Chapter 4:

Vector Autoregressive Models
Contents:

IV.1 Vector Autoregressive Models (VAR) .................................................................................................................. 3
  IV.1.1 Introduction .................................................................................................................................................. 3
  IV.1.2 Modelling of a VAR(1)-Models ................................................................................................................. 5
  IV.1.3 VAR(p)-Models with more than two Variables ............................................................................................ 8
  IV.1.4 Estimation of VAR-Models ........................................................................................................................ 9

IV.2 Granger Causality Test and Blockexogeneity .................................................................................................. 13

IV.3 Impulse Response .............................................................................................................................................. 16

IV.4 Orthogonal Impulse Response ........................................................................................................................ 20

IV.5 Variance Decomposition .................................................................................................................................. 23
IV.1 Vector Autoregressive Models (VAR)

IV.1.1 Introduction

In ARIMA models we only derive the actual value from past values for an endogenous variable. However, there is often no theoretical background available. Therefore, we can use Vector Autoregressive (VAR) Models.

1. The single equation approach explains an endogenous variable (e.g. private consumption) by a range of other variables (e.g., disposal income, property, interest rate)
   Assumption: Explanatory variables are exogenous.
   → But: Macroeconomic variables are often not exogenous (endogeneity problem).
2. In theory exists no assumptions about the dynamic adjustment.
   → Dynamic modelling of the consumption function, e.g. adjustment of consumption behaviour to substantial income, occurs only slowly.
Multivariate (linear) time series models eliminate both problems.

→ Development of a variable is explained by the development of potential explanatory variables.
→ Value is explained by its own history and simultaneously by considering various variables and their history.
→ e.g. Dependency of the long-term interest rate from the short run interest rate:
  ▪ Is difficult in univariate time series analysis, because one would need estimations of the development of short run interest rate. Therefore, it is not possible to make good forecast out of such a model.
  ▪ In contrast to a bivariate model, where both, the history of long and short term interest rate are taken into consideration.

With vector autoregressive models it is possible to approximate the actual process by arbitrarily choosing lagged variables. Thereby, one can form economic variables into a time series model without an explicit theoretical idea of the dynamic relations.
IV.1.2 Modelling of a VAR(1)-Models

The most easy multivariate time series model is the bivariate vector autoregressive model with two dependent variables $y_{1,t}$ and $y_{2,t}$, where $t = 1, ..., T$. The development of the series should be explained by the common past of these variables. That means, the explanatory variables in the simplest model are $y_{1,t-1}$ and $y_{2,t-1}$. The VAR(1) with lagged values for every variable is determined by:

\[ y_{1,t} = \alpha_{11}y_{1,t-1} + \alpha_{12}y_{2,t-1} + \varepsilon_{1,t} \]
\[ y_{2,t} = \alpha_{21}y_{1,t-1} + \alpha_{22}y_{2,t-1} + \varepsilon_{2,t} \]

Matrix Notation:

\[ y_t = A_1y_{t-1} + \varepsilon_t \]

\[ A_1 = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \]
Assumptions about the Error Terms:

1. The expected residuals are zero:
   \[ E[\varepsilon_{i,t}] = 0 \text{ with } i = 1, 2 \]

2. The error terms are not autocorrelated:
   \[ E[\varepsilon_{i,t} \cdot \varepsilon_{j,\tau}] = 0 \text{ with } t \neq \tau \]
Interpretation of VAR Models:

VAR-Models themselves do not allow us to make statements about causal relationships. This holds especially when VAR-Models are only approximately adjusted to an unknown time series process, while a causal interpretation requires an underlying economic model. However, VAR-Models allow interpretations about the dynamic relationship between the indicated variables.

