Methods of Empirical Macroeconomics

Emanuel

January 2021

AR(1)

$$x_t = \mu + \alpha x_{t-1} + \epsilon_t \tag{1}$$

$$\forall\ k = t-1 : x_t = \mu \sum_{i=0}^{t-1} \alpha^i + \alpha^t x_o + \sum_{i=0}^{t-1} \alpha^i \epsilon_{t-i} \tag{2}$$

Given that $|\alpha| < 1$, we have for $t \to \infty$:

$$x_t = \mu \frac{1}{1-\alpha} + \sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i} \tag{3}$$

or in terms of the general solution (initial value) via parts of the solution:

$$x_t = \mu \sum_{i=0}^{t-1} \alpha^i + A\alpha^t + \sum_{i=0}^{t-1} \alpha^i \epsilon_{t-i} \tag{4}$$

where you can plug in x_0 and solve for A to get an initial value.

The **parts** of the solution are:

- homogenous part: $x_t = \alpha x_{t-1}$ with $z = \alpha$ s.t $x_t = A\alpha^t$
- $x_t = \mu + \alpha x_{t-1}$ with proposal of $x_t = k$ or $x_t = kt$
- $x_t =_{t-1} + \epsilon_t$ with proposal $x_t = \sum_{i=0}^{\infty} c_i \epsilon_{t-i}$

AR(2)

$$x_t = \mu + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \epsilon_t \tag{5}$$

$$x_t = \frac{\mu}{1 - \sum_{i=1}^{2} \alpha_i} + \sum_{j=1}^{2} A_j z_j^t + \sum_{i=o}^{\infty} c_i \epsilon_{t-i} \tag{6}$$

- homogenous part: $x_t = A_1 z_1^t + A_2 z_2^t$
- $x_t = \mu + \alpha_1 x_{t-1} + \alpha_2 x_{t-2}$ with same proposal as for the AR(1)
- $x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \epsilon_t$ again the same as above with solution $x_t = \sum_{i=0}^{\infty} (B_1 z_1^i + B_2 z_2^i)\epsilon_{t-i}$

$\forall\ \alpha_i\ s.t\ \sum_{i_1}^{2} \alpha_i \neq 1$
$c_i = \alpha_1 c_{i-1} + \alpha_2 c_{i-2} = B_1 z_1^i + B_2 z_2^i$
$z_{1,2} = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 - 4\alpha_2}}{2}$

Finding the MA coefficients c_i:

$c_i = \alpha_1 c_{i-1} + \alpha_2 c_{i-2}$ which we find via the method of undetermined coefficients. From solving the homogeneous part, we know that we can solve that as follows:

$$c_0 = B_1 z_1^0 + B_2 z_2^0 = B_1 + B_2 = 1 \tag{7}$$

$$c_1 = B_1 z_1^1 + B_2 z_2^1 = B_1 z_1 + B_2 z_2 = \alpha_1 \tag{8}$$

which we can solve to obtain $B_1 = 1 - \frac{\alpha_1 - z_1}{z_2 - z_1}$ and $B_2 = \frac{\alpha_1 - z_1}{z_2 - z_1}$

AR(p)

If $z_1, z_2, ..., z_p$ all lie within the complex unit circle, the process is stationary and we have the following:

$$x_t = \mu \sum_{i=0}^{\infty} c_i + \sum_{i=0}^{\infty} c_i \epsilon_{t-i} \tag{9}$$

which we call a $MA(\infty)$ representation. From that, it follows that the conditional expectation and conditional variance converge to the unconditional ones:

$$\mathbb{E}(x_t) = \mu \sum_{i=0}^{\infty} c_i \text{ and } \text{Var}(x_t) = \sigma^2 \sum_{i=0}^{\infty} c_i^2 \text{ and } \gamma(j) = \sigma^2 \sum_{i=0}^{\infty} c_i c_{i+j}$$

Alternatively solving these processes via Lag-operators

Link between linear stochastic difference equations and polynomials.

$$\mathbf{L}^k x_t = x_{t-k} \tag{10}$$

Applying this operator to our $AR(p)$ process yields:

$$x_t = \sum_{k \in K} \alpha_k x_{t-k} + \epsilon_t = \sum_{k \in K} \alpha_k \mathbf{L}^k x_t + \epsilon_t = \left[1 - \sum_{k \in K} \alpha_k \mathbf{L}^k\right] x_t = \epsilon_t \tag{11}$$

We define $\left[1 - \sum_{k \in K} \alpha_k \mathbf{L}^k\right] x_t = \alpha(L) x_t$

$$x_t = \alpha(L)^{-1} \epsilon_t \tag{12}$$

Let us focus on the polynomial $\alpha(L)$. Given $\alpha(L)$ is invertible, we can derive a quite general solution to our problem. $\alpha(L)$ can - using the fundamental theorem of algabra - be factorized. The fundemental theorem of algebra states: $f(x) = a \times (x - c_1) \times ... \times (x - c_n)$

$$\alpha(L) = (1 - z_1 L) \times (1 - z_2 L) \; \forall \; z_i = \frac{1}{\lambda_i} \tag{13}$$

with λ_i solving the polynomial in lagged terms.
That allows us to solve the stochastic difference equation as follows (note that the roots of the lag-polynomial (λ_1) need to lay outside the unit circle as they stand in a inverse relation to the roots of the characteristic equation (z_i)):

$$x_t = \alpha(L)^{-1} \epsilon_t = (1 - z_1 L)^{-1} \times (1 - z_2 L)^{-1} \epsilon_t \tag{14}$$

