Multivariate Time Series Models

Motivation

- Natural extension of univariate TS models
- The most general class VECTOR ARMA model: VARMA
- Each variable explained by: own lags and errors, other variable's lags
- A very general representation of joint dynamics of analysed variables
- Little to no restrictions on parameters hence, in short samples, they can overfit and perform poorly in forecasting
- Since multivariate MA components are difficult to estimate practitioners mostly work with "VAR model". (remember, for univariate series an AR model with sufficient lag length can approximate ARMA model well)
- Further reading: **Lutkepohl (2007)**, Hamilton (1994) and for forecasting Clements and Hendry (1998)

Topics covered

- Representation
- Estimation
- Diagnostic Checking
- Forecasting
- Impulse Response Analysis

Representation

- Assume y_t stands for $m \times 1$ vector of m time series:

$$y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{bmatrix}$$

- VAR model with p endogenous lags, VAR(p):

$$\underbrace{y_t}_{m\times 1} = \underbrace{\mu}_{m\times 1} + \underbrace{\Phi_1}_{m\times m} y_{t-1} + \cdots + \underbrace{\Phi_p}_{m\times m} y_{t-p} + \underbrace{\epsilon_t}_{m\times 1} \; ; \qquad \epsilon_t \sim WN(\underbrace{0}_{m\times 1}, \underbrace{\Sigma}_{m\times m})$$

p intercepts   p own and other variables lags   p error terms

Each error term is zero mean and homoscedastic. However, they can be cross correlated (non-zero off diagonal elements in Σ).

Number of parameters: $m\ (const.) + p \times m^2 (ar\ mat.) + \frac{m(m+1)}{2}\ (varcov.mat)$

Example: AR(1), no constant, m=3

- *full exposition*

$$y_{1t} = \phi_{11}y_{1t-1} + \phi_{12}y_{2t-1} + \phi_{13}y_{3t-1} + \epsilon_{1t}$$
$$y_{2t} = \phi_{21}y_{1t-1} + \phi_{22}y_{2t-1} + \phi_{23}y_{3t-1} + \epsilon_{2t}$$
$$y_{3t} = \phi_{31}y_{1t-1} + \phi_{32}y_{2t-1} + \phi_{33}y_{3t-1} + \epsilon_{3t}$$

- *matrix exposition*

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix}$$

- *compact form*

$$y_t = \Phi_1 y_{t-1} + \epsilon_t$$
$$(1 - \Phi_1 L)y_t = \epsilon_t$$
$$\Phi(L) = \epsilon_t$$

- White noise reisudals

$$\begin{bmatrix}\epsilon_{1t}\\ \epsilon_{2t}\\ \epsilon_{3t}\end{bmatrix} \sim WN\left\{\begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{bmatrix}\right\}$$

- Can be cross-correlated(e.g. $E(\epsilon_{1t}\epsilon_{2t}) = \sigma_{12} \neq 0$) but not serialy correlated (e.g. $E(\epsilon_{1t}\epsilon_{1t-1}) = E(\epsilon_{1t}\epsilon_{2t-1}) = 0$).
- As before, we assume weak stationarity

$E(y_t) = \mu$ - stable long-run mean

$Var(y_t) = V < \infty$ - finite covariance matrix

$Cov(y_t, y_{t+k}) = V_k$; - V_k does not depend on t

Finite 1st and 2nd moments that are stable across time.

- WS condition analogue: roots of $\det(1 - \Phi z) = 0$ are larger than 1 in absolute value. (Can be relaxed under certain assumptions.) Equivalent: If all eigenvalues of Φ have modulus less than 1.
- Justification (multivariate analogue): Any WS y_t can be approximated with $y_t = C(L)\epsilon_t$ (Wold decomposition theorem), which can be approximated with $A(L)y_t = B(L)\epsilon_t$, which can be approximated with $\Phi(L)y_t = \epsilon_t$ ("VAR(p) can approximate any WS vector of processes with p sufficiently large").

Model Specification

- We assume y_t is weakly stationary (d in $I(d)$ is 0)
- We need to select p in $VAR(p)$
- Due to a large number of parameters in VARs it is desirable to keep p as small as possible without risking model's validity.
- Approaches to selection of p:
  - Multivariate variants of ACF and PACF. [almost never used in practice]
  - Top down approach: start with high p and remove lags according to the results of statistical test [sometimes used in practice]
  - Use multivariate versions of Information Criteria [most often used in practice]

Multivariate IC

- Select a high p_{max} and estimate ICs for all lags $(1, \dots, p_{max})$:

$$AIC(j) = \ln\left|\Sigma_j\right| + \frac{2}{T} j m^2, \qquad where\ j = 1, \dots, p_{max}$$

$$BIC(j) = \ln\left|\Sigma_j\right| + \frac{\ln T}{T} j m^2,$$

m – VAR dimmension

- Select p that minimizes desired IC. (BIC is consistent, AIC is not)
- For $T > 8$ BIC has a larger penalty. It will tend to select more parsimonious models.
- You can also use statistical tests. E.g. If y_{1t} is not significant in non- y_{1t} equations, simply drop it.

