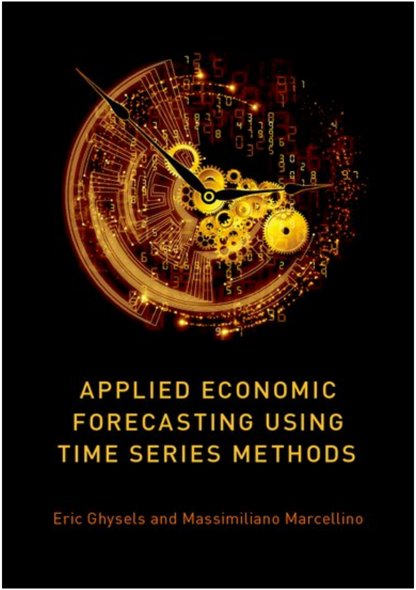
APPLIED ECONOMIC
FORECASTING USING
TIME SERIES METHODS
Eric Ghysels and Massimiliano Marcellino

Applied Economic Forecasting using Time Series Methods

Eric Ghysels and Massimiliano Marcellino

Companion Slides - Chapter 6 VAR Models

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

A VAR(p) model for the set of m variables $y_{1t}, \ldots, y_{mt}$ grouped in the $m \times 1$ vector $y_t = (y_{1t}, \ldots, y_{mt})'$ is:

$$\underset{m\times 1}{y_t} = \underset{m\times 1}{\mu} + \underset{m\times m}{\Phi_1}\, y_{t-1} + \ldots + \Phi_p y_{t-p} + \underset{m\times 1}{\varepsilon_t}\,, \quad \varepsilon_t \sim \text{WN}\left(0, \underset{m\times m}{\Sigma}\right). \tag{1}$$

We assume:

$$E(\varepsilon_{it}) = 0\;, i = 1, \ldots, m$$

$$E(\varepsilon_{it}\varepsilon_{jt-\tau}) = \begin{cases} \sigma_{ij} & \tau = 0, i \neq j \\ 0 & \tau \neq 0, i = j \end{cases}$$

$$E(\varepsilon_t\varepsilon_t') = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1m} \\ \sigma_{12} & \sigma_{22} & \ldots & \sigma_{2m} \\ \ldots & & \ldots & \\ \sigma_{1m} & \sigma_{2m} & \ldots & \sigma_{mm} \end{bmatrix}$$

Representation

- Note that the total number of parameters in a VAR(p) is m (the intercepts) plus m^2p (the coefficients of the lagged variables) plus $m(m+1)/2$ (the variances and covariances of the errors).

- Hence, the total number of parameters grows very fast with the number of variables m, usually we choose variables carefully based on economic theory considerations.

As an example, let us consider the expanded form of a VAR(1) for three variables, with no intercept. We have:

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

with

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \sim \text{WN}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \right).$$

Representation

- As in the univariate case, **weak stationarity** is an important property for representation and estimation. In a multivariate context, it requires that

$$\begin{aligned} E(y_t) &= c, \\ Var(y_t) &= V_0 < \infty, \\ Cov(y_t, y_{t+k}) &= V_k \text{ depends on } k \text{ but not on } t, \end{aligned}$$

  where c is now an $m \times 1$ vector and V_i an $m \times m$ matrix, $i = 0, 1, \ldots$.

- For a VAR(p) process, weak stationarity is verified **when all the roots z of $\det(I - \Phi_1 z - \ldots - \Phi_p z^p) = 0$ are larger than one in absolute value.**

Representation

VAR models are approximations to VMA(∞) models, if y_t is an m-dimensional weakly stationary process, from the Wold theorem it admits the representation

$$y_t = C(L)\varepsilon_t, \tag{2}$$

where $C(L)$ is an $m \times m$ matrix polynomial in L ($C(L) = I + C_1 L + C_2 L^2 + \ldots$), and ε_t is an $m \times 1$ vector white noise process, $\varepsilon_t \sim$ WN$(0, \Sigma)$ . .

- Under mild conditions, $C(L)$ can be approximated by $A^{-1}(L)B(L)$, so that we can rewrite (Formula 2) as

$$A(L)y_t = B(L)\varepsilon_t \tag{3}$$

  with $A(L) = I + A_1L + \ldots + A_pL^p$, $B(L) = I + B_0L + \ldots + B_qL^q$.
- Under slightly more stringent conditions, we can also assume that $q = 0$, such that (Formula 3) becomes a VAR(p) model.

Representation
Specification of the model
Estimation
Diagnostic checks
Forecasting
Impulse response functions
Forecast error variance decomposition
Structural VARs with long-run restrictions
VAR models with simulated data
Empirical examples
Concluding remarks

Select variables under analysis:

- The choice of the variables under analysis should be driven by economic theory, related to the specific problem of interest, and kept as limited as possible in order to avoid the curse of dimensionality.
- On the other hand, modeling too few variables can generate omitted variable problems.

Specification of the model

Select the deterministic component:

- Except intercept, additional variables, such as seasonal dummies, other types of dummy variables to capture potential parameter breaks, and trends can also be inserted.

Select number of lags:

- We can start with a high number of lags, and then reduce it by sequential testing for their significance using the Wald or the LR tests.
- We can use multivariate versions of the information criteria (IC): Compute the IC for a range of values, $j = 1, \ldots, p_{max}$, and select the number of lags that minimizes the IC.

Specification of the model

- The most common multivariate IC are

$$
\begin{aligned}
AIC(j) &= \ln|\Sigma_j| + \frac{2}{T} jm^2, \\
BIC(j) &= \ln|\Sigma_j| + \frac{\ln T}{T} jm^2, \quad j = 1, \ldots, p_{\max},
\end{aligned}
$$

  where AIC stands for **Akaike's information criterion** and BIC for **Bayesian information criterion**, also known as Schwarz information criterion.

- it can be easily seen that the BIC is consistent but the AIC is not.

- Since for $T \geq 8$ the penalty is larger for BIC than AIC, in general in empirical applications BIC will select a lower lag order than AIC.

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

Let us write the VAR(p) as follows:

$$y_t = Bx_t + \varepsilon_t$$

where

$$B \quad = (\mu \quad \Phi_1 \ldots \Phi_p) \quad (m \times (mp+1))$$

$$x_t \qquad = \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix} \qquad (mp+1) \times 1.$$

Therefore:

$$y_t | y_{t-1}, \ldots, y_{t-p} \sim N(Bx_t, \Sigma)$$

and the log-likelihood function is:

$$\begin{aligned} \mathcal{L}(y_1, \ldots, y_T; B, \Sigma) \quad &= \quad -\frac{Tm}{2} \log(2\pi) - \frac{T}{2} \log|\Sigma^{-1}| \\ &\quad -\frac{1}{2} \sum^{T} \left[(y_t - Bx_t) \Sigma^{-1} (y_t - Bx_t) \right] \end{aligned}$$

Estimation

- The resulting ML estimator of the VAR model parameters is:

$$\widehat{B} = \left[\sum_{t=1}^{T} y_t x_t'\right] \left[\sum_{t=1}^{T} x_t x_t'\right]$$

- The **ML estimator of the VAR parameters is equivalent to equation by equation OLS estimators.**

- OLS equation by equation is consistent and asymptotically efficient assuming that all the assumptions underlying the VAR are valid.

