Autoregressive distributed lags models. Vector Autoregression (VAR) models. Structural VAR.

Jakub Mućk
SGH Warsaw School of Economics

Introduction

- Time series y_t a series of observations indexed in time order, where $t = 1, 2, \ldots, T$.
- Lag operator L:

$$L(y_t) = y_{t-1}. \tag{1}$$

- Difference operator/first difference Δ:

$$\Delta(y_t) = (1 - L)\, y_t = y_t - y_{t-1}. \tag{2}$$

- Growth rates measure the percentage changes of y_t within a specific period:

$$g = \frac{y_t - y_{t-1}}{y_{t-1}}. \tag{3}$$

- Logarithmic growth rates:

$$\Delta \ln y_t = \ln y_t - \ln y_{t-1} = \ln \frac{y_t}{y_{t-1}} = \ln \frac{y_{t-1} \times (1+g)}{y_{t-1}} \approx g. \tag{4}$$

- Dynamic nature of some economic processes:

$$y_t = f(x_t, x_{t-1}, x_{t-2}, \ldots). \tag{5}$$

- Persistence/inertia of variables of interest.
- Inappropriate lag structure typically leads to serial correlation of the error term.
- Key assumption: stationary time series (i.e. non-trending variables, mean reversion).
- Models that accounts for persistence/ dynamic nature of relationship:
  - autoregressive models,
  - distributed lag models,
  - autoregressive distributed lag models,
  - error correction models (accounting for cointegration).
- In addition, multivariate models:
  - VAR models,
  - SVAR models.

Distributed lag model

Distributed lag model

- Distributed lags model of order K (denoted as DL(K)):

$$y_t = \mu + \sum_{i=0}^{K} \beta_i x_{t-i} + \varepsilon_t, \tag{6}$$

where
  - y_t – outcome/ dependent variable,
  - x_t – explanatory variable,
  - ε_t – the error term.

- **Short-run multiplier (β^{SR}):**

$$\beta^{SR} = \beta_0. \tag{7}$$

- **Long-run multiplier (β^{LR}):**

$$\beta^{LR} = \beta_0 + \beta_1 + \ldots + \beta_K. \tag{8}$$

- The parameters of equation (6) can be estimated with the least squares estimator.

Autoregressive model

- **Autoregressive model of order 1 (denoted as AR(1))**

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t \tag{9}$$

where ε_t is the error term and $\varepsilon_t \sim \mathcal{N}(0, \sigma)$.

- **Key assumption:** $|\rho| < 1$
- The parameter ρ measures the persistence/ inertia of y_t.
  - If ρ is close to 0 then effect of exogenous disturbances (measured by ε_t) is almost immediately absorbed.
  - If ρ is close to 1 then effect of exogenous slowly dies out.
- Selected properties of y_t when it follows AR(1) process:

$$\mathbb{E}(y_t) = \frac{\mu}{1-\rho}, \tag{10}$$

$$Var(y_t) = \frac{\sigma^2}{1-\rho^2}. \tag{11}$$

- Half-life:

$$hl = \frac{\ln(0.5)}{\ln(\rho)}. \tag{12}$$

- What is effect of the error term (ε_t) on dependent variable?
- Consider the simplified case $(\mu = 0)$ of AR(1) model:

$$y_t = \rho y_{t-1} + \varepsilon_t, \tag{13}$$

and assume that $\varepsilon_0 = 1$ and for $t > 1$, $\varepsilon_t = 0$. Then:

$$\begin{aligned} y_0 &= 0 \times \rho + 1 = 1 \\ y_1 &= y_0 \times \rho + 0 = 1 = \rho \\ y_2 &= y_1 \times \rho + 0 = \rho\rho = \rho^2 \\ &\dots \end{aligned}$$

or more generally:

$$y_t = \rho^t. \tag{14}$$

- Taking into account the fact that ε_0 is 0 in above example, the AR(1) model can be rewritten into moving-average representation:

$$y_t = \sum_{i=1}^{\infty} \rho^i \varepsilon_{t-i} + \varepsilon_t = \varepsilon_t + \sum_{i=1}^{\infty} \phi_i \varepsilon_{t-i}. \tag{15}$$