We can deduce properties for the coefficients s.t our process is stationary (s.t all the characteristic roots (of the non-lagged form) lay within the unit circle:

- $\sum_{j=1}^{p} \alpha_j < 1$ necessary
- $\sum_{j=1}^{p} |\alpha_j| < 1$ sufficient
- at least one unit root if $\sum_{j=1}^{p} \alpha_j = 1$

If there is at least one $\alpha_i = 1$, we call that a unit root process or a **random walk**

Solving that shit via Lag-operators done practical

$x_t = \alpha(L)^{-1} \epsilon_t = (1 + \alpha L + \alpha^2 L^2 + ...) \epsilon_t$
Idea of Proof: $(1 - z)^{-1} = 1 + z + z^2 + ... \forall \; |z| = |\alpha L| < 1$ as $S = 1 + z + z^2$ and $-zS = -z - z^2 - ...$ so we have $S - zS = 1$ and thus $S = \frac{1}{1-z} = (1 - z)^{-1}$

Let $z_1 z_2 = -\alpha_2$ and $z_1 + z_2 = \alpha_1$ which we can use to factorize $1 - \alpha_1 L - \alpha_2 L^2$ into $(1 - z_1 L)(1 - z_2 L)$.
We therefore have:

$$x_t = (1 - z_1 L)^{-1} (1 - z_2 L^2)^{-1} \epsilon_t = \left(\sum_{j=0}^{\infty} z_1^j L^j\right) \left(\sum_{j=0}^{\infty} z_2^j L^j\right) \epsilon_t$$

Summing cleverly, we get:

$$\left(\sum_{j=0}^{\infty} z_1^j L^j\right) \left(\sum_{j=0}^{\infty} z_2^j L^j\right) = \sum_{j=0}^{\infty} \left(\sum_{k=0}^{j} z_1^k z_2^{j-k}\right) L^j$$

More often used - at least for $AR(2)$ - are partial fractions:

$$\frac{1}{(1-z_1L)(1-z_2L)} = \frac{a}{(1-z_1L)} + \frac{b}{(1-z_2L)}$$

yielding $b = \frac{z_2}{z_2-z_1}$ and $a = \frac{z_1}{z_1-z_2}$ which we use to write:

$$x_t = \left(\frac{z_1}{z_1-z_2} \sum_{j=0}^{\infty} z_1^j L^j + \frac{z_2}{z_2-z_1} \sum_{j=0}^{\infty} z_2^j L^j \right) \epsilon_t = \left(\frac{z_1}{z_1-z_2} \sum_{j=0}^{\infty} z_1^j \epsilon_{t-j} + \frac{z_2}{z_2-z_1} \sum_{j=0}^{\infty} z_2^j \epsilon_{t-j} \right) \tag{15}$$

$$\sum_{j=0}^{\infty} \left(\frac{z_1}{z_1-z_2} z_1^j + \frac{z_2}{z_2-z_1} z_2^j \right) \epsilon_{t-j} \tag{16}$$

Remember that $a = B_1$ and $b = B_2$, which means that we have found the MA coefficients!

Impulse Responses for AR(1)

Describes the path x follows if it is kicked by a single shock ϵ_t.

$$\mathbb{E}_t[x_{t+k}] = \mathbb{E}\left[\mu \sum_{j=0}^{k-1} \alpha^j + \alpha^k x_t + \sum_{j=0}^{k-1} \alpha^j \epsilon_{t+k-j} \right]$$

$$= \mu \sum_{j=0}^{k-1} \alpha^j + \alpha^k x_t$$

That means we get:

$$IR(x_t, k) = \frac{\partial}{\partial \epsilon_t} \mathbb{E}_t[x_{t+k}] = \frac{\partial}{\partial \epsilon_t} \left(\mu \sum_{j=0}^{k-1} \alpha^j + \alpha^k (\mu + _{t-1} + \epsilon_t) \right) = \alpha^k$$

Here lays the intuition for the stationarity: if the α's are smaller then 1, those impulse responses decay geometrically and the system returns to its old state.

