in turn, statistical tests (t-tests, F-tests, etc.) can be applied.
- The results of the VAR should be checked with the usual diagnostic statistics. In particular, make sure that the residuals are *WN*, which would provide evidence that the VAR is correctly specified. At this stage, you will often have to proceed in an **iterative** way.
- VARs are quite demanding as regards the amount of data. When only short time-series are available, it is not sensible to use a VAR. In this case, univariate time-series techniques can be more appropriate.

2. Determining the Lag Length

- The lag length p can be determined according to the Akaike (AIC), the Schwarz-Bayesian (BIC), or other information criteria.
- Specifically, the so-called **general to specific approach** starts with a large number of lags and estimates the VAR with consecutively fewer lags until an information criterion reaches a minimum.

↷ Information criteria can produce conflicting results. In particular, the AIC is arguably statistically consistent but the BIC has better small sample properties in finding the correct lag length.

3. Stationary and Cointegration

- For standard coefficient tests to be meaningful and to mitigate against spurious results, variables should be stationary. Check this with a unit-root test (ADF-test) before introducing a variable into a VAR.

↷ Many macroeconomic variables grow over time or wander about and are, thus, not stationary (e.g. Nelson and Plosser, 1982).

- Two ways exist to deal with non-stationary data.

1. In many cases, simple transformations such as de-trending for trend-stationary or taking difference for integrated variables produces stationarity. Most macroeconomic variables are integrated of order one (e.g. I(1)) that is they are stationary in growth rates but not in levels.
2. Even if a set of time series $\mathbf{y_t}$ is non-stationary, it is possible that a combination between them is stationary. Then, they are said to be **cointegrated**, which reflects a scenario where the mean or variance of individual variables is time-variant, but in the long term they are tied together by some common trend. This case warrants the usage of cointegration models like vector error-correction models (see below).

Identification

Identification

Identification refers to the reconstruction of the deep parameters $\mathbf{\Phi}$ of the structural form from the estimated coefficients $\widehat{\mathbf{B}}$ and variance-covariance matrix $\widehat{\mathbf{\Sigma}}$ of the reduced form of a VAR.

- Though the reduced form coefficients $\mathbf{B}$ can be estimated by OLS, to uncover the parameters from the reduced form coefficients, that is $\mathbf{\Phi_l} = \mathbf{\Phi_0 B_l}$, the matrix $\mathbf{\Phi_0}$ must be known.
- Reduced-form residuals offer additional information to pin down the variance-covariance matrix $\widehat{\mathbf{\Sigma}}$. Recall that $\mathbf{Ae_t} = \mathbf{\Phi_0^{-1}} \varepsilon_t$. Hence, the structural and reduced form variance-covariance matrices are connected as follows:

Variance-Covariance Matrix

$$\mathbf{A}\widehat{\mathbf{\Sigma}}\mathbf{A}' = \mathbf{\Phi_0^{-1}} \mathbf{\Sigma}_\varepsilon \mathbf{\Phi_0^{-1}}'$$

- Nevertheless, an **identification problem** arises in the sense that the information of the reduced form does not permit to uniquely infer the underlying structure of a VAR. To see this, note that the parameters to be estimated include $n \times n \times p$ coefficients of $\mathbf{\Phi_l}$ and n constants of $\mathbf{d_t}$. For this, an equal number of estimates appear in, respectively, $\mathbf{B_l}$ and $\mathbf{B_0}$. Conversely, the symmetric matrix $\widehat{\mathbf{\Sigma}}$ offers only $n(n+1)/2$ entries to gauge the structure of $\mathbf{\Phi_0^{-1}\Sigma_\varepsilon\Phi_0^{-1\prime}}$ with n^2 elements of $\mathbf{\Phi_0}$ and n diagonal elements of $\mathbf{\Sigma_\varepsilon}$, or $n^2 + n = n(n+1)$ in total.

$\Rightarrow$ In sum, the reduced form VAR is **underidentified** and requires

$$\underbrace{(nnp - nnp)}_{Coefficients} + \underbrace{(n-n)}_{Contstants} + \underbrace{[n(n+1) - n(n+1)/2]}_{Variance-Covariance} = n(n+1)/2$$

additional **identifying restrictions** to extract a structure of $\mathbf{\Phi}$ from the reduced-form estimates $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{\Sigma}}$.