AR(1) and IRF II

- The moving average representation illustrates how the outcome variable reacts to the some exogenous disturbances over the time:

$$\frac{\partial \mathbb{E}(y_t)}{\partial \varepsilon_{t-i}} = \phi_i = \rho^i. \tag{16}$$

- **Impulse response function** describes the expected evolution of the outcome variable to a unit shock.

$$\{1, \phi_1, \phi_2, \ldots\}. \tag{17}$$

For the variable that follows AR(1):

$$\left\{1, \rho, \rho^2, \ldots\right\}. \tag{18}$$

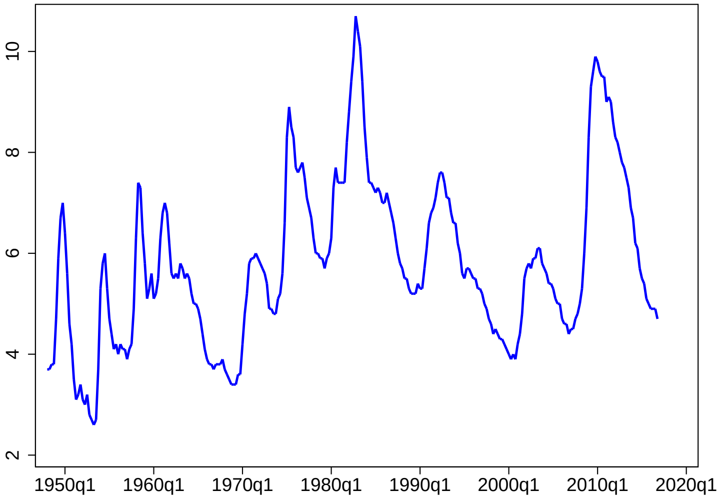
- **Autoregressive model of order P (denoted as AR(P))**

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \ldots + \rho_P y_{t-P} + \varepsilon_t. \tag{19}$$

- The parameters can be still estimated with least squares.
- The AR(P) models
  - are useful in studying complex dynamic properties of variable of interest,
  - are useful in forecasting.

Empirical example: unemployment rate U_t in the US

SGH



Source: FRED

- Estimates of AR(1) model

$$\hat{U}_t = \underset{(0.087)}{0.184} + \underset{(0.014)}{0.969} U_{t-1} \tag{20}$$

substantial/extreme persistence.
But: serially correlated residuals (correlation between residuals and its lag ≈ 0.66).

- Estimates of AR(2) model

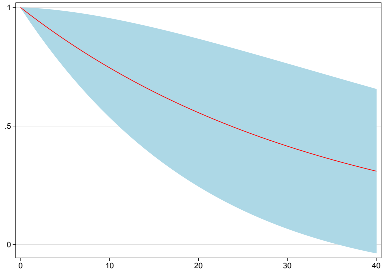
$$\hat{U}_t = \underset{(0.066)}{0.285} + \underset{(0.045)}{1.613} U_{t-1} - \underset{(0.045)}{0.661} U_{t-2} \tag{21}$$

How does unemployment rate react to exogenous shocks?

AR(1) model:

$$U_t = \underset{(0.087)}{0.184} + \underset{(0.014)}{0.969} U_{t-1}$$

Impulse response function:

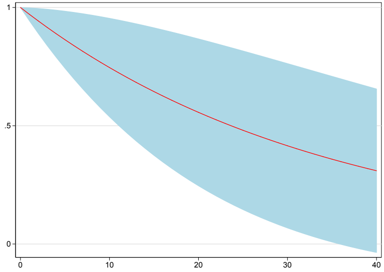


Empirical example: unemployment rate U_t in the US

Impulse response function:

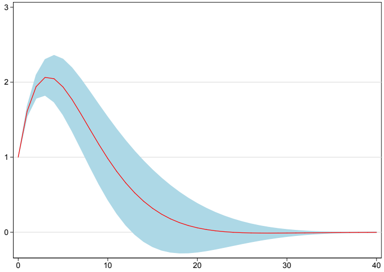
AR(1) model:

$$U_t = \underset{(0.087)}{0.184} + \underset{(0.014)}{0.969} U_{t-1}$$



AR(2) model:

$$U_t = \underset{(0.066)}{0.285} + \underset{(0.045)}{1.613} U_{t-1} - \underset{(0.045)}{0.661} U_{t-1}$$



Autoregressive distributed lag model

- **Autoregressive distributed lag model ADL(1,0)**:

$$y_t = \mu + \rho y_{t-1} + \beta_0 x_t + \varepsilon_t, \tag{22}$$

when $|\rho| < 1$.

- Assume that $y_0 = 0$, $\mu = 0$ and $\varepsilon_t = 0$ and consider a unit change in x at the period 0. Then,

$$\begin{aligned} y_0 &= 0 \times \rho + \beta_0 \times 1 + 0 = \beta_0, \\ y_1 &= \beta_0 \times \rho + \beta_0 \times 0 + 0 = \rho\beta_0, \\ y_2 &= \rho\beta_0 \times \rho + \beta_0 \times 0 + 0 = \rho^2\beta_0, \end{aligned}$$

more generally

$$y_t = \rho^t \beta_0.$$

- Short-run multiplier: β_0.
- Impulse response function for ith period:

$$\frac{\partial \mathbb{E}(y_t)}{\partial x_{t-i}} = \rho^i \beta_0.$$

- Cumulative response function for ith period:

$$\sum_{j=0}^{i} \frac{\partial \mathbb{E}\left(y_t\right)}{\partial x_{t-j}} = \sum_{j=0}^{i} \rho^j \beta_0 = \beta_0 + \beta_0 \rho + \beta_0 \rho^2 + \ldots + \beta_0 \rho^j .$$

- The long-run multiplier:

$$\sum_{j=0}^{\infty} \frac{\partial \mathbb{E}\left(y_t\right)}{\partial x_{t-j}} = \beta_0 \left(1 + \rho + \rho^2 + \ldots\right) = \frac{\beta_0}{1-\rho} .$$

- **Autoregressive distributed lag model ADL(P,K)**:

$$y_t = \mu + \sum_{i=1}^{P} \rho_i y_{t-i} + \sum_{i=0}^{K} \beta_i x_{t-i} + \varepsilon_t. \tag{23}$$

- **Short-run multiplier (β^{SR}):**

$$\beta^{SR} = \beta_0. \tag{24}$$

- **Long-run multiplier (β^{LR}):**

$$\beta^{LR} = \frac{\beta_0 + \beta_1 + \ldots + \beta_K}{1 - \rho_1 - \rho_2 - \ldots - \rho_P} = \frac{\sum_{i=0}^{K} \beta_i}{1 - \sum_{i=1}^{P} \rho_i}. \tag{25}$$

Choosing order of AR/DL/ADL models

The trade-off between:

- Risk of omitting important variables (when P and/or K are small).
- Efficiency (when P and/or K are large).

The most popular strategies:

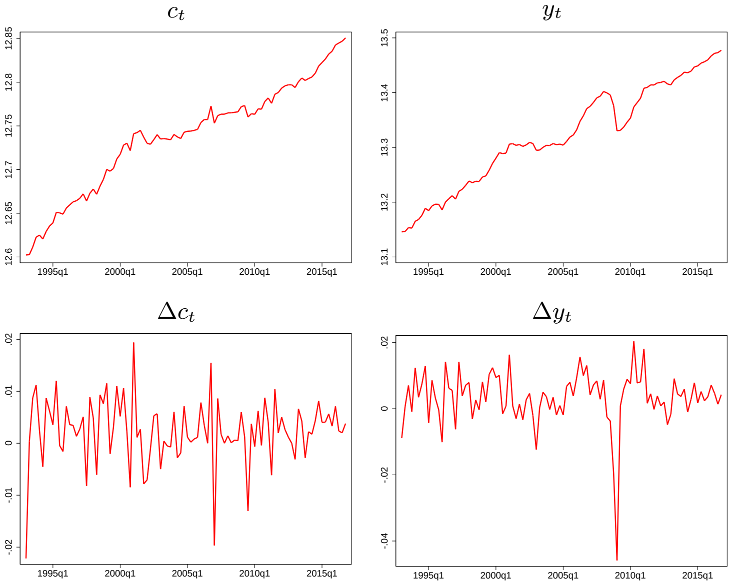
- From general to specific
- From specific to general