Impulse Responses for AR(2)

As before, we have:

$$IR(x_t, k) = \frac{\partial}{\partial \epsilon_t} \mathbb{E}_t[x_{t+k}]$$

$$\begin{aligned}
E_t[x_t] &= aE_t[x_{t-1}] + bE_t[x_{t-2}] + E_t[\epsilon_t] \quad \Rightarrow IR_0 = 1 \\
E_t[x_{t+1}] &= aE_t[x_t] + bE_t[x_{t-1}] \quad \Rightarrow IR_1 = aIR_0 = a \\
E_t[x_{t+2}] &= aE_t[x_{t+1}] + bE_t[x_t] \quad \Rightarrow IR_2 = aIR_1 + bIR_0 = a^2 + b \\
E_t[x_{t+3}] &= aE_t[x_{t+2}] + bE_t[x_{t+1}] \quad \Rightarrow IR_3 = aIR_2 + bIR_1 = a(a^2+b) + b(a) \\
&\vdots
\end{aligned}$$

This explains how we construct our impulse responses in the code. For the AR(2) process you have in general

$$\begin{aligned}
& IR_k = aIR_{k-1} + bIR_{k-2} \\
\Leftrightarrow \quad & IR_k = \begin{bmatrix} b & a \end{bmatrix} \begin{bmatrix} IR_{k-2} \\ IR_{k-1} \end{bmatrix}.
\end{aligned}$$

(Weak- or Covariance-) Stationarity

- the unconditional mean is time invariant, i.e $\mathbb{E}[x_t] = \mathbb{E}[x_{t-j}]$
- the uncond. variance is time invariant, i.e $\text{Var}(x_t) = \text{Var}(x_{t-j})$

- the uncond. covariance is time invariant, i.e $\text{Cov}(x_t, x_{t-j}) = \text{Cov}(x_{t+k}, x_{t+k-j}) = \gamma_j$
  That is very intuitive as the covariance will depend on the lag j but if the lag is kept constant and t is changed, the covariance should remain unchanged.

Important note: the $MA(\infty)$-representation of an AR(p) exists only if the process is covariance stationary.
Important note: the impulse responses of a stationary variable always equal the MA coefficients

Judging Stationarity by looking at the time series:

- If the IR does not converge to zero, the process is not stationary
- The permanent impact might be an indicator for unit roots (and if the IR explodes: $\alpha_i > 1$)
- periodic behavior of the IR is an indicator for complex roots!
- if at least one of the roots is negative, the IR will oscillate
- if $d = 0$, we have $z_{1,2} = \alpha_1/2$ and another solution $t(\frac{\alpha_1}{2})^t$ to get $c_i = B_1(\frac{\alpha_1}{2})^t + B_2(\frac{\alpha_1}{2})^t$ to generate a non-monotonic IR and oscillation if $\alpha_1 < 0$. See Slides Univariate Dynamics, p. 31

Complex Characteristic Roots

If there is a $MA(\cdot)$-representation to the time series $\{x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \epsilon_t\}_t$, we have that (as seen before):

$$x_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$$

We know that for an AR(2)-process, we have $c_j = \alpha_1 c_{j-1} + \alpha_2 c_{j-2}$, which is a Diff.eq that we can solve via $c_j = Az^j$
We get the following equation: $z^2 - \alpha_1 z - \alpha_2 = 0$ with two roots and the solution $c_j = A_1 z_1^j + A_2 z_2^j$. What happens if $z_{1,2}$ happen to be complex roots?
Let $z = u + iv$ with $u = \cos(\phi)\theta$ and $v = \sin(\phi)\theta$ and hence $z = \cos(\phi)\theta + i\sin(\phi)\theta = \theta(\cos(\phi) + i\sin(\phi)) \overset{Euler}{=} \theta\exp(i\phi)$
θ is the modulus of the complex number, i.e $\theta = \sqrt{u^2 + v^2}$. We saw in the lectures that $z_{1,2}$ must be complex conjugates, i.e $z_1 = \theta(\cos(\phi) + i\sin(\phi))$ and $z_2 = \bar{z_1}$. Using the derivation for powers of these complex roots, we get:

$$z_1^j = \theta^j(\cos(j\phi) + i\sin(j\phi))$$

and

$$z_2^j = \theta^j(\cos(j\phi) - i\sin(j\phi))$$

Note: $i = \sqrt{-1}$

Using $c_j = A_1 z_1^j + A_2 z_2^j = A_1\theta^j(\cos(j\phi) + i\sin(j\phi)) + A_2\theta^j(\cos(j\phi) - i\sin(j\phi)) = 2B\theta^j\cos(C + j\phi)$, which implies that $\theta \overset{!}{<} 1$, i.e the roots lay within the unit circle.

Unit Roots

We looked at how we can solve these $AR(p)$ processes and found out that $|\alpha_i| < 1$ has to hold. We are now interested in the case of $|\alpha_i| = 1$ because random walks (with or w.o drift) are hugely important in economics. Additionally, a process with a unit root is non-stationary.A unit root process is one that fulfills the following criteria:

$$x_t = x_{t-1} + \epsilon_t$$

and thus also

$$x_t(1 - \{z_1 \neq 1\}L)(1 - L) = \epsilon_t$$

Non-stationarity of Unit-Root $AR(1)$ process:

$$\mathbb{E}_t[x_t] = \mathbb{E}_t[x_{t-2} + \epsilon_{t-1} + \epsilon_t] = \mathbb{E}_t[x_0 + \sum_{i=0}^{t} \epsilon_{t-i}] = \mathbb{E}_t[x_0] \tag{17}$$

$$\text{Var}(x_t) = \text{Var}(x_0 + \sum_{i=0}^{t} \epsilon_{t-i}) \overset{i.i.d}{=} t\sigma^2 \tag{18}$$

which is not time-constant.
We now search a way to transform the non-stationary time-series (n-s because of unit roots) into a stationary one.