- Identifying restrictions add information (structure) to a VAR. Hence, within this context, the term **structural VAR** (or SVAR) is used.[3]

[3] Not to be confused with the structural form of a VAR.

- Different structural forms can correspond with the same reduced form VAR equations. Furthermore, alternative restrictions lead often to different results. Therefore, it is crucial to support the choice of identifying restrictions with economic theory or prevailing policy conditions!

Example 1:(*continued*) The structural form of the bivariate VAR(1) has 12 parameters ($\delta_1^t, \delta_2^t, \phi_{12}^t, \phi_{21}^t, \phi_{11}^{t-1}, \phi_{12}^{t-1}, \phi_{21}^{t-1}\phi_{22}^{t-1}, \sigma_1^2, \sigma_2^2, \sigma_{12}, \sigma_{21}$). The reduced form has 9 estimates ($b_{01}^t, b_{02}^t, b_{11}^{t-1}, b_{12}^{t-1}, b_{21}^{t-1}, b_{22}^{t-1}, \widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\sigma}_{12}$). Hence, 3 identifying restrictions are warranted (since n=2). We have already imposed that $\sigma_{12} = \sigma_{21} = 0$ reflecting that innovations are though to be uncorrelated shocks.

- Several sources of a priori information can be exploited to identify the deep parameters of a VAR. This leads to different **schemes to reach identification** including (1.) contemporaneous restrictions and (2.) the Cholesky decomposition.[4]

[4] Other identification schemes are long-term restrictions (Blanchard and Quah, 1989), sign restrictions (e.g. Uhlig, 2005), or identification through heteroscedasticity (e.g. Rigobon and Sack, 2004).

1. Contemporaneous (or Short-Term) Restrictions

- Contemporaneous restrictions assume that current changes of y_i^t do not immediately affect other endogenous variables in $\mathbf{y_t}$.

Example 1:(*continued*)

In the bivariate policy VAR(1), one identifying restriction was warranted. Assuming $\phi_{12}^t = 0$, that is policy does not react to contemporaneous changes of the target variable (possibly because such information is unavailable), or $\phi_{21}^t = 0$, that is policy does not have an immediate effect on the target variable, are examples for contemporaneous restrictions.

- ↷ Whether contemporaneous restrictions are reasonable depends often on the frequency of the data. Assuming away contemporaneous feedbacks is more reasonable with monthly than, say, yearly data.
- ↷ Monetary policy regimes give rise to different contemporaneous restrictions (Bernanke and Mihov, 1995). E.g. inflation targeting or the adoption of a fixed exchange rate embodies certain restrictions with respect to macroeconomic variables.

2. Cholesky Decomposition

- According to the Cholesky decomposition, every positive-definite matrix such as $\widehat{\mathbf{\Sigma}}$ can be decomposed into two lower triangular matrices $\mathbf{P}$, that is

$$\widehat{\mathbf{\Sigma}} = \mathbf{P}\mathbf{P}'.$$

- Recall that $\mathbf{P}\mathbf{P}' = \mathbf{\Phi}_0^{-1}\mathbf{\Sigma}_\varepsilon\mathbf{\Phi}_0^{-1'}$. Normalising $\mathbf{\Sigma}_\varepsilon = \mathbf{I}$ and setting the elements above the principal diagonal in $\mathbf{\Phi}_0$ to zero, that is

$$\mathbf{\Phi}_0^{-1} = \mathbf{S} = \begin{pmatrix} s_{11}^t & 0 & \dots & 0 \\ s_{21}^t & s_{22}^t & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ s_{n1}^t & s_{n2}^t & \dots & s_{nn}^t \end{pmatrix} \overset{\text{endogeneity}}{\downarrow} \quad \mathbf{\Sigma}_\varepsilon = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0\dots & \dots & 1 \end{pmatrix}$$

yields precisely the $n(n+1)/2$ required restrictions to reach identification. Identification by the Cholesky decomposition imposes a recursive order where the variables are ranked according to the degree of endogeneity. Therefore, this approach is also called **recursive VAR**.