- The OLS estimators are also asymptotically normal distributed (with the usual $\sqrt{T}$ rate of convergence).
- Inference on the parameters can be conducted by standard Wald-type statistics, which will have χ^2 asymptotic distributions.

Estimation

- The **ML estimator for the innovation variance is:**

$$\widehat{\Sigma} = (1/T) \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

where $\hat{\varepsilon}_t$ = y_t - $\widehat{B} x_t$.

- The ML estimator of the variance is consistent, but it is biased in small samples, so it is common to use the variance estimator adjusted by the number of degrees of freedom:

$$\widetilde{\Sigma} = \frac{T}{(T - mp - m)} \widehat{\Sigma}$$

Estimation

- If we want to estimate the VAR parameters of VARMA models, leaving the MA parameters unspecified, except for the order q,
- then x_t is no longer a valid instrument for the estimation of the VAR parameters in B since lagged, say y_{t-1} are correlated with the MA process determining ε_t.

Estimation

- However, assuming q known, y_{t-q-k}, $k \geq 1$ are valid instruments, and therefore we can estimate via instrumental variables

$$\widetilde{B} = \left[\sum_{t=1}^{T} y_t z_t'\right] \left[\sum_{t=1}^{T} x_t z_t'\right]$$

where the instrument set is

$$z_t = \begin{bmatrix} 1 \\ y_{t-q-1} \\ y_{t-q-2} \\ \vdots \\ y_{t-q-p} \end{bmatrix}.$$

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

Diagnostic checks

- So far, we have assumed that the **errors are (multivariate) white noise, namely, uncorrelated and homoskedastic.** These properties can be tested equation by equation, using the testing procedures we have briefly discussed.
- As an alternative, there exist multivariate versions of the LM (Breusch-Godfrey) test for no correlation, of the White test for homoskedasticity, and of the Jarque-Bera test for normality.

Diagnostic checks

- From a forecasting point of view, **parameter stability** is particularly important (both in- and out-of-sample).
- we can consider recursively estimated parameters and confidence intervals, or we can conduct Chow tests in the case where the break dates are known, equation by equation.

Diagnostic checks

If the assumptions are rejected we may want to reconsider the model specification, e.g.,

- By increasing the lag length,
- Including additional (possibly dummy) variables,
- Changing the specification of the deterministic component,
- Modifying the estimation sample.

VAR Models

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

Forecasting

- The **optimal forecast** in the MSFE sense for a VAR(p) is:

$$\widehat{y}_{T+h} = \Phi_1 \widehat{y}_{T+h-1} + \ldots \Phi_p \widehat{y}_{T+h-p} \tag{4}$$

  where $\widehat{y}_{T+h-j} = y_{T+h-j}$ for $h - j \leq 0$,

- Hence, to compute the forecast for $\widehat{y}_{T+h}$, we calculate $\widehat{y}_{T+1}$, use it to obtain $\widehat{y}_{T+2}$ and keep iterating until we obtain $\widehat{y}_{T+h}$.

- As an example, if y is VAR(1),

$$\begin{aligned}
\widehat{y}_{T+1} &= Ay_T, \\
\widehat{y}_{T+2} &= A\widehat{y}_{T+1} = A^2 y_T, \\
&\ldots \\
\widehat{y}_{T+h} &= A\widehat{y}_{T+h-1} = A^h y_T.
\end{aligned}$$

- The **direct forecasting method** presented in the previous chapter provides an alternative forecast for y_{T+h}. In the VAR(p) context, the direct model takes the form

$$y_t = A_1 y_{t-h} + \ldots + A_p y_{t-h-p} + u_t,$$

with forecast

$$\widetilde{y}_{T+h} = A_1 y_T + \ldots + A_p y_{T-p}. \tag{5}$$

- Under correct specification, $\widehat{y}_{T+h}$ is more efficient than $\widetilde{y}_{T+h}$. However, in the presence of model mis-specifications, $\widetilde{y}_{T+h}$ can be more robust than $\widehat{y}_{T+h}$.

- Given a VAR written in the form:

$$\begin{bmatrix} a_{11}(L) & a_{12}(L) \\ a_{21}(L) & a_{22}(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}, \tag{6}$$

  x_t does not Granger-cause y_t if $a_{12}(L) = 0$, that is y_t does not depend on the lags of x_t. Similarly, y_t does not Granger-cause x_t if $a_{21}(L) = 0$.

- Since the variables can be contemporaneously correlated through the errors, it is possible to have causality even in the presence of Granger non-causality.

From formula 2, the MA(∞) representation of y_t, we can also write the optimal forecast as

$$\widehat{y}_{T+h} = \sum_{j=0}^{\infty} C_{j+h}\varepsilon_{T-j}, \tag{7}$$

with associated forecast error

$$e_{T+h} = \sum_{j=0}^{h-1} C_j \varepsilon_{T+h-j}. \tag{8}$$

Hence, the **variance covariance matrix of the forecast error** is

$$V(e_{T+h}) = \Sigma + C_1\Sigma C_1' + \ldots + C_{h-1}\Sigma C_{h-1}'.$$

- The elements on the diagonal of $V(e_{T+h})$ contain the variances of the forecast errors for each specific variable, $V(e_{1T+h}),\ldots,V(e_{mT+h})$.
- Under the additional assumption of normal errors, the interval forecast for variable $\widehat{y}_{jT+h}$, $j = 1, \ldots, m$, takes the form

$$\left[\widehat{y}_{jT+h} - c_{\alpha/2}\sqrt{V(e_{jT+h})}, \widehat{y}_{jT+h} + c_{\alpha/2}\sqrt{V(e_{jT+h})}\right], \tag{9}$$

where $c_{\alpha/2}$ is the proper critical value from the standard normal distribution.

The formula for the optimal forecast in (formula 7) can be rearranged into

$$\widehat{y}_{T+h} = C_h \varepsilon_T + \sum_{j=0}^{\infty} C_{j+h+1} \varepsilon_{T-1-j} = \widehat{y}_{T+h|T-1} + C_h \varepsilon_T,$$

and using (formula 8) gives

$$\widehat{y}_{T+h} = \widehat{y}_{T+h|T-1} + C_h \underbrace{(y_T - \widehat{y}_{T|T-1})}_{forecast\ error}. \tag{10}$$

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

A stationary VAR for the m variables grouped in the vector y_t can be written in MA(∞) form as

$$y_t = \Phi^{-1}(L)\,\varepsilon_t = \Theta(L)\,\varepsilon_t, \quad \varepsilon_t \sim \text{WN}(0, \Sigma)\,. \tag{11}$$

Since Σ is positive definite, there exists a non-singular matrix P such that

$$P\Sigma P' = I. \tag{12}$$

We can therefore rewrite equation (11) as

$$\begin{aligned} y_t &= \Theta(L)\,P^{-1}P\varepsilon_t = \Psi(L)\,v_t, \qquad (13)\\ v_t &= P\varepsilon_t, \\ E(v_t) &= 0, \quad E\left(v_t v_t'\right) = P\Sigma P' = I. \end{aligned}$$

Equation (formula 13) is the MA(∞) representation of the model

$$P\Phi(L)\,y_t = v_t, \tag{14}$$

which is typically known as a **Structural VAR (SVAR)**, as there are contemporaneous relationships among the variables because of the P matrix. The orthogonal errors v_t are usually interpreted as **structural (economic) shocks**, e.g., demand or supply shocks.