If a process has $k$unit roots, the process is $I(k)$, i.e $\{x_t\}_t \sim I(k)$. We can transform the time series into a stationary one by differencing. That means that:

$$\{x_t\}_t \sim I(k) \iff \Delta x_t \sim I(k-1)$$

Motivation behind Differencing: We see that the Google stock price (a) is a non-stationary time-series but the changes in the stock prices (b) build a stationary time series. That means by differencing once, we get a stationary time series. In theory, we can do that a many times as we like to obtain a stat. time-series. If $k = 1$, the process is **difference-stationary**.

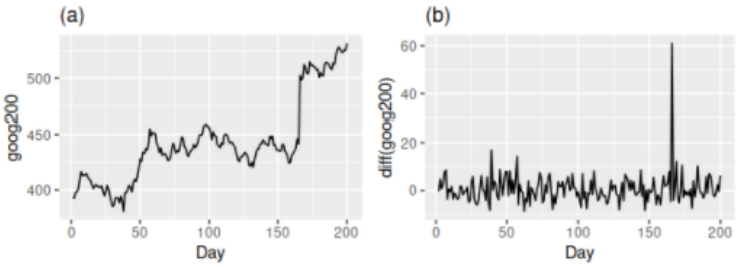
From first difference models to level models and vice versa

Let $\{x_t\}_t$ be a time series with $x_t = 1.2x_{t-1} - 0.2x_{t-2} + \epsilon_t$ and we subtract x_{t-1} to get:

$$x_t - x_{-1} = \Delta x_t = 1.2x_{t-1} - x_{t-1} - 0.2x_{t-2} + \epsilon_t$$

$$\Delta x_t = 0.2\Delta x_{t-1} + \epsilon_t$$

and if it does not work that nicely ($\alpha_1 - 1 = \alpha_2$), just use factorization. One always loses one lag when going from level to difference. The same has to be done in the other direction.



Beveridge-Nelson Decomposition

Decomposing any $\{x_t\}_t$ that is difference-stationary (exactly one unit root) into a random walk part and a stationary MA part. To derive the parts explicitly, see examples from page 15 onwards.

$$x_t = \underbrace{c(1)\sum_{i=0}^{t}\epsilon_i}_{\text{random walk } x_t^P} + \underbrace{c^*(L)\epsilon_t}_{\text{stationary MA } x_t^T}$$

If there are unit roots in time-series, the estimates of the coefficients will no longer be asymptotically normal. Thus we conduct a Unit-root test and difference if needed. Problem: These tests have low power.

Derivation:

$$\Delta x_t = A(L)^{-1}\epsilon_t = C(L)\epsilon_t$$

where $C(L)\epsilon_t = C(1)\epsilon_t + (C(L) - C(1))\epsilon_t = C(1)\epsilon_t + C^*(L)(1-L)\epsilon_t$ as we can factor out $(1-L)$ as $C(L) = \sum_{i=0}^{t}(CL)^i$ and $C(1) = \sum_{i=0}^{t} C^i$. We then divide by $(1-L)$ to get

$$x_t = C(1)(1-L)^{-1}\epsilon_t + C^*(L)\epsilon_t$$

As we know that $(1-L)^{-1} = 1 + L + L^2 + ...$, we get our result:

$$x_t = C(1)\sum_{i=0}^{t}\epsilon_i + C^*(L)\epsilon_t$$

Dickey-Fuller Test

1. Specifying the model to estimate (time trend and drift excluded or included; a and bt)
2. $x_t = \rho x_{t-1} + \epsilon_t \to \Delta x_t = \gamma x_{t-1} + \epsilon_t$, wobei $\gamma = \rho - 1$
   We thus want to test whether $\gamma = 0$. We need to difference as we cant test using a non-stationary time series. We thus cant use a t-test to test $H_0 : \gamma = 0$. More formally: As $\rho \to 1$, our OLS estimator is no longer asymptotically normal.
3. use the critical values calculated by Dickey in 1981 to test

4. fundamental problem: knowing whether to indclude a time trend and/or drift require knowing whether there are unit roots (e.g running the different models via. OLS and choosing the best). See PS 6 for the auxiliary regression to run for the DF test.

5. If there is no evidence for a unit root, there might just be stochastic or deterministic trends in the data. Clarification: Stochastic trend: Drift μ and deterministic trend: bt. Interesting: As Random Walks are very persistent, a RW might visually fool one into thinking there is a trend.