Estimation

- In principle we should use “system” estimation method that takes into account that VAR errors are cross-correlated.
- However, it can be shown that ML estimator of VAR parameters matrix Φ coincides with equation-by-equation OLS estimator.
- Equation-by-equation OLS estimator is equivalent to the OLS based system estimator which is called SUR estimator (seemingly unrelated regression).
- ML error variance estimator is consistent but biased in small samples. Use corrected estimator (pre-programed in stat. software)

Diagnostics

- We have assumed that errors are uncorrelated and homoscedastic.
- We can use the same tests as we used in the univariate case, equation by equation.
- We can instead use the multivariate versions of those tests that we employed in the univariate case.
- No autocorrelation: Multivariate Breusch-Godfrey LM test.
- Heteroscedasticity: Multivariate White test.
- Normality: Multivariate JB-test.
- Parameter stability: Chow test for breaks, recursive estimation,…
- If test we fail we reconsider the model: increase p, add trend, add (dummy) variables, modify sample, transform variables,…

Forecasting

- Optimal (MSFE error) forecast is of the same form as in the univariate case (no parameter uncertainty):

$$\hat{y}_{t+h} = \Phi_1 \hat{y}_{t+h-1} + \cdots + \Phi_p \hat{y}_{t+h-1}$$

- Example with VAR(1) iterated forecast (no constant, parameters are estimated):

$$\hat{y}_{T+1} = \mathrm{A}_1 y_T$$
$$\hat{y}_{T+2} = \mathrm{A}_1 \hat{y}_{T+1}$$
$$\vdots$$
$$\hat{y}_{T+h} = \mathrm{A}_1 \hat{y}_{T+h-1} = \mathrm{A}_1^h y_T$$

Forecasting cont.

- As in the univariate case, one can also estimate the direct forecast. Estimate:
$$y_t = \Phi_1 y_{t-h} + \cdots + \Phi_p \hat{y}_{t-h-p} + u_t$$
- And forecast:
$$\tilde{y}_{T+h} = A_1 y_T + \cdots + A_p y_{T-p}$$
- As in the univariate case, if the model is correctly specified $\hat{y}_{T+1}$ is more efficient. $\tilde{y}_{T+h}$ can be more robust if the model is miss-specified.

Granger causality (should a variable be included in a VAR)

- Caution: The names is miss-leading. Granger causality does not imply causality. It only implies that one variable is useful for forecasting another variable once we control for past dynamics (own lags).
- Assume we have x_t $(T \times 1)$ and y_t $(T \times 1)$ and estimate:

$$\begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{xt} \end{bmatrix}$$

$\phi(L) = 0$ - "x_t does not Granger-cause y_t "

$\phi(L) \neq 0$ - "x_t Granger-causes y_t", x_t can be used to improve forecast for y_t

Why does Granger-causality not imply causality? Because even if $\phi_{12}(L) = 0$ (x_t lag polynomial does not load into the equation for y_t), ϵ_{xt} might be causing movements in y_t. Also, ϵ_{xt} is not known at the time of the forecast. Causal analysis, on the other hand, is typically done in-sample so ϵ_{xt} is known.

Forecast errors and variances

- Intuition is the same as in the univariate case.
- Analogous expressions are complex so we omit them.
- A simple shortcut to constructing error bands around the forecast is to observe (VAR transformed into VMA):

$\hat{y}_{T+h} = \hat{y}_{t+h|T-1} + C_h(y_T - y_{T|T-1})$ eq. (1)

Where C_h is the coefficient from regressing y_{T+h} on in-sample errors ($y_{T+h} = C_h \epsilon_T + \sum_{j=0}^{\infty} C_{j+h+1} \epsilon_{T-1-j}$)

- Eq. 1 shows that optimal forecast for $T+h$ made in period T can be obtained by adding optimal forecast for $T+h$ made in $T-1$ and the one step ahead forecast error made in forecasting y_T in $T-1$.
- For details regarding constructing forecasts and confidence bands for VARs see Lutkepohl (2007, Chapter 3). Statistical software constructs appropriate bands by default.