Impulse response functions

It is interesting to compute how the variables under analysis react to the structural shocks. We can write

$$
\begin{aligned}
\Psi(L) &= P^{-1} - \Theta_1 P^{-1} L - \Theta_2 P^{-1} L^2 - \ldots \\
&= \Psi_1 - \Psi_2 L - \Psi_3 L^2 - \ldots
\end{aligned}
\tag{15}
$$

A shock is a vector with one element equal to one and all the others equal to zero, e.g.,

$$
v_{1,t+1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, v_{2,t+1} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, v_{m,t+1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{16}
$$

Impulse response functions

The response in period $t+i$ of y to a shock in period $t+1$ will be:

$$\begin{aligned}\frac{\partial y_{t+1}}{\partial v_{t+1}} &= P^{-1} = \Psi_1, \\ \frac{\partial y_{t+2}}{\partial v_{t+1}} &= -\Psi_2, \\ \frac{\partial y_{t+3}}{\partial v_{t+1}} &= -\Psi_3, \\ & \dots\end{aligned} \tag{17}$$

Impulse response functions

For example,

$$\Psi_1 = \begin{bmatrix} \frac{\partial y_{1,t+1}}{\partial v_{1,t+1}} & \cdots & \frac{\partial y_{1,t+1}}{\partial v_{m,t+1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{m,t+1}}{\partial v_{1,t+1}} & \cdots & \frac{\partial y_{m,t+1}}{\partial v_{m,t+1}} \end{bmatrix}. \quad (18)$$

The term $\partial y_{k,t+i}/\partial v_{jt+1}$ is known as the impulse response in period $t+i$ of variable k to the shock j, with $k, j = 1, \ldots, m$ and $i = 1, 2, \ldots$. The collection of all impulse responses is knows as the **impulse response function (IRF)**.

Impulse response functions

- There are many invertible matrices P that satisfy $P\Sigma P' = I$.
- Since Σ, the variance matrix of ε_t, has $m(m+1)/2$ distinct elements, this is the maximum number of unrestricted elements in P.
- **Typically, P is chosen as a triangular matrix:**

$$P = \begin{bmatrix} p_{11} & 0 & \dots & 0 \\ p_{12} & p_{22} & \dots & 0 \\ \vdots & & \ddots & \\ p_{1m} & p_{2m} & \dots & p_{mm} \end{bmatrix},$$

- In this context, different identification schemes are associated to different orderings of the m variables in y_t.

Impulse response functions

- Empirically, the elements of P are unknown and, once its form is chosen, they must be estimated starting from those of $\widehat{\Sigma}$.
- Estimates of P are combined with those of the other VAR parameters $\Phi(L)$ in (formula 14) to obtain estimates for $\Theta(L)$ and $\Psi(L)$ in (formula 13).

Impulse response functions

- Appropriate standard errors for the IRF can also be derived, either analytically or by means of **Monte Carlo or Bootstrap methods.**
- Typically, the elements of the estimated IRF and their associated **confidence bands** are reported in multi-panel graphs.

As an example,let us assume that

$$y_t = \begin{bmatrix} \text{output gap} \\ \text{inflation} \\ \text{short term interest rate} \end{bmatrix} = \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix}.$$

Let us assume that y_t follows as VAR(1) process:

$$y_t = A_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN}\left(0, \Sigma\right)$$

In order to identify the structural shocks, as discussed above we assume that

$$v_t = P\varepsilon_t, \quad P\Sigma P' = I = Var\left(v_t\right),$$

with

$$P = \begin{bmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix}.$$

It follows that

$$P^{-1} = \begin{bmatrix} \alpha_{11} & 0 & 0 \\ \alpha_{21} & \alpha_{22} & 0 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} \tag{19}$$

$$= \begin{bmatrix} \frac{1}{p_{11}} & 0 & 0 \\ -\frac{1}{p_{11}}\frac{p_{21}}{p_{22}} & \frac{1}{p_{22}} & 0 \\ \frac{(p_{21}p_{32}-p_{22}p_{31})}{p_{11}p_{22}p_{33}} & -\frac{1}{p_{22}}\frac{p_{32}}{p_{33}} & \frac{1}{p_{33}} \end{bmatrix},$$

$$v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix} = \begin{bmatrix} p_{11}\varepsilon_{1t} \\ p_{21}\varepsilon_{1t} + p_{22}\varepsilon_{2t} \\ p_{31}\varepsilon_{1t} + p_{32}\varepsilon_{2t} + p_{33}\varepsilon_{3t} \end{bmatrix},$$

and the SVAR can be written as

$$y_t = A_1 y_{t-1} + P^{-1} v_t. \tag{20}$$

Impulse response functions

- We interpret v_1 as a demand shock, v_2 as a supply shock, and v_3 as a monetary policy shock.
- Given the structure of the matrix P, and of P^{-1}, we assume that:
  (1) the demand shock influences all the three variables contemporaneously,
  (2) the supply shock affects current inflation and interest rate but the output gap only with a delay,
  (3) the monetary shock has a delayed effect on both output gap and inflation, while the interest rate immediately reacts.

Impulse response functions

- Formula (20) implies that (1) y_{1t} is affected by $\alpha_{11}v_{1t}$, (2) y_{2t} is affected by $\alpha_{21}v_{1t} + \alpha_{22}v_{2t}$, (3) y_{3t} is affected by $\alpha_{31}v_{1t} + \alpha_{32}v_{2t} + \alpha_{33}v_{3t}$.
- Hence, another way to interpret our identification scheme is that:
(1) the output gap is only affected contemporaneously by its own shock,
(2) inflation by its own shock plus the output gap shock,
(3) interest rate by its own shock plus the output gap shock, and the inflation shock.

Impulse response functions

The MA(∞) representation of the model is

$$y_t = \varepsilon_t + A_1\varepsilon_{t-1} + A_1^2\varepsilon_{t-2} + A_1^3\varepsilon_{t-3} + \ldots = \Theta\left(L\right)\varepsilon_t$$

Hence, the impulse response functions are

$$\frac{\partial y_{t+i}}{v_t} = A^i P^{-1} = \Psi_i = \begin{bmatrix} \Psi_{11} & \Psi_{12} & \Psi_{13} \\ \Psi_{21} & \Psi_{22} & \Psi_{23} \\ \Psi_{31} & \Psi_{32} & \Psi_{33} \end{bmatrix},$$

where

$$\Psi_{qj,i} = \frac{\partial y_{qt+i}}{v_{jt}}, \quad q = 1,2,3 \quad j = 1,2,3, \quad i = 1,2,\ldots \tag{21}$$

and

$$v_{1t} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, v_{2t} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, v_{3t} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Impulse response functions

To conclude, we note that there exist more complex SVAR structures linking VAR residuals and structural shocks, typically expressed as

$$Bv_t = A\varepsilon_t. \tag{22}$$

It is again necessary to recover the parameters in the B and A matrices from those of the variance of the VAR residuals, Σ, with

$$A^{-1}BB'\left(A^{-1}\right)' = \Sigma.$$

Therefore, we need to impose a sufficient number of a priori restrictions on A and B to identify their remaining free parameters from those of Σ.