Vector Autoregressions (VARs)

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} a_{11}^1 & a_{22}^1 \\ a_{21}^1 & a_{22}^1 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{11}^2 & a_{22}^2 \\ a_{21}^2 & a_{22}^2 \end{pmatrix} \begin{pmatrix} x_{1,t-2} \\ x_{2,t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

So we have e.g two dependent variables that are influenced by one another in different time periods. So x_1 is for example a linear combination of its and x_2 past values and the error term $\epsilon_{1,t}$

We can easily generalise from the AR models that the solution to this stochastic DE has to be of the form:

$$\mathbf{A}(L)\mathbf{x}_t = \epsilon_t$$

with $A(L) = I - \sum_{i=1}^{p} A_i L^i$. How do we get $A(L)^{-1}$ to solve the VAR?

$$A(L)^{-1} = C(L) = \sum_{i=0}^{\infty} A^i L^i \text{ and } A(1)^{-1} = \sum_{i=0}^{\infty} A^i$$

which can be found on p.9, VAR.
We can also derive the important concepts being expectations, variances, impulse responses and stationarity:

	$\mathbf{x}_{t+j}$	$\mathbf{E}_t(\mathbf{x}_{t+j})$
$j = 0$	$\mathbf{x}_t$	$\mathbf{x}_t$
$j = 1$	$\boldsymbol{\mu} + \mathbf{A}\mathbf{x}_t + \boldsymbol{\varepsilon}_{t+1}$	$\boldsymbol{\mu} + \mathbf{A}\mathbf{x}_t$
$j = 2$	$\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\mu} + \mathbf{A}^2\mathbf{x}_t + \mathbf{A}\boldsymbol{\varepsilon}_{t+1} + \boldsymbol{\varepsilon}_{t+2}$	$\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\mu} + \mathbf{A}^2\mathbf{x}_t$
	$\vdots$	
$j = k$	$\sum_{i=0}^{k-1} \mathbf{A}^i\boldsymbol{\mu} + \mathbf{A}^k\mathbf{x}_t + \sum_{i=0}^{k-1} \mathbf{A}^i\boldsymbol{\varepsilon}_{t+k-i}$	$\sum_{i=0}^{k-1} \mathbf{A}^i\boldsymbol{\mu} + \mathbf{A}^k\mathbf{x}_t$
$k \to \infty$		$(\mathbf{I} - \mathbf{A})^{-1}\boldsymbol{\mu}$

	$\mathbf{x}_{t+j} - \mathbf{E}_t(\mathbf{x}_{t+j})$	$\mathbf{E}_t[(\mathbf{x}_{t+j} - \mathbf{E}_t(\mathbf{x}_{t+j}))^2]$	$\partial\mathbf{E}_t(\mathbf{x}_{t+j})/\partial\boldsymbol{\varepsilon}_t$
$j = 0$	0	0	$\mathbf{1}$
$j = 1$	$\boldsymbol{\varepsilon}_{t+1}$	$\boldsymbol{\Sigma}$	$\mathbf{A}$
$j = 2$	$\mathbf{A}\boldsymbol{\varepsilon}_{t+1} + \boldsymbol{\varepsilon}_{t+2}$	$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' + \boldsymbol{\Sigma}$	$\mathbf{A}^2$
	$\vdots$		
$j = k$	$\sum_{i=0}^{k-1} \mathbf{A}^i\boldsymbol{\varepsilon}_{t+k-i}$	$\sum_{i=0}^{k-1} \mathbf{A}^i\boldsymbol{\Sigma}(\mathbf{A}^i)'$	$\mathbf{A}^k$
$k \to \infty$		$(\mathbf{I} - (\mathbf{A} \otimes \mathbf{A}))^{-1}\text{vec}(\boldsymbol{\Sigma})$	$\mathbf{0}$

Companion Form

In order to invert VARs of higher order, we want to turn the VAR into a matrix structure to be able to handle it more easily. We begin with our $VAR(p)$ that is given by $x_t = \mu + \sum_{i=1}^{p} A_i x_{t-i} + \epsilon_t$. By adding identities, we obtain the following:

$$\begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \cdots \\ \mathbf{x}_{t-(p-1)} \end{pmatrix} = \begin{pmatrix} \mu \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I} & \mathbf{0} & & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \mathbf{x}_{t-3} \\ \vdots \\ \mathbf{x}_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

$$\mathbf{Z}_t = \mathbf{d} + \mathbf{A}_0 \mathbf{Z}_{t-1} + \mathbf{u}_t$$

We can invert the process to get our usual solution and we can retrieve our original VAR (w.o the added identities) by doing this:

$$(\mathbf{I} - \mathbf{A}_0 L)\mathbf{Z}_t = \mathbf{d} + \mathbf{u}_t$$

$$\mathbf{Z}_t = (\mathbf{I} - \mathbf{A}_0 L)^{-1}(\mathbf{d} + \mathbf{u}_t) = \sum_{i=0}^{\infty} (\mathbf{A}_0 L)^i (\mathbf{d} + \mathbf{u}_t) = (\mathbf{I} - \mathbf{A}_0)^{-1}\mathbf{d} + \sum_{i=0}^{\infty} \mathbf{A}_0^i \mathbf{u}_{t-i}$$