→ Granger Causality

Be careful, Granger Causality is a much weaker argument than normal causality. If \( y_2 \) has any influence on \( y_1 \) (\( \alpha_{12} \neq 0 \)), one can say \( y_2 \) Granger causes \( y_1 \). On the other hand, if \( y_1 \) has any influence on \( y_2 \) (\( \alpha_{21} \neq 0 \)), one can say \( y_1 \) Granger causes \( y_2 \).
IV.1.3 VAR(p)-Models with more than two Variables

An VAR(p)-Model, with $p$ variables, is given as:

$$ y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t $$

If one wants to expand the equation with a trend, intercept or seasonal adjustment, it will be necessary to augment the Vector $x_t$, which includes all the deterministic components, and the matrix $B$ (VARX-Model):

$$ y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + B x_t + \varepsilon_t $$
IV.1.4 Estimation of VAR-Models

Specifications:

1. Determination of endogenous variables according to economic theory, empirical evidence and experience.
2. Transformation of time series (take logs or log-returns).
3. Insert seasonal component, especially for macro data.

→ Estimation according to LS-Regression

→ If the error terms between the variables are uncorrelated, the estimation will be unbiased and efficient.
Determination of Lag Length:

The determination of lag length is a trade-off between the curse of dimensionality and abbreviate models, which are not appropriate to indicate the dynamic adjustment.

If the lag length is too short, autocorrelation of the error terms could lead to apparently significant and inefficient estimators. Therefore, one would receive wrong results.

With the so-called curse of dimensionality we understand, that even with a relatively small lag length a large number of parameters if required. On the other hand, with increasing number of parameters, the degrees of freedom decrease, which could possibly result in significant of inefficient estimators.

Information Criteria:

The idea of information criteria is similar to the trade-off discussed above. On the one hand, the model should be able to reflect the observed process as precise as possible (error terms should be as small as possible) and on the other hand, too many variables lead to inefficient estimators. Therefore, the information criteria are combined out of the squared sum of residuals and a penalty
term for the number of lags. In detail, for T observations we chose the lag length p in a way that the reduction of the squared residuals after augmenting lag p+1, is smaller than the according boost in the penalty term.

Squared residuals:

$$\ln \left( \frac{\hat{e}' \hat{e}}{T} \right)$$

Penalty terms:

AIC:

$$\frac{2p}{T}$$

SIC:

$$\frac{p \cdot \ln(T)}{T}$$

HQ:

$$\frac{2 \cdot c \cdot p \cdot \ln(T)}{T} \text{ with constant } c > 1$$
Statistical Tests:

An alternative to the global preparation are local statistic tests. For example, one could use Log Likelihood-Ratio Test (LR) to specify the lag length. Based on a VAR-Model with lag length $p$, one can check if the explanatory power of the model increased after taken lag $p+1$ into consideration. With the support of the LR test, one can control if the change in the power is significant or if it has only a power comparable to a random variable. Formally the LR Test is defined as,

$$
\lambda = \frac{L(\beta_r)}{L(\beta_{ur})}
$$

where $r$ stands for restricted and $ur$ for unrestricted model. Hence, it is tested whether it is possible to restrict all $p+1$ lags to zero. Additional lags cannot increase the power of estimation if $\lambda$ is close to one, but if $\lambda$ is close to zero. Therefore, the usual Wald-Test, which is adequate for an $\chi^2$-distribution can be used.

Unfortunately, the LR-test is not sufficient for models with missing lag structures.
IV.2 Granger Causality Test and Blockexogeneity

Granger (1969) developed a test approach to proof if a time series $X$ contribute to the prediction of another series $Y$.

Granger Causality is exists if the mean squared forecast error (MSE) by using the series $X$ in the forecast model is smaller than without consideration of $X$:

\[
\text{MSE}_Y(\hat{Y}_{t+h} \mid l_t) < \text{MSE}_Y(\hat{Y}_{t+h} \mid (l_t \text{ without } X_{t-s}, s = 1, \ldots, n))
\]

for at least one $h = 1, 2, \ldots, m$

with: $\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$ and $\hat{Y}_i$ equals the forecast in time point $i$ and $N$ equals the number of forecasts, $h$ equals the forecast horizon.
In the bivariate approach, it follows:

\[ Y_t = \alpha + \sum_{i=1}^{n} \beta_i \cdot Y_{t-i} + \sum_{j=1}^{m} \delta_j \cdot X_{t-j} + \varepsilon_t \]