VAR Models

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

We have seen before that the **h-steps ahead forecast error** can be written as

$$e_{T+h} = \varepsilon_{T+h} + \Theta_1 \varepsilon_{T+h-1} + \ldots + \Theta_{h-1} \varepsilon_{T+1},$$

so that

$$Var\left(e_{T+h}\right) = \Sigma + \Theta_1 \Sigma \Theta_1' + \ldots + \Theta_{h-1} \Sigma \Theta_{h-1}'.$$

Let us now write $Var\left(e_{T+h}\right)$ as

$$\begin{aligned} Var\left(e_{T+h}\right) &= P^{-1} P \Sigma P' (P^{-1})' + \Theta_1 P^{-1} P \Sigma P' (P^{-1})' \Theta_1' + \\ &\quad + \ldots + \Theta_{h-1} P^{-1} P \Sigma P' (P^{-1})' \Theta_{h-1}' \\ &= \Psi_1 \Psi_1' + \Psi_2 \Psi_2' \ldots + \Psi_h \Psi_h' \end{aligned} \tag{23}$$

since $P \Sigma P' = I$ and $\Theta_{i-1} P^{-1} = \Psi_i$.

Forecast error variance decomposition

It follows that

$$\Psi_{ij,1}^2 + \Psi_{ij,2}^2 + \ldots + \Psi_{ij,h}^2, \tag{24}$$

represents the contribution of the innovations in the j^{th} variable in explaining the h-steps ahead forecast error variance for y_i, $i, j = 1, \ldots, m$. This is the so-called **forecast error variance decomposition (FEVD).**

Forecast error variance decomposition

Let us consider in more detail, as an example,

$$\begin{bmatrix} Var\left(e_{1,T+1}\right) & Cov\left(e_{1,T+1}, e_{2,T+1}\right) & \dots & Cov\left(e_{1,T+1}, e_{m,T+1}\right) \\ Cov\left(e_{2,T+1}, e_{1,T+1}\right) & Var\left(e_{2,T+1}\right) & \dots & Cov\left(e_{2,T+1}, e_{m,T+1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ Cov\left(e_{m,T+1}, e_{1,T+1}\right) & Cov\left(e_{m,T+1}, e_{2,T+1}\right) & \dots & Var\left(e_{m,T+1}\right) \end{bmatrix} = \Psi_1 \Psi_1',$$

where

$$\Psi_1 = \begin{bmatrix} \Psi_{11} & \dots & \Psi_{1m} \\ \Psi_{21} & \dots & \Psi_{2m} \\ \vdots & \ddots & \vdots \\ \Psi_{m1} & \dots & \Psi_{mm} \end{bmatrix}$$

Forecast error variance decomposition

The one-step ahead forecast error variance for the first variable in the system can be decomposed

$$Var\left(e_{1,T+1}\right) = \Psi_{11}^2 + \Psi_{12}^2 + \ldots + \Psi_{1m}^2. \tag{25}$$

Similarly, for the second variable it is

$$Var\left(e_{2,T+1}\right) = \Psi_{21}^2 + \Psi_{22}^2 + \ldots + \Psi_{2m}^2.$$

And so on for the other variables.

Forecast error variance decomposition

Going back to the example where

$$y_t = \begin{bmatrix} \text{output gap} \\ \text{inflation} \\ \text{short term interest rate} \end{bmatrix},$$

we have that, e.g., Ψ_{21}^2 is the contribution of the “structural” shock 1 (demand shock) in explaining the variance of the one-step ahead forecast error of variable 2 (inflation), while Ψ_{22}^2 and Ψ_{23}^2 are the contributions of, respectively, the supply and monetary shocks.

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

Structural VARs with long-run restrictions

- In this section, we will consider the role of assumptions on the long-run behavior of some variables and/or effects of some shocks.
- We illustrate the method by means of an example, where we want to identify demand and supply shocks exploiting the generally accepted proposition that demand shocks do not have long-run effects, while supply shocks do. See Blanchard and Quah(1989) for more details.

Structural VARs with long-run restrictions

We consider a bivariate VAR(1) for GDP growth and unemployment, defined as

$$y_t = Ay_{t-1} + Bv_t, \tag{26}$$

where v_t are the structural shocks, and $B = P^{-1}$ in the previous notation.
The MA representation is

$$y_t = (I - AL)^{-1} Bv_t, \tag{27}$$

from which the cumulative long-run response of y to the shocks in v is

$$(I - A)^{-1} B = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}. \tag{28}$$

The i^{th}, j^{th} element of this matrix captures the long-run effects of shock j on variable i.

For identification purpose, we impose a priori restriction $b_{21} = 0$.

If

$$y_t = \begin{bmatrix} \text{growth}_t \\ \text{unemployment}_t \end{bmatrix},$$

then the demand shock v_{1t} has no long-run effects on growth if

$$\pi_{11} b_{11} + \pi_{12} b_{21} = 0. \tag{29}$$

In fact, with this restriction it is

$$(I - AL)^{-1} B v_t = \begin{bmatrix} 0 & * \\ * & * \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \tag{30}$$

Hence, rather than imposing $b_{21} = 0$ as with the Cholesky approach, we can assume that b_{11} and b_{21} are linked by the linear relationship in (29).

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

VAR models with simulated data

- We now illustrate the use of VAR models using simulated data. Specifically, we generate 600 observations using a bivariate VAR(1) DGP, and in order to avoid dependence on the starting values, we discard the first 100 observations.
- We will use the methods described in this chapter to specify and estimate an appropriate VAR model, and then to provide forecasts for the two series under analysis by using both deterministic and stochastic simulation.
- The computed forecasts will be evaluated in comparison with those obtained by an equivalent structural equation model.

VAR models with simulated data

- **Deterministic forecast** means the iterated approach is applied to compute the forecasts (the model is solved forward deterministically).
- The **stochastic forecast** means bootstrapped errors are added to the forward solution in each of a set of simulations, and the average across all the simulations is treated as the forecast.

VAR models with simulated data

The DGP is a simple bivariate model with the following specification:

$$\begin{aligned} y_t &= 1 + 0.8 y_{t-1} + \varepsilon_{yt} \\ x_t &= 1 + 0.5 x_{t-1} + 0.6 y_{t-1} + \varepsilon_{xt} \end{aligned} \tag{31}$$

with

$$v_t = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} \varepsilon_t \quad \text{and} \quad v_t \overset{iid}{\sim} N(0.I). \tag{32}$$

The pattern of the simulated series is reported in Figure (1).