VAR in »companion« form

- In what follows we work with a VAR(1). This simplifies the derivations.
- This does not limit our analysis to VAR(1) models because a VAR(p) can always be transformed into a VAR(1) model.
- VAR(p) written as a VAR(1) model is called the companion form.
- Statistical software does all its calculations in the companion form.
- Example on the board: VAR(2) -> VAR(1)

Impulse response analysis

- VAR can be expressed as an MA(∞) model:

$$y_t = \Phi^{-1}(L)\epsilon_t, \qquad \epsilon_t \sim WN(0, \Sigma)$$

- Note that this implies that y_t is a dynamic system driven by residuals:

$$y_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_1 \epsilon_{t-2} + \cdots \quad (3)$$

- Therefore we can analyze how a shock to ϵ_{1t} (say unexpected change in the interest rate) affects all variables in y_t (say interest rate, rGDP and inflation) over time.
- Since Σ positive definite (symmetric with positive eigenvalues) therefore there exists P such that $\mathrm{P}\Sigma\mathrm{P}' = I$. We can therefore rewrite 3 as:

Simplified example on the board: 1) VAR(1)+unit schock. 2) Equality with the MA representation. 3) Is unit shock ok?

$$y_t = \Phi^{-1}(L)P^{-1}P\,\epsilon_t = \Psi(L)v_t \qquad (4)$$

Where $v_t = P\epsilon_t$ with $E(v_t) = 0$ and $E(v_t v_t') = P\Sigma P' = I$.

- This is typically called a "structural" VAR representation. v_t are called "structural or deep shocks" (demand, supply, monetary policy shocks,… see also Ramey (2016)) because they are orthogonal (deep drivers that do not depend on anything else). Matrix P determines how structural shocks affect variables in y_t.
- To arrive at (4) we therefore need to estimate P. The most simple method is to assume $P = chol(\Sigma)$ since we know that $chol(\Sigma) \times chol(\Sigma) = \Sigma$. However, note that by Cholesky definition P is lower triangular and therefore determines which variables are affected by which "deep" shock:

$$\begin{bmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{bmatrix} = \Phi^{-1}(L)P^{-1} \begin{bmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{2t} \end{bmatrix} \quad (5)$$

- A shock is a vector with all elements except one equal to zero:

$$v_{1t+1} = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} \cdots v_{mt+1} = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$$

- Therefore, to inspect how variables respond to a specific deep shock we simply take (5) and insert desired deep shock:

$$\begin{bmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{bmatrix} = \Phi(L)P^{-1} \begin{bmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} 0 \\ v_{kt} \\ 0 \end{bmatrix} \quad (6)$$

Response of y_t to shock v_{kt} in period t.

Sum of infinite geometric series:

$$\frac{1}{1-r} = \sum_{n=1}^{\infty} r^{n-1}$$

$$\boldsymbol{y_t = \Phi^{-1}(L)P^{-1}P\,\epsilon_t = \Psi(L)v_t}$$

- To investigate the response of the system in time we simply inspect (6):

$$\frac{\partial y_{t+1}}{\partial v_{t+1}} = P^{-1} = \Psi_1$$

$$\frac{\partial y_{t+2}}{\partial v_{t+1}} = \Phi_1 P^{-1} = -\Psi_2$$

$$\frac{\partial y_{t+2}}{\partial v_{t+1}} = \Phi_2 P^{-1} = -\Psi_3$$

How to estimate P

- To arrive at (4) we need to estimate P.
- The most simple method is to assume $P = chol(\Sigma)$ since we know that $chol(\Sigma) \times chol(\Sigma) = \Sigma$. However, note that by Cholesky definition P is lower triangular and therefore determines which variables are affected by which "deep" shock:

$$\begin{bmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{bmatrix} = \Phi(L)P^{-1} \begin{bmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{2t} \end{bmatrix} \quad (5)$$

- In principle, any invertible matrix that satisfies $P\Sigma P' = I$ will do. And there are many (different decompositions, rotations of matrices, etc…). Typically we use economic theory restrictions when selecting which type of matrix P to choose from.
- Note that one can re-order variables in y_t and he/she will get a slightly different response of the system to deep shocks. Therefore, one needs to be cautious.
- There are many way to handle these issues (generalized impulse responses, sign restrictions, magnitude restrictions, exclusion restrictions, short and long term restrictions,…) which we do not further inspect here.