The test statistic of the Wald test (W) equals to \( J \cdot F \). Thereby, \( J \) is the number of restrictions to test (in the above case \( J = m \)). \( F \) denotes the value of the F statistic with

\[
F = \frac{(\hat{\varepsilon}^\prime \hat{\varepsilon} - \hat{\varepsilon}_r^\prime \hat{\varepsilon}_r) / J}{\hat{\varepsilon}_{\hat{\varepsilon}}^\prime \hat{\varepsilon}_{\hat{\varepsilon}} / (T - K)}
\]

where \( \hat{\varepsilon}_r \), equals the sum of the squared residuals by imposition of restrictions \( (\delta_1 = \delta_2 = \ldots = \delta_m = 0) \) and \( \hat{\varepsilon}_{\hat{\varepsilon}} \) is the sum of squared residuals of the estimation without restrictions.
T is the number of observations and K the number of regressors of the model. The test statistic W is asymptotically $\chi^2$-distributed with J degrees of freedom.

**Test for Blockexogeneity:**

Hypothesis:

- $y_3$ is (block) exogenous for $y_1$ and $y_2$
  
  $H_0: \alpha_{13} = \alpha_{23} = 0$

- $y_2$ and $y_3$ are (block) exogenous for $y_1$
  
  $H_0: \alpha_{12} = \alpha_{13} = 0$

\[
\begin{align*}
y_{1,t} &= \alpha_{11}y_{1,t-1} + \alpha_{12}y_{2,t-1} + \alpha_{13}y_{3,t-1} + \varepsilon_{1,t} \\
y_{2,t} &= \alpha_{21}y_{1,t-1} + \alpha_{22}y_{2,t-1} + \alpha_{23}y_{3,t-1} + \varepsilon_{2,t} \\
y_{3,t} &= \alpha_{31}y_{1,t-1} + \alpha_{32}y_{2,t-1} + \alpha_{33}y_{3,t-1} + \varepsilon_{3,t}
\end{align*}
\]
**IV.3 Impulse Response**

The dynamic adjustment of reciprocal dependency is immediately not considerable. The impulse response test shows the effects of an exogenous shock on the whole process over time. Therefore, one can detect the dynamic relationships over time.

**Idea:**

Initially, look at the adjustment of the endogenous variables over time, after a hypothetical shock in t. This adjustment is compared with the time series process without a shock, i.e. the actual process. The impulse response sequences plot the difference between this two time paths.
To illustrate this, we assume a two dimensional VAR(1)-Model:

\[ y_{1,t} = \alpha_{11} y_{1,t-1} + \alpha_{12} y_{2,t-1} + \varepsilon_{1,t} \]
\[ y_{2,t} = \alpha_{21} y_{1,t-1} + \alpha_{22} y_{2,t-1} + \varepsilon_{2,t} \]

Initially, in \( t = 1 \) we assume a shock in the error term \( \varepsilon_{1,1} \) of the first equation. This shock has a direct effect on \( y_{1,1} \) of exactly the same amount. Whereas \( y_{2,1} \) is not effected, assuming that \( \varepsilon_{2,t} = 0 \) with \( t = 1, \ldots, T \). In the second Period (\( t = 2 \)), the original shock has still an effect over the lagged value of \( y_1 \). The effect on \( y_{1,2} \) is \( \alpha_{11} \varepsilon_{1,1} \) and the effect on \( y_{2,2} \) is \( \alpha_{21} \varepsilon_{1,1} \). In the third period the effect on \( y_{1,3} \) is not only \( \alpha_{11} (\alpha_{11} \varepsilon_{1,1}) \), but also \( \alpha_{12} (\alpha_{21} \varepsilon_{1,1}) \). Accordingly, the effect on \( y_{2,3} \) is \( \alpha_{21} (\alpha_{21} \varepsilon_{1,1}) + \alpha_{22} (\alpha_{21} \varepsilon_{1,1}) \). Thus, it is possible to obsess the effect of a non-recurring shock in one variable, to all variables over time. One could summarise the result in:

\[ y_{t} = \sum_{k=0}^{\infty} c_k \varepsilon_{t-k} \]

with \( c_0 = I \) (Vector-Moving-Average Process) and where \( c_k \) are the weight of past stocks.
Figure 1: Impulse-Response Sequences for Industrial Production and Orders Received
Figure 1 shows the adjustment of the impulse response sequence for the example of industrial production and orders received, based on a shock in the amount of the standard errors in both variables. On the left side of the figure are plotted the reactions on a shock in industrial production, on the right side the reactions on a shock in orders received. According to our assumption, the industrial production has no effect on orders received and the orders received have no influence on industrial production in the first period. Thus, these graphs start at the point of origin. In contrast, a shock in one variable has an instant effect on its present value. Therefore, the upper-left and lower-right graph begin at the respective standard error. The effects in the following periods depend on the dimension of the coefficients. If the sum of all coefficients in one equation is smaller than one, the effects will decrease over time and will revert to a value close to zero after a certain period.
IV.4 Orthogonal Impulse Response