Selection criteria:

- Serial correlation of residuals,
- Information criteria.
- Significance.

- Data: time series from 1993Q1to 2016Q.
- **Dependent variable:**
  c_t - the logged real consumption expenditures (in constant prices.).
- **Explanatory variable:**
  y_t - the logged real GDP (in constant prices).

Empirical example: consumption function for Germany



The DL models:

$$\Delta c_t = \alpha_0 + \sum_{i=0}^{K} \beta_i \Delta y_{t-i} + \varepsilon_t, \tag{26}$$

K	0	1	2
μ	0.002	0.002	0.002
	(0.001)	(0.001)	(0.001)
β_0	0.278	0.328	0.295
	(0.092)	(0.096)	(0.087)
β_1		-0.135	-0.200
		(0.096)	(0.091)
β_2			0.142
			(0.087)
β^{LR}	0.278	0.193	0.236
	(0.092)	(0.114)	(0.120)
BIC	-705.547	-696.699	-705.430

The ADL models:

$$\Delta c_t = \mu + \sum_{j=1}^{P} \rho_j \Delta c_{t-j} + \sum_{i=0}^{K} \beta_i \Delta y_{t-i} + \varepsilon_t, \tag{27}$$

K	0	1	2	0
P	0	0	0	1
μ	0.002	0.002	0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
ρ_1				-0.290
				(0.073)
β_0	0.278	0.328	0.295	0.204
	(0.092)	(0.096)	(0.087)	(0.074)
β_1		-0.135	-0.200	
		(0.096)	(0.091)	
β_2			0.142	
			(0.087)	
β^{LR}	0.278	0.193	0.236	0.205
	(0.092)	(0.114)	(0.120)	(0.068)
BIC	-705.547	-696.699	-705.430	-707.569

Error correction model

- If x_t and y_t are cointegrated then

$$e_t = y_t - \beta_0 - \beta_1 x_t \tag{28}$$

the residuals, e_t, measure deviation from a common stochastic trend (or long-run equilibrium between variables).

- **Long-run elasticity** equals β_1 in (28).
- **[Error correction model]** In the **short-run** dynamics the deviation from long-run relationship between variables can be taken into account by using **the lagged residuals from the long-run equation, i.e.** e_{t-1}.

$$\Delta y_t = \mu + \delta e_{t-1} + \sum_{i=1}^{P} \rho_i \Delta y_{t-i} + \sum_{i=0}^{K} \beta_i \Delta x_{t-i} + \varepsilon_t, \tag{29}$$

where parameter δ measures the pace of adjustment toward long-run equilibrium and $\delta \in (-1, 0)$.

- Half-life:

$$hl = \frac{\ln(0.5)}{\ln(1+\delta)}. \tag{30}$$

- Alternatively, one might directly account for **an adjustment toward the long-run relationship**.
- This can be captured by replacing the lagged residuals, i.e., e_{t-1}, in (29) by the variables in levels, i.e. y_{t-1} and x_{t-1}:

$$\Delta y_t = \mu + \phi_1 y_{t-1} + \phi_2 x_{t-1} + \sum_{i=1}^{P} \rho_i \Delta y_{t-i} + \sum_{i=0}^{K} \beta_i \Delta x_{t-i} + \varepsilon_t, \qquad (31)$$

where:

- ϕ_1 measures the pace of adjustment toward long-run equilibrium and if variables of interest are cointegrated then $\phi_1 \in (-1, 0)$
- The long-run elasticity of y_t with respect to changes in x_t: $-\phi_2/\phi_1$.