How to estimate P

- Choice of P, that map residuals (ϵ_t) into structural/deep shocks (v_t), is not unique.
- Why? Σ has $m(m+1)/2$ unique elements (since it is symmetric), while P contains m^2 elements. Therefore we need to impose at least $\frac{m^2-m}{2}$ restrictions to uniquely define P. Cholesky decomposition of Σ produces a lower triangular matrix with $\frac{m^2-m}{2}$ (number of restrictions needed!) above diagonal elements set to 0:

$$P = \begin{bmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \qquad \left(\frac{m^2-m}{2} = \frac{m^2-m}{2} = 3\right)$$

Cont.

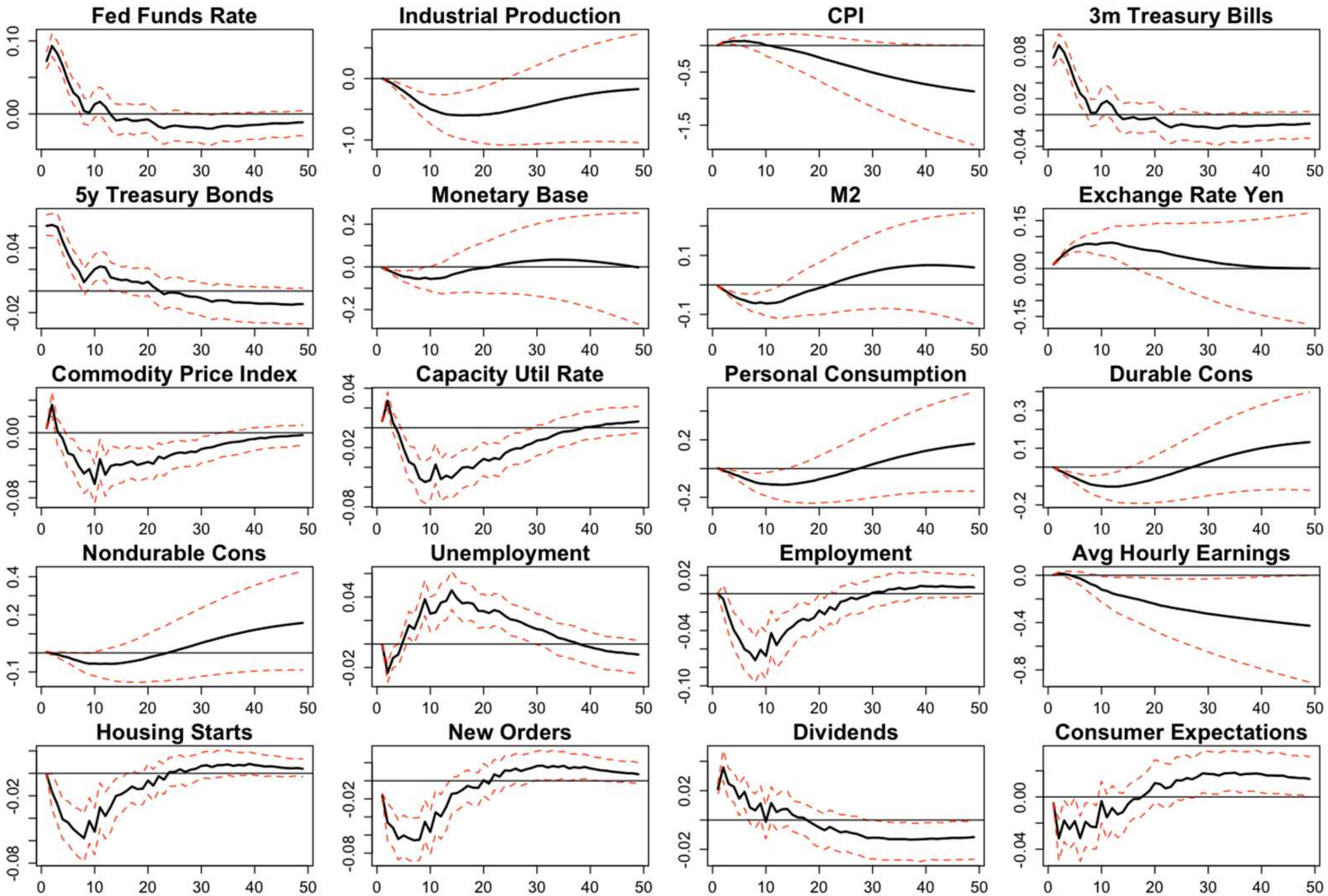
- Such P matrix is said to impose identification via "exclusion or short-run restrictions." Why?