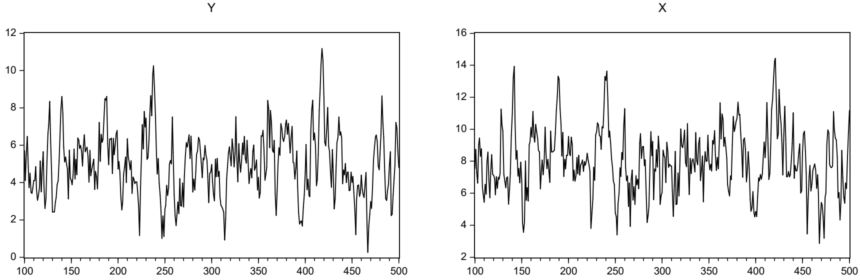

Figure 1: *Simulated series x and y*

From the figure, the variables look stationary and tend to co-move. We can think of y could be the growth rate of consumption and x of disposable income.

VAR models with simulated data

- We can start with a sufficiently long **lag length** (a general model) and then we have two options to reduce it:

  (1) either we use the multivariate version of the BIC and AIC information criteria

  (2) or we perform a sequence of LR tests for the significance of the various lags, starting with testing the significance of p versus $p-1$ lags.

- In this example, all the information criteria and the LR test agree on selecting p = 1, as we can see from Table 1. We estimate both a VAR(1) and a VAR(4) model, to assess the effects of over-parameterization.

VAR Lag Order Selection Criteria

Lag	*LogL*	*LR*	*FPE*	*AIC*	*SC*	*HQ*
0	-1596.550	NA	9.945	7.973	7.993	7.981
1	-1120.800	944.396*	0.946*	5.620*	5.680*	5.644*
2	-1118.250	5.030	0.953	5.627	5.727	5.667
3	-1117.630	1.230	0.969	5.644	5.783	5.699
4	-1114.200	6.699	0.972	5.647	5.826	5.718
5	-1113.940	0.515	0.990	5.666	5.885	5.752
6	-1111.720	4.291	0.999	5.674	5.933	5.777
7	-1109.430	4.396	1.007	5.683	5.982	5.801
8	-1108.140	2.468	1.021	5.696	6.035	5.831
9	-1105.210	5.590	1.027	5.702	6.080	5.852
10	-1103.130	3.936	1.036	5.711	6.130	5.877
11	-1101.080	3.878	1.047	5.721	6.179	5.903
12	-1098.540	4.756	1.054	5.728	6.226	5.926

* Indicates lag order selected by the criterion. LR: sequential test (each at 5% level), FPE: final prediction error, AIC: Akaike information criterion, SC: Schwarz, HQ: Hannan-Quinn

Table 1: *Results of the lag length selection tests*

VAR models with simulated data

Vector Autoregression Estimates

	Y	X
Y(-1)	0.790	0.597
	(0.035)	(0.031)
	[22.695]	[19.380]
X(-1)	-0.047	0.523
	(0.030)	(0.027)
	[-1.555]	[19.429]
C	1.421	0.862
	(0.242)	(0.214)
	[5.863]	[4.022]

Table2: *Estimation results of the VAR(1)*

VAR models with simulated data

Vector Autoregression Estimates		
Standard errors in() and t-statistics in []		
R-squared	0.594	0.758
Adj. R-squared	0.591	0.757
Sum sq. resids	485.316	379.840
S.E. equation	1.104	0.977
F-statistic	290.570	624.674
Log likelihood	-607.258	-558.125
Akaike AIC	3.044	2.799
Schwarz BIC	3.074	2.829
Mean dep	4.972	8.014
S.D. dep	1.728	1.983
Determinant resid covariance (dof adj.)		0.932
Determinant resid covariance		0.918
Log likelihood		-1120.801
Akaike information criterion		5.620
Schwarz criterion		5.680

Table 2: *Estimation results of the VAR(1)*

Vector Autoregression Estimates		
	Y	X
Y(-1)	0.839	0.648
	(0.056)	(0.050)
	[14.936]	[13.080]
Y(-2)	-0.097	-0.105
	(0.094)	(0.083)
	[-1.027]	[-1.264]
Y(-3)	0.095	-0.061
	(0.094)	(0.083)
	[1.004]	[-0.733]
Y(-4)	-0.146	0.166
	(0.078)	(0.069)
	[-1.866]	[2.403]

Table3: *Estimation results of the VAR(4)*

Vector Autoregression Estimates

	Y	X
X(-1)	0.011	0.577
	(0.064)	(0.056)
	[0.165]	[10.246]
X(-2)	-0.047	-0.049
	(0.074)	(0.065)
	[-0.633]	[-0.759]
X(-3)	0.112	-0.071
	(0.074)	(0.065)
	[1.524]	[-1.089]
X(-4)	-0.036	0.024
	(0.044)	(0.039)
	[-0.812]	[0.606]
C	1.215	0.943
	(0.309)	(0.272)
	[3.936]	[3.467]

Table3: *Estimation results of the VAR(4)*

Vector Autoregression Estimates		
Standard errors in() and t-statistics in []		
R-squared	0.598	0.763
Adj. R-squared	0.590	0.759
Sum sq. resids	479.443	372.030
S.E. equation	1.106	0.974
F-statistic	73.024	158.072
Log likelihood	-604.816	-553.960
Akaike AIC	3.061	2.808
Schwarz BIC	3.151	2.897
Mean dep	4.972	8.014
S.D. dep	1.728	1.983
Determinant resid. covariance (dof adj.)		0.929
Determinant resid.covariance		0.888
Log likelihood		-1114.202
Akaike information criterion		5.647
Schwarz criterion		5.826
Standard errors in() and t-statistics in[]		

Table 3: *Estimation results of the VAR(4)*

VAR models with simulated data

- The overall fit of the two models is quite similar, although the information criteria slightly favor the VAR(1).
- Note also that the inverse roots of the characteristic polynomials of the two models in Figure 2 do not signal any problem of non-stationarity, as expected.

VAR models with simulated data

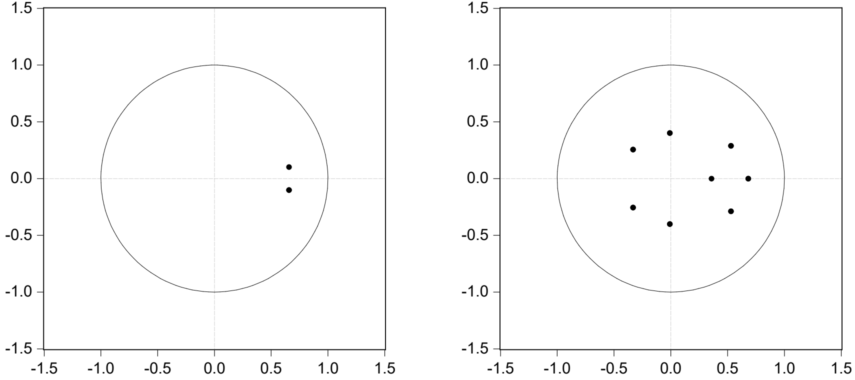

Figure 2: *Inverse roots of the AR characteristic polynomial for the VAR(1) (left panel) and the VAR(4) (right panel)*

VAR models with simulated data

Let us consider **diagnostic tests** on the VAR(1) residuals in Tables 4 - 6.