We can retrive to original process by premultiplying by the $n \times np$ dimensional matrix $\mathbf{J} = \begin{pmatrix} \mathbf{I}_n & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix}$ and using $\mathbf{J}'\mathbf{J}\mathbf{u}_t = \mathbf{u}_t$ and $\mathbf{J}\mathbf{u}_t = \varepsilon_t$

$$\mathbf{x}_t = \mathbf{J}\mathbf{Z}_t = \mathbf{J}(\mathbf{I} - \mathbf{A}_0)^{-1}\mathbf{d} + \sum_{i=0}^{\infty} \mathbf{J}\mathbf{A}_0^i \mathbf{J}' \varepsilon_{t-i}$$

$$= \overline{\mu} + \sum_{i=0}^{\infty} \mathbf{C}_i \varepsilon_{t-i}$$

where

$$x_t = \bar{\mu} + \sum_{i=0}^{\infty} C_i \epsilon_{t-i}$$

is called the MA-rep. of the VAR (and as always only exists if the process is stationary).

Stationarity: The roots (eigenvalues) of $\det(A_0 - \lambda A_0)$ need to lay inside the unit circle. That is the same as saying that the roots of $\det(I - \lambda A_0)$ need to lie outside the unit circel. That implies that all eigenvalues $|\lambda_i| \overset{!}{<} 1$.

Structural Model vs. Reduced Form

Motivation for the structural vs. reduced form problem:

Reduced Form: functional or stochastic mapping with inputs consisting of (i) exogenous variables and (ii) unobservables and with outputs being endogenous variables, i.e $Y = g(X, Z, U)$

Structural Form: The Structural model represents the (economic) theory and "tries" to state the data-generating process.
Let us look at an example of Supply and Demand:
Let

$$Q = D(P, X, U_D)$$
$$P = MC(Q, Z, U_S)$$

be the structural economic model of how demand an supply are generated. As we now want to actually estimate something, we solve for the equilibrium and get

$$P = p(Z, X, U_S, U_D)$$
$$Q = q(Z, X, U_s, U_D)$$

which are the reduced form relations. Generally, one can say the solving the model for the endogenous variable and possibly combining two (or more) correlated structural form terms to remove dependency yields the reduced form.

Terminology:

- e_t are the structural shocks, e.g a fiscal policy shock. They are uncorrelated and zero-mean, s.t $\text{Var}(e_t) = I_{n\times n}$
- ϵ_t are the reduced form errors, i.e the "inaccuracy" of estimation

We now have an economic (structural) model of how the world is supposed to work (maybe concerning inflation and GDP-growth). It looks as follows:

$$b_{11}^0 x_{1,t} = d - b_{12}^0 x_{2,t} + b_{11}^1 x_{1-t-1} + b_{12}^1 x_{2,t-1} + e_{1,t}$$

$$b_{22}^0 x_{2,t} = d - b_{21}^0 x_{1,t} + b_{21}^1 x_{1-t-1} + b_{22}^1 x_{2,t-1} + e_{2,t}$$

The problem here is that they have a contemporaneous impact on each other, e.g $e_{1,t} \uparrow \rightarrow x_{1,t} \uparrow \rightarrow x_{2,t} \uparrow$ and $e_{2,t} \uparrow$. We can write our eq. in matrix notation by taking everything with subscript t on the LHS:

$$B_0 x_t = d + B_1 x_{t-1} + e_t$$

By multiplying with B_0^{-1}, we solve for the endogenous variable and obtain a **reduced form**

$$x_t = B_0^{-1} d + B_0^{-1} B_1 x_{t-1} + B_0^{-1} e_t = c + A x_{t-1} + \epsilon_t$$

we are clearly especially interested in $\epsilon_t = B_0^{-1} e_t = S e_t$. It is easy to see that the reduced form error is a linear combination of the fundamental economic shocks and as $\mathbb{E}[e_t] = 0$ they are also zero-mean. They are not uncorrelated though as for example $\epsilon_{1,t} = s_{11} e_{1,t} + s_{12} e_{2,t}$

We now want to simply estimate our reduced form $x_t = c + A x_{t-1} + \epsilon_t$ via OLS which gives us:

- $\hat{c}, \hat{A}, \hat{\Sigma}$ (9 parameters)
- the structural model has 10 parameters, which means we cannot simply recover the structural model via solving the system of equations. The big problem here is the symmetry of the covariance matrix. The process of trying to return to the structural form is called identification.

(Ignoring $\hat{\cdot}$ from now on). From $\epsilon_t = S e_t$ and $\Sigma = \mathbb{E}(\epsilon_t \epsilon_t') = E(S e_t (S e_t)') = S S'$, we have:

$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} s_{11}^2 + s_{12}^2 & s_{11}s_{21} + s_2 s_{22} \\ s_{11}s_{21} + s_{12}s_{22} & s_{21}^2 + s_{22}^2 \end{pmatrix}$ That are 3 ind. equations in 4 unknowns. Formalized: There are $\frac{N(N-1)}{2}$ redundant equations.