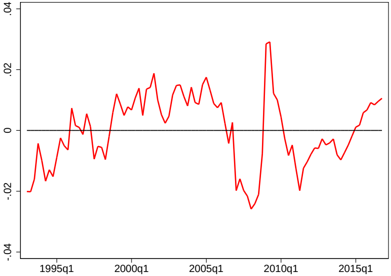
Both c_t and y_t are integrated in order order 1.

Long-run relationship:

$$\hat{c}_t = \underset{(0.174)}{3.984} + \underset{(0.013)}{0.657} y_t \tag{32}$$

What is the long-run elasticity?

Residuals from the long-run regression



The ADF statistics: -3.52.

Error Correction Model (simplified version):

$$\Delta\hat{c}_t = \underset{(0.050)}{-0.153} ec_{t-1} + \underset{(0.066)}{0.283}\Delta y_t \tag{33}$$

where ec_{t-1} is the error correction, i.e., lagged residuals from regression for variables in levels.

Estimated parameter δ that describe pace of adjustment to the long-run equilibrium is statistically significant and negative.

half-life: ≈ 4.13 quarters ($4.13 \approx \ln(0.5)/\ln(1-0.153)$).

Vector Autoregression (VAR) models

- **Vector Autoregression (VAR) models** are atheoretical/agnostic multivariate models that allow to capture dynamic properties of several variables with an interplay between them.
- VAR(2,1) model:

$$\mathbf{y}_t = \mathcal{A}_0 + \mathcal{A}_1 \mathbf{y}_{t-1} + \varepsilon_t, \tag{34}$$

where $\mathbf{y}_t$ is the vector of two endogenous variables, ε_t is the vector of the error terms, $\mathcal{A}_0$ is the vector of constants and $\mathcal{A}_1$ is the matrix of coefficients that captures dynamic relationship.
- In the standard VAR model:

$$\varepsilon_t \sim \mathcal{N}(0, \Sigma),$$

where Σ is the symmetric but possibly not diagonal matrix.
- VAR(2,1) model in matrix form:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}. \tag{35}$$

- VAR(K,P) model:

$$\mathbf{y}_t = \mathcal{A}_0 + \mathcal{A}_1 \mathbf{y}_{t-1} + \ldots + \mathcal{A}_P \mathbf{y}_{t-P} + \varepsilon_t, \tag{36}$$

where P is the number of lags while K is the number of endogenous variables.

- VARX(K,P) model:

$$\mathbf{y}_t = \mathcal{A}_0 + \mathcal{A}_1 \mathbf{y}_{t-1} + \ldots + \mathcal{A}_P \mathbf{y}_{t-P} + \mathcal{D} x_t + \varepsilon_t, \tag{37}$$

where P is the number of lags while K is the number of endogenous variables and x_t is the vector of exogenous variables.

- General notation:

$$\mathcal{A}(L) \mathbf{y}_t = \varepsilon_t, \tag{38}$$

where $\mathcal{A}(L) = \left(I - \mathcal{A}_0 - \mathcal{A}_1 L - \ldots - \mathcal{A}_P L^P\right)$

Impulse response function I

- Key assumption is **stability**.The VAR model is stable if:
  - VAR(K,1): $det(I - \mathcal{A}_1 z) \neq 0$ and $|z| < 1$.
  - VAR(K,P): $det(I - \mathcal{A}_1 z - \ldots - \mathcal{A}_P z^P) \neq 0$ and $|z| < 1$.

  or if eigenvalues of below matrix are below zero:

$$\begin{bmatrix} \mathcal{A}_1 & \ldots & \ldots & \mathcal{A}_P \\ I & 0 & \ldots & 0 \\ 0 & I & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & I & 0 \end{bmatrix}. \tag{39}$$