- The multivariate version of the normality test does not reject the hypothesis that the residuals have a joint normal distribution.
- The multivariate version of the White heteroskedasticity test does not find any sign of neglected residual heteroskedasticty.
- The serial correlation LM test which in both cases cannot reject the null of no serial correlation up to lag 12.

VAR models with simulated data

VAR residual normality tests

Component	Skewness	Chi-sq	df	Prob.
1	0.110	0.812	1	0.367
2	-0.260	4.501	1	0.034
Joint		5.313	2	0.070
Component	**Kurtosis**	**Chi-sq**	**df**	**Prob.**
1	2.911	0.133	1	0.715
2	2.806	0.631	1	0.427
Joint		0.764	2	0.682
Component	**Jarque-Bera**	**df**	**Prob.**	
1	0.945	2	0.623	
2	5.132	2	0.077	
Joint	6.077	4	0.193	

Table 4: *VAR(1) residual normality test*

VAR heteroskedasticity tests

Joint test:					
Chi-sq	df	Prob.			
11.059	12	0.524			
Individual components:					
Dependent	R-squared	F(4,396)	Prob.	Chi-sq(4)	Prob.
res1*res1	0.014	1.429	0.223	5.708	0.222
res2*res2	0.009	0.867	0.484	3.482	0.480
res2*res1	0.002	0.180	0.949	0.727	0.948

Table 5: *VAR(1) residual heteroskedasticity test*

VAR residual serial correlation LM tests		
Lags	LM-Stat	Prob
1	4.470	0.346
2	1.066	0.900
3	6.659	0.155
4	0.721	0.949
5	4.546	0.337
6	1.563	0.816
7	3.588	0.465
8	5.211	0.266
9	4.231	0.376
10	2.308	0.679
11	8.047	0.090
12	1.878	0.758

Table 6: *VAR(1) residual serial correlation LM tests*

VAR models with simulated data

- A usual problem with VAR models is **overparameterization**, which yields very good in-sample fit but bad out-sample forecasts: in our example both the VAR(1) and VAR(4) do very well in terms of in-sample fit, so we now try to assess comparatively their predictive capabilities.
- We use the observations 501-600 as the forecast sample, and construct both **static (one-step ahead)** and **dynamic (one- to 100-steps ahead) forecasts** for the two variables y and x.

VAR models with simulated data

- In the **deterministic setting**, the inputs to the model are fixed at known values, and a single path is calculated for the output variables, using the theoretical formula.
- In the **stochastic environment**, uncertainty is incorporated into the model by adding a random element. To simulate the distributions of future values, the model object uses a **Monte Carlo approach**.

The stochastic forecasts for both y and x (static and dynamic) generated from the VAR(1) specification are graphed in Figure 3 against the respective actual values.

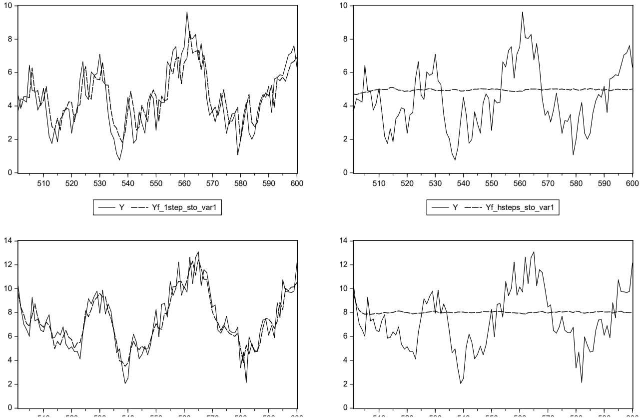

Figure 3: *Static one-step and dynamic h-steps ahead stochastic forecasts against the actuals, VAR(1)*

Table 7 provides the detailed **forecast evaluation statistics** for all the series.

VAR(1)

	x			y	
Stochastic setting					
	h-steps ahead	one-step ahead		h-steps ahead	one-step ahead
RMSFE	2.482	0.975	RMSFE	1.952	1.107
MAFE	2.071	0.788	MAFE	1.643	0.890
Deterministic setting					
	h-steps ahead	one-step ahead		h-steps ahead	one-step ahead
RMSFE	2.477	0.974	RMSFE	1.939	1.107
MAFE	2.063	0.788	MAFE	1.632	0.893

Table7: *Detailed forecast evaluation criteria for VAR(1) and VAR(4)*

VAR models with simulated data

VAR(4)

	x			y	
Stochastic setting					
	h-steps ahead	one-step ahead		h-steps ahead	one-step ahead
RMSFE	2.473	1.003	RMSFE	1.917	1.105
MAFE	2.068	0.794	MAFE	1.617	0.896
Deterministic setting					
	h-steps ahead	one-step ahead		h-steps ahead	one-step ahead
RMSFE	2.479	0.998	RMSFE	1.939	1.100
MAFE	2.068	0.789	MAFE	1.633	0.894

Table 7: *Detailed forecast evaluation criteria for VAR(1) and VAR(4)*

VAR models with simulated data

- If we consider first the one-step ahead forecasts, both the RMSFE and the MAFE are lower for the VAR(1) specification than for the VAR(4), both for the deterministic and the stochastic settings, though the differences are small.
- The pattern is reversed for the dynamic one- to 100-step(s) ahead forecasts (indicated by h-steps ahead in the table), but the differences remain small.

VAR models with simulated data

We want to compare the VAR model with two seperate ARMA models.

$$\begin{aligned} x_t &= 7.97 + 0.73x_{t-1} + 0.15\varepsilon_{t-1} + 0.09\varepsilon_{t-2} + 0.16\varepsilon_{t-3} \\ R^2_{adj} &= 0.71, \quad DW = 1.99 \end{aligned} \tag{33}$$

$$\begin{aligned} y_t &= 4.96 + 0.77y_{t-1} + u_t \\ R^2_{adj} &= 0.59, \quad DW = 1.98 \end{aligned} \tag{34}$$

VAR models with simulated data

We can now produce static and dynamic forecasts from the above models. The evaluation criteria reported in Table 8

	$x \sim$ ARMA(1,3)		$y \sim$ AR(1)
	h-steps ahead		h-steps ahead
RMSFE	2.505	RMSFE	1.942
MAFE	2.107	MAFE	1.641
	one-step ahead		one-step ahead
RMSFE	1.471	RMSFE	1.086
MAFE	1.180	MAFE	0.883

Table 8: *Forecast evaluation criteria for the ARMA models specified for x and y*

VAR models with simulated data

- Next, we use the **Cholesky decomposition** of the variance matrix of the VAR residuals to identify the **structural shocks** and then study their dynamic transmission on the variables x and y.
- Since the Cholesky decomposition is not unique, there is one for each possible ordering of the VAR variables, we have computed two sets of **Impulse response functions (IRFs)**.