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \Phi(L)P^{-1}\begin{bmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix}\begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix}$$

- Because deep shock v_{1t} affects/moves y_{1t}, y_{2t} and y_{3t} in period t. However, deep shock v_{2t} impacts only y_{2t} and y_{3t} but not y_{1t} in t. And deep shock v_{3t} only impacts y_{3t}!
- One needs to use economic/financial theory to justify such restrictions. E.g. if y_{3t} is central bank interest rate and y_{1t} and y_{2t} inflation and rGDP, one can reasonable assume that unexpected shocks to the interest rate (v_{3t}) in period t does not affect rGDP and inflation in the same period (it takes time for the banks to start issuing loans at the new interest rate and for those loans to affect production and consumer behaviour).

- Empirically, once the form of P is chosen we can estimate it from $\hat{\Sigma}$ and since we also estimated AR parameters $\widehat{\Phi}$ we can easily calculate $\widehat{\Psi}$ and the related "impulse" response of the system to deep shock $\mathrm{v}_{k,t}$.
- Confidence intervals can be derived analytically (see Lutkepohl (2007)). Alternatively we can use bootstrap methods to obtain them.
- Typically, impulse responses are presented graphically in a multi-panel graph.

Example: Replica of Boivin, Bernanke and Eliasz (2005, QJE). Response of economy to a monetary policy shock (deep, structural shock) that increases FFR.



Source: https://colab.research.google.com/github/jbduarte/blog/blob/master/_notebooks/2020-04-24-FAVAR-Replication.ipynb#scrollTo=qrLqN6XNURT3

Example: Simple «monetary» VAR

- Similar albeit simpler model whose IRs we saw on the previous slide.

$$y_t = \begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix} = \begin{bmatrix} inflation \\ rGDP \\ interest\ rate \end{bmatrix}$$

- Assume that the estimated model is VAR(1):

$$y_t = A\, y_{t-1} + \epsilon_t; \qquad \epsilon_t \sim WN(0, \Sigma)$$

$$\begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ y_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix}$$

- ASSUMPTION: monetary policy shock = unexpected (=not explained by past values of variables) change in the interest rate which changes the interest rate on impact (=in the period in which shock occurs) but does not affect other variables in the period in which it occurs.
- Given the above assumptions on how a monetary policy shock affects the economy we can use Cholesky decomposition to identify P. Why?

$$\begin{bmatrix} \pi_t \\ rGDP_t \\ r_t \end{bmatrix} = \mathrm{A}(L)P^{-1} \begin{bmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_t^{mon.pol.} \end{bmatrix}$$

- We estimated $\{\hat{A}, \hat{\Sigma}\}$ via OLS and $\hat{P} = chol(\hat{\Sigma})$.
- We can now re-write the model into the VMA form:

$$y_t = \hat{A} y_{t-1} + \hat{\epsilon}_t$$
$$y_t = \hat{A} y_{t-1} + P^{-1} \hat{v}_t$$
$$(I - \hat{A}L) y_t = P^{-1} \hat{v}_t$$
$$y_t = \left(I - \hat{A}L\right)^{-1} P^{-1} \hat{v}_t$$
$$y_t = \left[\left(\hat{A}L\right)^0 + \left(\hat{A}L\right)^1 + \left(\hat{A}L\right)^2 + \cdots\right] P^{-1} \hat{v}_t$$
$$y_t = \left[I + \hat{A}L + \hat{A}^2 L^2 + \cdots\right] P^{-1} \hat{v}_t$$

Sum of infinite geometric series:

$$\frac{a}{1-r} = \sum_{n=1}^{\infty} a r^{n-1}$$

- To arrive at:

$$y_t = P^{-1}\hat{v}_t + \hat{A} P^{-1} \hat{v}_{t-1} + \hat{A}^2 P^{-1} \hat{v}_{t-2} + \cdots$$
$$= \hat{v}_t + \hat{\Psi}_1 \hat{v}_{t-1} + \hat{\Psi}_2 \hat{v}_{t-2} + \cdots$$

Elements in $\hat{\Psi}_k$ can be easily calculated since we know $\hat{A}$ and $\hat{P}$. The impulse(s) are (in matrix notation): $\frac{\partial y_{t+i}}{\partial v_t} = \Psi_i = A^i P^{-11}$.