Example 2: Monetary Policy VAR (Stock and Watson, 2001)

Consider a monetary policy VAR with inflation π, unemployment u, and the money market interest rate r, which is the monetary policy instrument. This yields a three equations VAR with structural and reduced form

$$\mathbf{\Phi_0 y_t} = \mathbf{d_t} + \sum_{l=1}^{p} \mathbf{\Phi_l y_{t-l}} + \varepsilon_t$$

$$\mathbf{y_t} = \mathbf{B_0} + \sum_{l=1}^{p} \mathbf{B_l y_{t-l}} + \mathbf{e_t}$$

where $\mathbf{y_t}' = [\pi, u, r]$ and $\varepsilon_t' = [\epsilon^\pi, \epsilon^u, \epsilon^r]$. Imposing e.g. the recursive ordering that $\pi_t \rightarrow u_t \rightarrow r_t$ means that

i. Inflation is contemporaneously only affected by current expectations,
ii. Unemployment is only affected by its natural rate and an inflation shock, and
iii. Interest rates are contemporaneously affected by shocks of variables.

Monetary Policy Analysis with VARs

- The results of a VAR can be used in many ways including:[5]
  1. **Structural analysis** between policy instruments and goals. Probably the most important tool is the **impulse-response function** that traces out the dynamic reaction of policy goals to a policy intervention.[6]
  2. **Forecasting** the future development of policy goals such as inflation.
- The quality of an analysis depends on the proper choice of variables, estimation, and identifying restrictions. Recall that large VARs are not necessarily superior to parsimonious specifications.
- The strength of VARs lies in forecasting and structural analysis for the short-term (e.g. a time horizon of several month/quarters).

[5] VARs can also be used to run Granger causality tests to gauge e.g. the direction of impact between policy instruments and goals. However, the lack of a distinction between structural and reduced form introduce caveats to the Granger causality (Geweke, 1984).

[6] Other tools are the forecast error variance decomposition—reflecting the proportion of movements in y_i due to its own shock relative to shocks from other variables in $\mathbf{y_t}$—or the historical decomposition—reflecting the movements of $\mathbf{y_t}$ if e.g. only monetary policy shocks had driven the data.

1. Impulse Response Function

- Impulse response functions (IRF) trace out the adjustment process of the endogenous variable y_t to a one-unit structural shock representing e.g. a monetary policy intervention. Recall that the VMA, that is $\mathbf{y_t} = \mathbf{\Psi_0}\boldsymbol{\varepsilon_t} + \mathbf{\Psi_1}\boldsymbol{\varepsilon_{t-1}} + \mathbf{\Psi_2}\boldsymbol{\varepsilon_{t-2}} + \dots$, relates current observations to an array of past structural shocks. Thus, the (i,j) element in matrix $\mathbf{\Psi_l}$ reports the response of the i^{th} variable to the j^{th} shock at lag l.

Impulse Response

$$\psi_{ij}^{l} = \frac{\partial y_t^i}{\partial \epsilon_{t-l}^j}$$

- Impulse responses are normally reported as function of $l = 1 \dots T$.

↷ IRFs are estimated from the structural form and depend, hence, on the adopted identifying restrictions. Therefore, the identification scheme should always be reported with the IRF!

2. Forecasting

- The reduced form VAR suggests that past data contain information about the future and can, hence, be used to forecast the value $\widetilde{\mathbf{y}}_{t+h|t}$ h periods ahead given the information available at time t.
- A one-period ahead forecast can be obtained by introducing the p most recent values of $\mathbf{y_t}$ into the reduced form VAR, calibrated with the estimated reduced from coefficients $\widehat{\mathbf{B}}$, this is:

$$\widetilde{\mathbf{y}}_{t+1|t} = \widehat{\mathbf{B}}_0 + \widehat{\mathbf{B}}_1 \mathbf{y_t} + \widehat{\mathbf{B}}_2 \mathbf{y_{t-1}} \cdots + \widehat{\mathbf{B}}_p \mathbf{y_{t-p+1}}$$

- Using $\widetilde{\mathbf{y}}_{t+1|t}$ permits to calculate the two-period ahead forecast with

$$\widetilde{\mathbf{y}}_{t+2|t} = \widehat{\mathbf{B}}_0 + \widehat{\mathbf{B}}_1 \widetilde{\mathbf{y}}_{t+1|t} + \widehat{\mathbf{B}}_2 \mathbf{y_t} \cdots + \widehat{\mathbf{B}}_p \mathbf{y_{t-p+2}}$$

and etc. for h-period ahead forecasts.