In the previous impulse response model we assumed that the error terms of the different equation are uncorrelated. However, this assumption is quite restricted. A hypothetical shock in only one equation does not respond a realistic adjustment process. To control for correlation between error terms we have to use the orthogonal impulse response sequences. The idea is to modify the original moving-average construction in a way that the residuals are uncorrelated, i.e. the residuals have to be orthogonal to each other. Therefore, we can write

\[ y_t = \sum_{k=1}^{\infty} \tilde{c}_k v_{t-k} \]

with \( \tilde{c}_k = C_k \cdot G \), where \( G \) is a transformation matrix with the property \( G^{-1} \Omega G^{-1} = I \) (Cholesky-Decomposition).
The error terms of the modified system are \( \nu_{t-k} = G^{-1}\varepsilon_{t-k} \). The variance-covariance matrix of \( \nu_{t-k} \) is diagonal, according to the properties of \( G \). However, the \( G \) matrix is not clearly defined by the Cholesky decomposition \( \hat{\Omega} = G^{-1}G^{-1} \), where \( \hat{\Omega} \) is the original variance-covariance matrix. Moreover, we have to specify the order of the variables. The chosen order presumes the causal relationship between the variables. The results of the impulse response can depend highly on the order of the variables, especially when they are highly correlated.

Figure 2 shows the orthogonal impulse response sequence for the variable order industrial production, orders received and vice versa. For the direction industrial production and orders received the joint component of error terms are only assigned to industrial production. The impulse response and the orthogonal impulse response with respect to a shock in industrial production are therefore the same. However, the effects on a change in orders received are different. Due to the high correlation, parts of the effect can be assigned to industrial production immediately.

As long as the economic theory gives no explicit information about the order of the causal relationships between variables, there is no unique solution.
Figure 2: Two-Sided Orthogonal Impulse Response for Industrial Production and Orders Received
IV.5 Variance Decomposition

An alternative of impulse response, to receive a compact overview of the dynamic structures of a VAR Model, are variance decomposition sequences. This method is also based on a vector moving average model and orthogonal error terms. In contrast to impulse response, the task of variance decomposition is to achieve information about the forecast ability. The idea is, that even a perfect model involves ambiguity about the realisation of $y_{i,t}$, because the error terms associate uncertainty. According to the interactions between the equations, the uncertainty is transformed to all equations. The aim of the decomposition is to reduce the uncertainty in one equation to the variance of error terms in all equations.
Assume a two-dimensional VAR model (Moving average representation):

\[
y_{1,t} = \tilde{c}_{11} v_{1,t} + \tilde{c}_{12} v_{2,t} + \tilde{c}_{13} v_{1,t-1} + \tilde{c}_{14} v_{2,t-1} + \ldots
\]

\[
y_{2,t} = \tilde{c}_{21} v_{1,t} + \tilde{c}_{22} v_{2,t} + \tilde{c}_{23} v_{1,t-1} + \tilde{c}_{24} v_{2,t-1} + \ldots
\]

As the development of the lagged error terms is already known, it exists only uncertainty about the present error terms \( v_{1,t} \) and \( v_{2,t} \).

A compact illustration is shown in figure 3, where the share of the single error terms is plotted against time. For the industrial production it is shown that it depends mainly on its own error terms, independent of the order of the variables. However, after a while the orders received gets more and more important and outdistance the industrial production in period 17 or 6, respectively. On the other hand, the variance decomposition of the orders received replies a clear picture. The own variance dominates the variance of industrial production even in the long run.
Figure 3: Variance Decomposition of Industrial Production and Orders Received