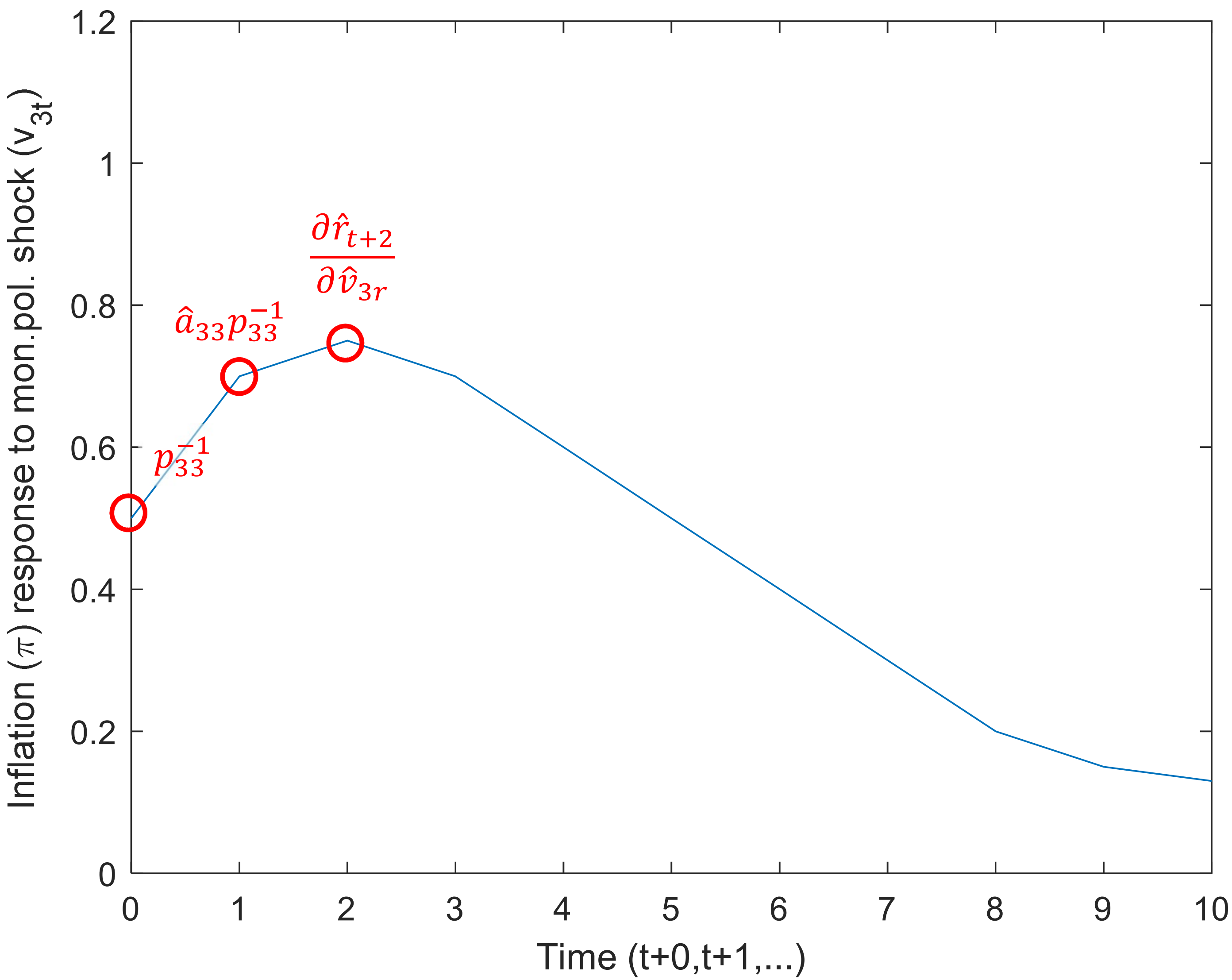
- Assume now that in period t we have a unit-valued monetary policy shock ($v_t = [0,0,1]'$) and trace the system dynamics:

$$\mathrm{y_t} = \begin{bmatrix} inflation_t \\ rGDP_t \\ r_t \end{bmatrix} = \hat{v}_t + \widehat{\Psi}_1 \underbrace{\hat{v}_{t-1}}_{=0} + \widehat{\Psi}_2 \underbrace{\hat{v}_{t-2}}_{=0} + \cdots = P^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ p_{33}^{-1} \end{bmatrix} \quad \text{in } t$$