Choleski-Factorization to solve Identifcation

If matrix A is symmetric, then $A = LL'$ with L being a lower triangular matrix and thus L' a upper triangular matrix. As there are $\frac{N(N-1}{2}$ zeros in the Choleski (here S), we have solved our problem. That is equivalent to saying that $s_{12} = 0$ (in our example), which is a so-called short-term restriction. We cant assume structural shocks to be e.g supply and demand shocks as we do not know the ordering. Explanation: We should be able to exchange x_1 and x_2 and get the same results. But as we always assume $s_{12} = 0$, the ordering of the variables is important.

Long-Run Restrictions to solve Identification

Not a restriction on the short-run (contemporaneous) but on the long run. Based on the fact that demand shocks should not have an impact on y_t in the long run. We start with $\begin{pmatrix} \Delta y_t \\ \Delta p_t \end{pmatrix} = C(L) S e_t$
We then apply the BN-Decomposition (which can be easily generalized for VARs):

$$\begin{pmatrix} y_t \\ p_t \end{pmatrix} = C(1) S \sum_{i=0}^{t} e_i + C^*(L) S e_t$$

and given that the error term $e_t = \begin{pmatrix} e_t^D \\ e_t^S \end{pmatrix}$, we need that $C(1)S = \begin{pmatrix} x_i & 0 \\ x_j & x_z \end{pmatrix}$ How do we solve for S now?
That implies that $A(1)^{-1} \sum (A(1)^{-1})' = A(1)^{-1} S S' (A(1)^{-1})' = C(1) S S' C(1)' = R R'$ with R being lower triangular and the Choleski-Factor of $A(1)^{-1} \sum (A(1)^{-1})'$ and we have:

$$C(1)S = R \rightarrow A(1)^{-1} S = R \rightarrow S = A(1) R$$

Impulse Responses for VARs

$$\phi(n,i) = IR(n,i,k) = \frac{\partial \mathbb{E}_t(x_{n,t+k})}{\partial e_{i,t}}$$

meaning the response of the n-th variable in x_t to the i-th shock.

important note: $\mathbb{E}_t[\epsilon_{t+k}] = 0 \ \forall \ k > 0$ and $\mathbb{E}_t[\epsilon_{t+k}] = \epsilon_{t+k} \ \forall \ k \leq 0$ as we observed the past and current shocks and the ones in the future are assumed to be zero-mean.
To derive to IRs, we start at the MA-rep.:

$$x_{t+k} = \sum_{i=0}^{\infty} C_i \epsilon_{t+k-i} \rightarrow \mathbb{E}_t[x_{t+k}] = \sum_{i=0}^{\infty} C_i \mathbb{E}_t[\epsilon_{t+k-i}] = \sum_{i=k}^{\infty} C_l \epsilon_{t+k-i} \tag{19}$$

as $\forall \ i < k$, we have are either in the present or future so the ϵ_is are supposed to have zero-mean. Rewriting gives:

$$= \sum_{i=k}^{\infty} C_l S e_{t+k-i} = \sum_{i=k}^{\infty} D_i e_{t+k-i} \tag{20}$$

and hence

$$\frac{\partial \mathbb{E}_t[x_{t+k}]}{\partial e_{i,t}} = D_k$$

In words: a shock in variable i which was k periods ago causes an impulse response that can be found in the i-th column of D_k. As D_k can de decomposed into components conatining S, the identifcation of S plays a huge role in identifying impulse responses. **Decomposition of D_i:** $D_i = JG^iJ'S$ with G being the companion form matrix. It is easy to see that $D_0 = S$.
Accumulated IRs Total impact over n periods: Summing over all the MA-coefficient matrices, i.e:

$$\sum_{k=o}^{\infty} D_k$$

which gives the matrix of total multipliers or the matrix of long-run effects. Multiply the matrix $D(1) = C(1)S = (I - AL)^{-1}$ with the shock, e.g $(I - AL)^{-1}(0,1)'$ gives the long run shock on x_t to a orthogonal unit shock in y_t.

Variance Decompositions

$x_{t+k} = \sum_{i=0}^{\infty} D_i e_{t+k-i} \rightarrow \mathbb{E}(x_{t+K}) = \sum_{i=k}^{\infty} C_l \epsilon_{t+k-i}$
Hence, the prediction error is $\sum_{i=0}^{k-1} D_i e_{t+k-i}$ and the variance of the predicition error $MSPE(k)$ is

$$MSPE(k) = \mathbb{E}\{[x_{t+k} - \mathbb{E}(x_{t+k})][x_{t+k} - \mathbb{E}(x_{t+k})]'\} = \sum_{i=0}^{k-1} D_i D_i'$$

where $D_p = [d_p(1)d_p(2)...d_p(N)]$ with $k \in \ \{1, ..., N\}$ Rewriting gives us

$$MSPE(k) = \sum_{p=0}^{k-1} \left(\sum_{i=1}^{N} d_p(i) d_p(i)' \right) = \sum_{i=1}^{N} \left(\sum_{p=0}^{k-1} d_p(i) d_p(i\prime \right)$$

where the term in the last bracket gives us the relative contribution of the i-th shock to the n-th endogen. variable.