- A **conditional forecast** pre-specifies some variables in $\widetilde{\mathbf{y}}_{t+h|t}$ whilst, contingent on this, the remaining variables are forecasted. In central banking, this can be useful when comparing the impact of different paths of the monetary policy.
- Forecasting does not necessitate a structural identification.
- Structural breaks due to e.g. economic, political, or financial crises undermine the accuracy of forecasts.
- With stationary variables, a VAR is **mean-reverting** in the sense that long-term forecasts have a tendency to return to the unconditional mean of a variable.
- Forecasting with a VAR is susceptible to the **Lucas critique** when the estimated coefficients are a function of the policy regime. Specifically, economic agents can learn and revise their expectations and reactions to policy interventions. Therefore, the usefulness of VARs lies mainly in short-term forecasting.

Cointegration and Vector Error Correction Models

- Stationarity is a key property for a VAR. Non-stationary variables might warrant some de-trending to avoid **spurious results**.
- Yet, eliminating trends is inappropriate when they genuinely tie the variables together around some equilibrium. Put differently, whether to difference depends on whether the variables are cointegrated.
- The multivariate context lends itself naturally to cointegration, but also adds twists as there might be several cointegrating relationships.

Cointegration

When the variables of $\mathbf{y_t}$ are integrated of order one I(1), but there exists a combination $\beta'\mathbf{y_t}$ that is stationary, the variables are cointegrated.

- Cointegration suggests that the variables of $\mathbf{y_t}$ are tied together by some equilibrium relationship, from which individual time series of y_t can depart in the **short-term**, but to which they revert in the **long-term**. The equilibrium should be associated with economic theory.

- Hence, cointegration calls for a framework that embodies both the long-term trend and the short-term deviations. To obtain this, subtract $\mathbf{y_{t-1}}$ from a reduced form VAR(p), which yields

$$\Delta \mathbf{y_t} = \mathbf{B}_0^* + \mathbf{B}_1^* \Delta \mathbf{y_{t-1}} + \mathbf{B}_2^* \Delta \mathbf{y_{t-2}} + \cdots + \mathbf{B}_\mathbf{p}^* \Delta \mathbf{y_{t-p}} + \mathbf{\Pi y_{t-1}} + \mathbf{e_t}$$

where $\mathbf{\Pi} \equiv -\mathbf{\Psi}(\mathbf{1}) = (\mathbf{I} - \sum_{j=1}^{p} \mathbf{B_j})\mathbf{y_{t-1}}$ is a **disequilibrium error**.
- Cointegration manifests in $\mathbf{\Pi}$ having a reduced rank, in which case it can be disentangled into a $r \times k$ **cointegration vector** β and a $r \times k$ **loading matrix** α, that is $\mathbf{\Pi} = \alpha\beta'$. Substituting this yields:

Vector-Error-Correction Model (VECM)

$$\Delta \mathbf{y_t} = \mathbf{B}_0^* + \mathbf{B}_1^* \Delta \mathbf{y_{t-1}} + \mathbf{B}_2^* \Delta \mathbf{y_{t-2}} + \cdots + \mathbf{B}_\mathbf{p}^* \Delta \mathbf{y_{t-p}} + \alpha\beta' \mathbf{y_{t-1}} + \mathbf{e_t}$$

↷ The error correction mechanism implies that deviations from the co-integrating relationship $\beta' \mathbf{y_{t-1}}$ tend to be followed by a correction of $\Delta \mathbf{y_t}$ with α reflecting the **adjustment speed** towards the equilibrium.

- The **Johansen approach** tests the rank of $\mathbf{\Pi}$—and thus the number of columns in β—which defines the number of cointegrating vectors.

Example 3: Purchasing Power Parity Theory

The purchasing power parity (PPP) theory suggests that tradable goods should cost the same when expressing prices in the same currency, that is $S_t = P_t/P_t^*$ where S_t is the exchange rate and P_t and P_t^* are, respectively, the domestic and foreign prices index. In logarithms, we have

$$s_t = p_t - p_t.$$

Econometrically, PPP suggests that prices and exchange rates should be tied together by a cointegrating relationship. In a bivariate context, this could be tested via the regression