$$\mathrm{y_{t+1}} = \begin{bmatrix} inflation_{t+1} \\ rGDP_{t+1} \\ r_{t+1} \end{bmatrix} = \underbrace{\hat{v}_{t+1}}_{=0} + \widehat{\Psi}_1 \hat{v}_t + \widehat{\Psi}_2 \underbrace{\hat{v}_{t-1}}_{=0} + \cdots = \widehat{\Psi}_1 \hat{v}_t = \hat{A} P^{-1} v_t = \begin{bmatrix} \hat{a}_{13} p_{33}^{-1} \\ \hat{a}_{23} p_{33}^{-1} \\ \hat{a}_{33} p_{33}^{-1} \end{bmatrix} \quad \text{in } t+1$$

$$\mathrm{y_{t+2}} = \begin{bmatrix} inflation_{t+2} \\ rGDP_{t+2} \\ r_{t+2} \end{bmatrix} = \quad \ldots \quad = \hat{A}^2 P^{-1} v_t = \begin{bmatrix} * \\ * \\ \frac{\partial \hat{r}_{t+2}}{\partial \hat{v}_{3r}} \end{bmatrix} \text{in } t+2$$

In example, we see that after a unit monetary policy shock ($\hat{v}_{3t} = 1$) the interest rate (in red) increases by p_{33}^{-1} in t, by $\hat{a}_{33} p_{33}^{-1}$ in t+1 … We typically plot impulse responses.