VAR models with simulated data

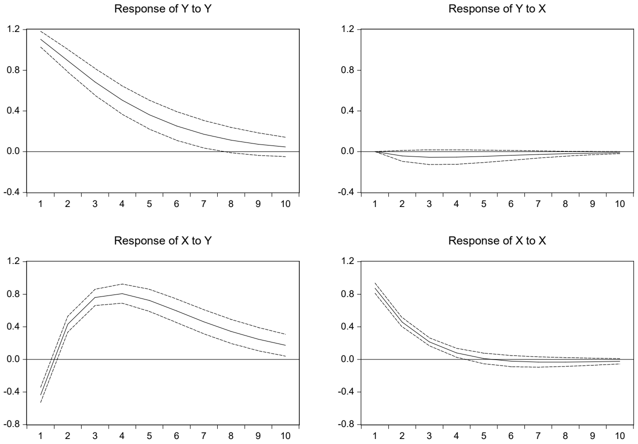

Figure 4: *Impulse response functions for the VAR(1) model. Response to Cholesky one s.d. innovations* $\pm$ *2 s.e. The values are computed using a* ***lower triangular mapping matrix P***

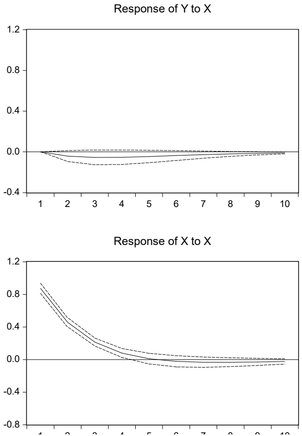

Figure 5: *Impulse response functions for the VAR(1) model. Response to Cholesky one s.d. innovations* $\pm$ *2 s.e. The values are computed using an* ***upper triangular mapping matrix P***

VAR models with simulated data

- The results are similar for both ordering of the variables: y does not seem to respond very much to structural shocks coming from x, while x instead is very sensitive to shocks coming from y.
- Furthermore, the figures provide additional evidence that both systems are **stationary**, as the IRFs tend to go to zero, as the impact of the shock vanishes after 10 periods.

Further, let us consider the **variance decomposition** of the two models,

- The variance decomposition splits the forecast error variance of each endogenous variable, at different forecast horizons, into the components due to each of the shocks, thereby providing information on their relative importance.
- Since the two types of Cholesky decompositions generate similar IRFs, we only consider the (y, x) ordering.

VAR models with simulated data

- For both models, shocks to y itself dominate both in the short and in the long-run. The shocks to x contribute much less in explaining y .
- Also for x its own shock dominates in the short run, but the effects of the y shock increase over time while those of x decrease, so that the two shocks have comparable importance at longer horizons.

VAR Models

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

GDP growth in the Euro area

- In this example we apply the VAR estimation and forecasting techniques to study the relationship among the GDP growth rates of three major countries in the Euro area, namely, France, Germany and Italy, using the United States as an exogenous explanatory variable.
- The data set contains 108 quarterly observations from 1986Q1 to 2012Q4. The series are all adjusted for seasonality and labeled as g_i where $i = fr, ger, it, us$.

GDP growth in the Euro area

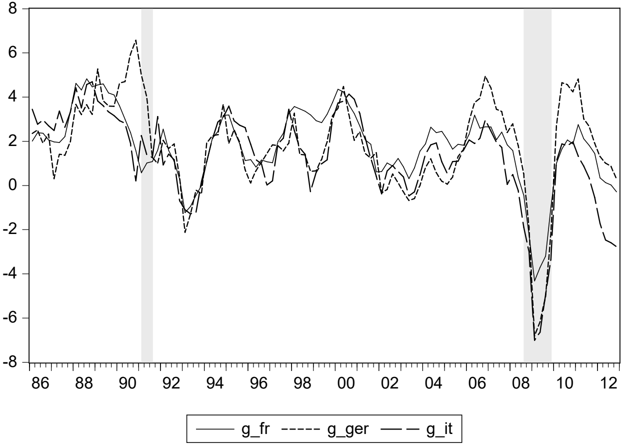

Figure 6: *Real GDP growth rates for France, Germany, and Italy during the period 1986Q1 - 2012Q4. The gray shaded areas indicate the NBER recessions for the United States*

GDP growth in the Euro area

- Preliminary analysis on the series confirmed that g_fr, g_us, and g_ger can be considered as stationary at the 5% confidence level, while g_it cannot, and it is probably characterized by a unit root.
- However, from an economic point of view, it is surprising that the growth rate is integrated, and therefore we will treat it anyway as stationary.

Let us start by selecting the appropriate **lag length** p for our VAR model. Table 9 reports the information criteria associated with the various lags of our VAR model.

VAR Lag Order Selection Criteria

Lag	*LogL*	*LR*	*FPE*	*AIC*	*SC*	*HQ*
0	-324.200	NA	3.034	9.623	9.721	9.662
1	-217.244	201.331	0.170	6.742	7.134*	6.898
2	-198.730	33.217*	0.129*	6.463*	7.148	6.734*
3	-193.902	8.233	0.146	6.585	7.564	6.973
4	-183.820	16.310	0.143	6.553	7.827	7.058

* Indicates lag order selected by the criterion. LR: sequential test (each at 5% level), FPE: final prediction error, AIC: Akaike information criterion, SC: Schwarz, HQ: Hannan-Quinn

Table 9: *Choice of the appropriate lag length p through information criteria*

- The AIC, the Hannan-Quinn criterion, the Final Prediction Error (FPE), and the LR criteria all suggest $p = 2$. while the more parsimonious BIC indicates $p = 1$.
- As a consequence, we will estimate two rival VAR models with, respectively, $p = 1$ and $p = 2$, and we will evaluate their relative performance.

GDP growth in the Euro area

Vector Autoregression Estimates

	G_(FR)	G_(GER)	G_(IT)
G_FR(-1)	0.668	0.007	0.063
	(0.074)	(0.158)	(0.135)
	[9.064]	[0.047]	[0.470]
G_GER(-1)	0.009	0.745	0.122
	(0.042)	(0.089)	(0.076)
	[0.214]	[8.352]	[1.606]
G_IT(-1)	0.107	0.091	0.623
	(0.057)	(0.121)	(0.104)
	[1.893]	[0.751]	[6.008]
G_US	0.172	0.105	0.099
	(0.030)	(0.065)	(0.055)
	[5.695]	[1.632]	[1.794]

Table10: *Estimation results of the VAR(1) for the period 1986 - 2003*

GDP growth in the Euro area

Vector Autoregression Estimates

	G_(FR)	G_(GER)	G_(IT)
D_1988Q1	1.068	0.838	1.213
	(0.301)	(0.644)	(0.550)
	[3.551]	[1.301]	[2.206]
D_1993Q1	-1.624	-2.141	-1.604
	(0.310)	(0.663)	(0.566)
	[-5.247]	[-3.229]	[-2.834]
D_2000Q1	0.711	0.723	1.079
	(0.304)	(0.651)	(0.556)
	[2.339]	[1.110]	[1.942]