- **Stability implies stationarity**.
- If VAR model is stable then, according to the Wold theorem, it has infinite moving average MA(∞) representation:

$$\mathbf{y}_t = \mathcal{A}(L)^{-1} \varepsilon_t = \mathcal{C}(L) \varepsilon_t = \sum_{i=0}^{\infty} \mathcal{C}_i \varepsilon_{t-i}. \tag{40}$$

- The moving average representation is useful tool in an investigation of dynamic reaction of endogenous variables to some disturbances. In particular, **impulse response function** measures dynamic effects of variables to a change in the error term in given equation:

$$IRF_{i,j,h} = \frac{\partial y_{i,t+h}}{\partial \varepsilon_{j,t}}, \tag{41}$$

  and, since VAR model is stable, $IRF_{i,j,h}$ tends to 0 as $h \to \infty$. Alternative measure is **cumulative impulse response function**:

$$CIRF_{i,j,H} = \sum_{h=0}^{H} \frac{\partial y_{i,t+h}}{\partial \varepsilon_{j,t}}. \tag{42}$$

- **Forecast variance error decomposition:** shows the share of disturbances in overall forecasting error in given horizon.

- The VAR models can be estimated consistently equation-by-equation using ordinary least squares estimator.
- Key problem: the lag length. Natural trade-off
  - **[High lag length]:** a loss in efficiency but smaller risk of omitting important lags (endogenity).
  - **[Low lag length]:** a high risk of omitting important lags but smaller loss in efficiency.
- Criteria for selecting lag length:
  - **Serial correlation**.
  - Information criteria.

Structural VAR

- **VAR models are atheoretical**. One of the assumption is that the error terms can be correlated between equations:

$$\varepsilon_t \sim \mathcal{N}\left(0, \Sigma\right),$$

where Σ is the non-diagonal matrix.

- In the **Structural VAR models** the errors are not correlated and, therefore, can be interpreted as structural shocks:

$$\mathbf{u}_t \sim \mathcal{N}\left(0, I\right),$$

where I is the identity matrix.

- Using non-sample information (theory) is essential in order to identify structural shocks.

Popular schemes of the structural shocks identification

- Cholesky's decomposition
- AB-model/short-run restrictions
- Long-run restrictions
- Sign restrictions
- Identification through heteroskedasticity

- Let us rewrite the variance of the errors:

$$\Sigma = PP', \tag{43}$$

where P is the lower triangular matrix.

- Now, define the new error which is transformation of the error from the reduced form:

$$u_t = P^{-1}\varepsilon_t, \tag{44}$$

and clearly

$$var(u_t) = P^{-1}var(\varepsilon_t)(P')^{-1} = P^{-1}P(P')^{-1}P' = I. \tag{45}$$

- The orthogonalized IRF (at given time horizon):

$$OIRF_{.,.,h} = P^{-1} \times IRF_{.,.,h}. \tag{46}$$

- The ordering of variables matters.

- **The AB-model:**

$$A\varepsilon_t = Bu_t, \tag{47}$$

where A is the matrix describing contemporaneous relationship between endogenous variables while matrix B measures the short-run impact of structural shock on variables, u_t is the vector of structural shocks.

- **Key problem: identification problem**. The AB model can be rewritten as:

$$\varepsilon_t = A^{-1}Bu_t, \tag{48}$$

and it could be shown that there are needed $K(K+1)/2$ restrictions while K is the number of parameters and the total number of parameters is $2K^2$.

- The parameters can be estimated with the ML estimator while possible overidentifying restriction can be tested with LR test.

- The long-run impact (matrix C):

$$C = \lim_{h\to\infty} CIRF_{,,h} = \sum_{h=0}^{\infty} IRF_{.,.,h} = \sum_{h=0}^{\infty} \mathcal{C}_h. \tag{49}$$

where $\mathcal{C}_h$ is the matrix from the moving average representation.

- The identification of structural shocks can base on the long-run neutrality. In other words, this translates into zero restrictions on elements of matrix C.