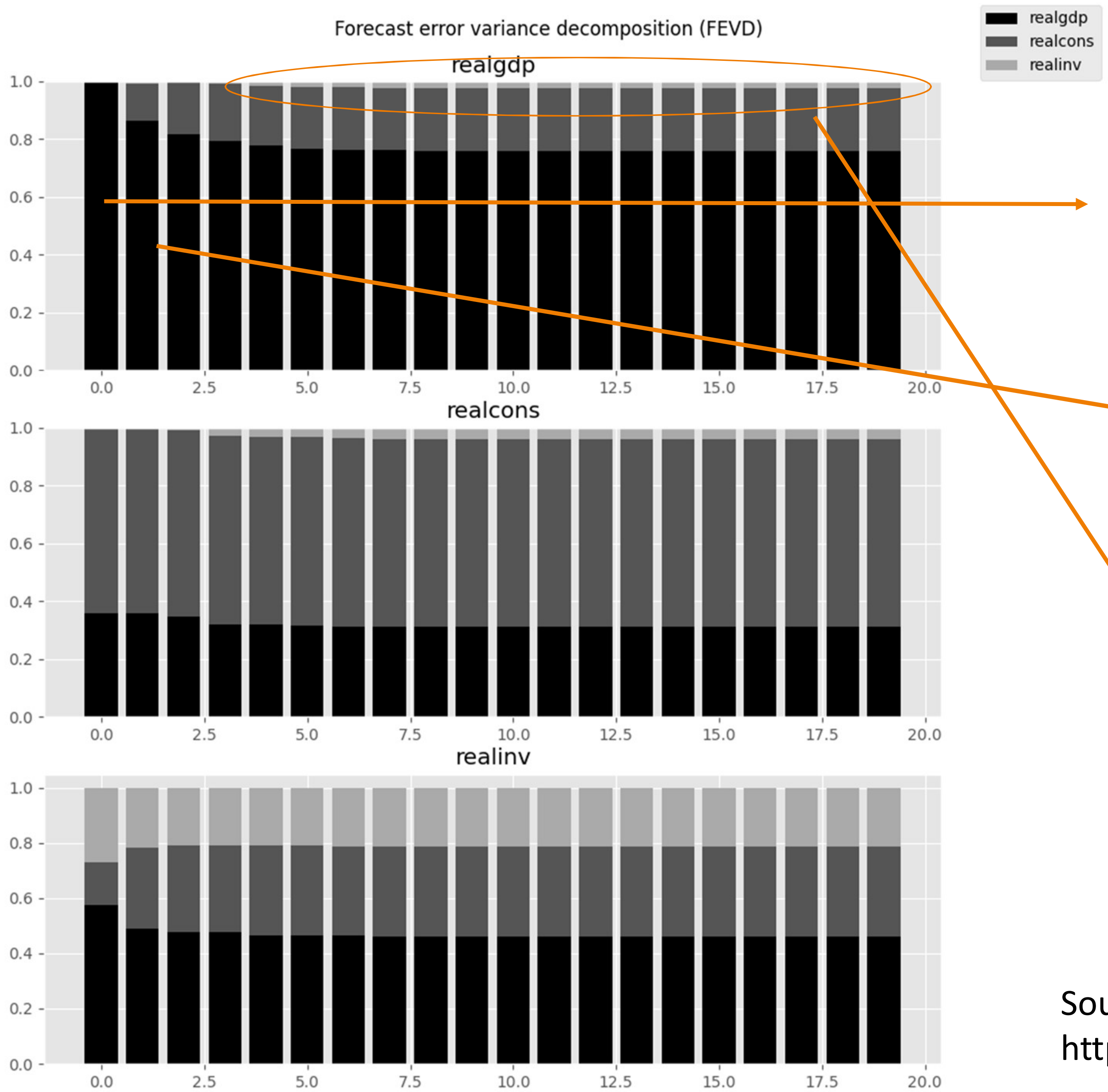


Inflation (π) response to mon.pol. shock (v_{3t})
Time (t+0,t+1,...)
p_{33}^{-1}
$\hat{a}_{33}p_{33}^{-1}$
$\frac{\partial \hat{r}_{t+2}}{\partial \hat{v}_{3r}}$
0
0.2
0.4
0.6
0.8
1
1.2
0
1
2
3
4
5
6
7
8
9
10

- There exist more elaborate structures to identify deep shocks.
- Sometimes we are not interested in causality and identified structural shocks (such as mon. policy shock in the example).
- Sometimes we are only interested in correlations and co-movements. For example, we might want to know how rGDP co-moves with interest rates, regardless of what combination of deep shocks causes the co-movement. This can be useful for forecasting.
- In these cases we often evaluate generalized impulse responses.

Forecast error variance decomposition

- We can use the system of equations derived to decompose the forecast error variance of a variable (FEVD).
- FEVD tells us what share of forecast error variance for horizon «h» is due to uncertainty driven by a specific deep shock or a combination of reduced shocks related to a specific variable.
- E.g. we can answer questions like «what share of forecast error variance for rGDP is driven by uncertainty related to monetary policy shocks?»
- E.g. we can answer questions like «what share of forecast error variance for rGDP is driven by uncertainty related errors in interest rate equation?»
- We can therefore make judgment of what drives our forecast errors! This is useful as it tells us which variable we need to model better to improve the forecast.



EXAMPLE

On impact (period 0), 100% of variance of rGDP is driven by shocks to rGDP. Why? Think of P.

In period 1, about 90% of variance of rGDP is driven by shocks to rGDP. About 10% is driven by shocks to real consumption.

Real investment does not contribute much to the variance of rGDP (not realistic).

Source: https://www.statsmodels.org/stable/vector_ar.html

- H-step ahead forecast error and variance (general):

$$e_{T+h} = \epsilon_{T+h} + \Theta_1 \epsilon_{T+h-1} + \cdots + \Theta_{h-1} \epsilon_{T+1}$$

$$Var(e_{T+h}) = \Sigma + \Theta_1 \Sigma \Theta_1' + \cdots + \Theta_{h-1} \Sigma\, \Theta_{h-1}'$$

- Substitute in VAR(1) parameters:

$$Var(e_{T+h})$$
$$= \mathrm{P}^{-1}\mathrm{P}\Sigma\mathrm{P}'(\mathrm{P}^{-1})' + \Theta_1 \mathrm{P}^{-1}\mathrm{P}\Sigma\mathrm{P}'(\mathrm{P}^{-1})'\Theta_1' + \cdots + \Theta_{h-1}\mathrm{P}^{-1}\mathrm{P}\Sigma\mathrm{P}'(\mathrm{P}^{-1})'\Theta_{h-1}' =$$
$$= \Psi_1 \Psi_1' + \Psi_2 \Psi_2' + \cdots + \Psi_h \Psi_h'$$

- It follows (use $P\Sigma P' = I$ and $\Theta_{i-1} P^{-1} = \Psi_i$) "i,j-th" element of $\Psi_{ij,h}$ represents the contribution of j-th variable in explaining the h-step ahead forecast error variance for i-th variable (y_i).

- For example:

$$\mathrm{y_t} = \begin{bmatrix} inflation_t \\ rGDP_t \\ r_t \end{bmatrix}$$

Ψ_{13}^2 is the contribution of the monetary policy shock (sh. 3) in explaining the variance of the on-step ahead forecast error of inflation (var. 1). If this value is high (relative to $\Psi_{11}^2 + \Psi_{12}^2$) then monetary policy is important for inflation and we might want to spend more time modelling monetary policy well.

Advanced topics

- ECM
- «Big» data: FVARs, FAVARs, BVARs, Global VARs
- Mixed frequency models: MF-VAR, MF-BVAR,..
- Time varying parameters: TVP-VAR, MS-VAR, T-VAR,…
- Time varying variances: V-ARCH, V-GARCH, SV-VAR
- ML methods used in conjuncture with standard econometric models: Double-IV, ML-FVAR,…