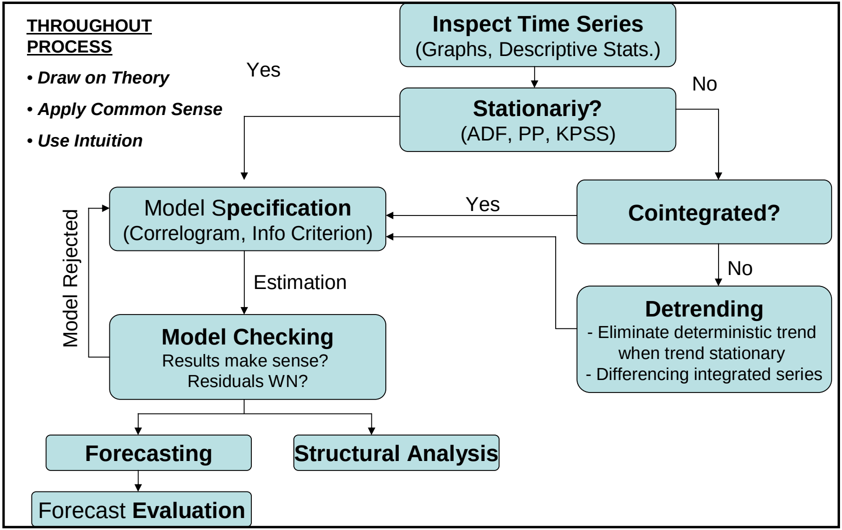
$$s_t = \beta_0 + \beta_1 p_t - \beta_2 p_t + \epsilon_t.$$

Under cointegration, the error term ϵ_t should be stationary. In a multivariate context, with $y_t = (s_t, p_t, p_t^*)'$, the VECM

$$\mathbf{\Delta y_t = B_0^* + \sum B_l^* \Delta y_{t-l} + \Pi y_{t-1} + e_t}.$$

permits to run a trace or maximum eigenvalue test on the roots of $\mathbf{\Pi}$. For the long term (e.g. several decades), there is empirical support for PPP (in particular when average inflation is high and exchange rates are fixed).

Summary and Conclusions



- VAR models offer many advantages when it comes to data description and short-term forecasting. In particular, they are a powerful tool to trace the dynamic interaction that can become cumbersome even with a relatively small number of jointly determined variables.
- Structural inference with impulse-response functions necessitates an adequate identification scheme and is, hence, subject to judgement.
- To obtain a compelling analysis of the effect of e.g. monetary interventions, it is vital to combine theory and empirical results. As regards the VAR methodology, this means that the endogenous variables correspond with the question to be analysed, identifying restrictions are firmly embedded in economic theory, and reflect the policy regime of a given country during the period under investigation.

References

- Bernanke, B.S. and I. Mihov, 1998. Measuring monetary policy, Quarterly Journal of Economics 113(3), 869-902.
- Blanchard, O.J. and D. Quah, 1989. The Dynamic Effects of Aggregate Demand and Supply Disturbances, American Economic Review 79(4), 655-673.
- Geweke, J., 1984. Inference and causality in economic time series models. In: Griliches, Z. and M.D. Intriligator (eds.) Handbook of Economitrics, Vol II, Elsevier.
- Nelson, C. and C. Plosser, 1982. Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications, Journal of Monetary Economics 10, 139-162.
- Johansen, S., 1995. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press.
- Rigobon, R., and B. Sack, 2004, The impact of monetary policy on asset prices, Journal of Monetary Economics, 51, 1553-1575.
- Sims, C., 1980. Macroeconomics and Reality, Econometrica 48(1), 1-48.
- Stock, J.H. and M.W. Watson, 2001. Vector Autoregressions, Journal of Economic Perspectives 15(4), 101-115.
- Uhlig, H., 2005. What are the effects of monetary policy on output? Results from an agnostic identification procedure, Journal of Monetary Economics 52, 381-419.

Textbooks

- Canova, F, 2007, *Methods for Applied Macroeconomic Research*, Princeton University Press.
- Davidson, J., 2000. *Econometric Theory*, Blackwell Publishers.
- Enders, W., 2010. Applied Econometric Time Series, John Wiley and Sons, Ch. 5 and 6.
- Hamilton, J.D., 1994, *Time Series Analysis*, Princeton Uni. Press.
- Lütkepohl, H., 2007. *New Introduction to Multiple Time Series Analysis*, Springer.
- Lütkepohl, H., 2003. Vector Autoregressions, in A Companion to Theoretical Econometrics, ed. B.H. Baltagi, ch. 32.
- Teräsvirta, T, D. Tjøstheim D., and C.W.J. Granger, 2010. *Modelling nonlinear economic time series*, Oxford University Press.
- Watson, M.W., 1994. Vector Autoregressions and Cointegration, in *Handbook of Econometrics*, Vol VI, ed. R.F. Engle and D.L. McFadden.