Forecast Variance

If there is no intercept term (no drift), we have by forward iteration:

$$y_{t+h} = A^h y_t + \sum_{j=0}^{h-1} A^j \epsilon_{t+h-i}$$

and hence the predicition error is given by

$$y_{t+h} - \mathbb{E}(y_{t+h}) = \sum_{j=0}^{h-1} A^j \epsilon_{t+h-j} = \sum_{j=0}^{h-1} A^j S e_{t+h-j} = \sum_{j=0}^{h-1} D_j e_{t+h-j}$$

The step one ahead forecast error is thus

$$y_{t+1} - \mathbb{E}(y_{t+1}) = D_0 e_{t+1}$$

because we are interested in the first forecast error $(t+1)$ with a step of 1.

$$\text{Var}(D_0 e_{t+1})$$

$$\begin{aligned}
\text{VAR}(\mathbf{D}_0\boldsymbol{e}_{t+1}) &= E\left(\mathbf{D}_0\boldsymbol{e}_{t+1}\boldsymbol{e}_{t+1}'\mathbf{D}_0'\right) = \\
&= E\left(\begin{pmatrix}\mathbf{d}_0(1) & \mathbf{d}_0(2)\end{pmatrix}\boldsymbol{e}_{t+1}\boldsymbol{e}_{t+1}'\begin{pmatrix}\mathbf{d}_0(1)'\\ \mathbf{d}_0(2)'\end{pmatrix}\right) = \\
&= E\left(\left(\mathbf{d}_0(1)e_{1,t} + \mathbf{d}_0(2)e_{2,t}\right)\left(e_{1,t}\mathbf{d}_0(1)' + e_{2,t}\mathbf{d}_0(2)'\right)\right) = \\
&= E\left(\mathbf{d}_0(1)e_{1,t}^2\mathbf{d}_0(1)' + \mathbf{d}_0(2)e_{2,t}e_{1,t}\mathbf{d}_0(1)' + \mathbf{d}_0(1)e_{1,t}e_{2,t}\mathbf{d}_0(2)' + \mathbf{d}_0(2)e_{2,t}^2\mathbf{d}_0(
\end{aligned}$$

Using $E(e_ie_j)$ is 0 if $i \neq j$ and 1 if $i = j$:

$$\begin{aligned}
\text{VAR}(\mathbf{D}_0\boldsymbol{e}_{t+1}) &= \mathbf{d}_0(1)\mathbf{d}_0(1)' + \mathbf{d}_0(2)\mathbf{d}_0(2)' = \\
&= \begin{pmatrix}d_{1,1}^0\\ d_{2,1}^0\end{pmatrix}\begin{pmatrix}d_{1,1}^0 & d_{2,1}^0\end{pmatrix} + \begin{pmatrix}d_{1,2}^0\\ d_{2,2}^0\end{pmatrix}\begin{pmatrix}d_{1,2}^0 & d_{2,2}^0\end{pmatrix} \\
&= \begin{pmatrix}(d_{1,1}^0)^2 & d_{1,1}^0d_{2,1}^0\\ d_{1,1}^0d_{2,1}^0 & (d_{2,1}^0)^2\end{pmatrix} + \begin{pmatrix}(d_{1,2}^0)^2 & d_{1,2}^0d_{2,2}^0\\ d_{1,2}^0d_{2,2}^0 & (d_{2,2}^0)^2\end{pmatrix} = \\
&= \begin{pmatrix}(d_{1,1}^0)^2 + (d_{1,2}^0)^2 & d_{1,1}^0d_{2,1}^0 + d_{1,2}^0d_{2,2}^0\\ d_{1,1}^0d_{2,1}^0 + d_{1,2}^0d_{2,2}^0 & (d_{2,1}^0)^2 + (d_{2,2}^0)^2\end{pmatrix}
\end{aligned}$$

Hence the total one-step ahead forecast error variance of variable one is $(d_{1,1}^0)^2 + (d_{1,2}^0)^2$, the share explained by shock one is $\dfrac{(d_{1,1}^0)^2}{(d_{1,1}^0)^2 + (d_{1,2}^0)^2}$ and by the second shock it is $\dfrac{(d_{1,2}^0)^2}{(d_{1,1}^0)^2 + (d_{1,2}^0)^2}$.
Accordingly the total one-step ahead forecast error variance of variable two is $(d_{2,1}^0)^2 + (d_{2,2}^0)^2$, the share explained by shock one is $\dfrac{(d_{2,1}^0)^2}{(d_{2,1}^0)^2 + (d_{2,2}^0)^2}$ and by the second shock it is $\dfrac{(d_{2,2}^0)^2}{(d_{2,1}^0)^2 + (d_{2,2}^0)^2}$.

Can also be done in matrix notation as under *variance Decompositions*.

Two-Step ahead Forecast error:

$$y_{t+2} - \mathbb{E}(y_{t+2}) = D_0e_{t+1} + D_1e_{t+2}$$