Table10: *Estimation results of the VAR(1) for the period 1986 - 2003*

Vector Autoregression Estimates			
Standard errors in () and t-statistics in []			
R-squared	0.924	0.761	0.776
Adj. R-squared	0.917	0.739	0.755
Sum sq. resids	10.809	49.584	36.139
S.E.equation	0.411	0.880	0.751
F-statistic	129.747	33.982	37.044
Log likelihood	-33.923	-88.000	-76.771
Akaike AIC	1.153	2.676	2.360
Schwarz BIC	1.376	2.899	2.583
Mean dep	2.202	1.923	1.898
S.D.dependent	1.426	1.722	1.520
Determinant resid covariance (dof adj.)			0.066
Determinant resid covariance			0.048
Log likelihood			-194.456
Akaike information criterion			6.069
Schwarz criterion			6.738

Table 10: *Estimation results of the VAR(1) for the period 1986 - 2003*

GDP growth in the Euro area

- Note that both models include the same set of exogenous variables, namely, the growth rate of US GDP and the **dummy variables** necessary to remove some outliers (in 1988Q1, 1993Q1, and 2000Q1).
- Both VARs display a good overall performance and in general the exogenous variables we chose are significant.
- On the other hand, while in the VAR(1) model the significance of some lagged variables is very low, the same is generally not true for the VAR(2) model, suggesting that the former is probably mis-specified.

Tables 11 - 13 report results from the **diagnostic tests** for the VAR(1); results for the VAR(2) are not reported since they are very similar.

- The residuals of both models look quite "clean," although from their correlograms, it is even more evident that the VAR(2) captures some autocorrelation that the VAR(1) does not.
- There is no sign of heteroskedasticty or non-normality in any of the two models' residuals, and this is partly due to the dummies we have introduced.
- Note that the inverse of the AR roots are all well within the unit circle, supporting our choice of treating the growth rate of Italy as stationary.

GDP growth in the Euro area

Lags	LM-Stat	Prob
1	19.908	0.018
2	12.179	0.203
3	24.141	0.004
4	31.144	0.000
5	14.147	0.117
6	12.920	0.166
7	17.845	0.037
8	11.977	0.215
9	13.564	0.139
10	16.843	0.051
11	10.129	0.340
12	11.498	0.243

Table 11: *VAR(1) residuals serial correlation LM tests*

VAR Heteroskedasticity Tests:
Joint test:

Chi-sq	df	Prob.
72.340	66	0.277

Individualcomponents:

Dependent	R-squared	F(11,59)	Prob.	Chi-sq(11)	Prob.
res1*res1	0.113	0.684	0.748	8.031	0.711
res2*res2	0.174	1.132	0.354	12.373	0.336
res3*res3	0.132	0.812	0.628	9.339	0.591
res2*res1	0.203	1.365	0.214	14.406	0.211
res3*res1	0.130	0.804	0.636	9.252	0.599
res3*res2	0.300	2.302	0.020	21.321	0.030

Table 12: *VAR(1) residuals heteroskedasticity test*

VAR Residual normality Tests

Component	Skewness	Chi-sq	df	Prob.
1	-0.040	0.019	1	0.890
2	0.144	0.246	1	0.620
3	-0.333	1.311	1	0.252
Joint		1.576	3.000	0.665

Component	Kurtosis	Chi-sq	df	Prob.
1	2.541	0.623	1	0.430
2	3.543	0.873	1	0.350
3	4.112	3.661	1	0.056
Joint		5.157	3.000	0.161

Table13: *VAR(1) Residuals normality test*

VAR Residual normality Tests

Component	Jarque-Bera	df	Prob.
1	0.642	2	0.725
2	1.118	2	0.572
3	4.972	2	0.083
Joint	6.732	6	0.346

Table 13: *VAR(1) Residuals normality test*

GDP growth in the Euro area

- Table 14 reports results from **Granger causality test**, it seems that the shocks are not significantly dynamically transmitted from Germany to France and Italy, while the reverse is not true and there are cross-linkages across France and Italy.
- As the Granger causality test only considers lagged relationships, in order to better understand if there is contemporaneous correlation across the variables after conditioning on their lags, we can use IRF.

GDP growth in the Euro area

Pairwise Granger Causality Tests

Null Hypothesis:	Obs	F-Statistic	Prob.
G_GER does not Granger cause G_FR	70	0.605	0.549
G_FR does not Granger cause G_GER		4.449	0.016
G_IT does not Granger cause G_FR	70	6.403	0.003
G_FR does not Granger cause G_IT		6.275	0.003
G_IT does not Granger cause G_GER	70	6.511	0.003
G_GER does not Granger cause G_IT		0.780	0.463

Table 14: *Granger causality test among the three variables using the lag length* $p = 2$

GDP growth in the Euro area

- **IRF results** suggest that indeed the contemporaneous correlation is limited, in line with economic reasoning that suggests that it takes some time for the shocks to be transmitted across countries.
- The response functions also indicate that there are positive spillovers across all countries, some of which become significant only 2 - 3 quarters after the shock.

GDP growth in the Euro area

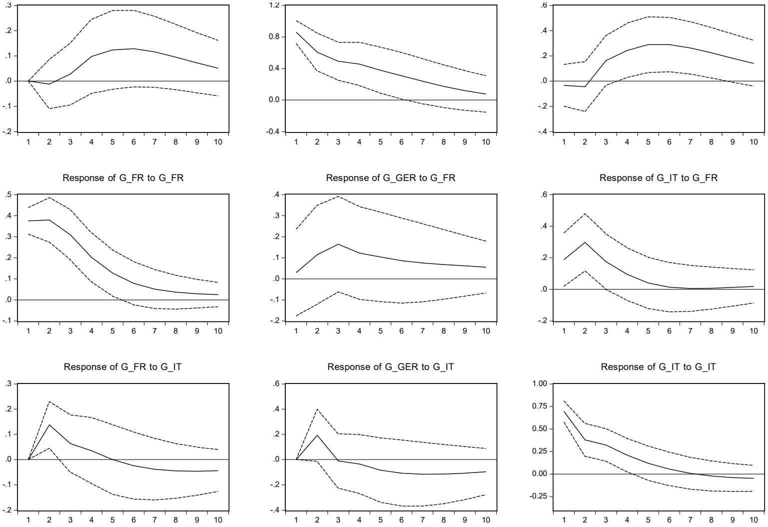

Figure 7: *Impulse response function from the VAR(1) model with the ordering (g_fr,g_ger,g_it)*

GDP growth in the Euro area

Which of the two estimated models (VAR(1) or VAR(2)) will perform better from a **forecasting** point of view?

We answer the question first for the forecast sample, 2004 - 2006. Then, we re-estimate the models until 2006 and use the second forecast sample, spanning from 2007 to 2012, to assess the robustness of the results and the performance of the models during problematic times.

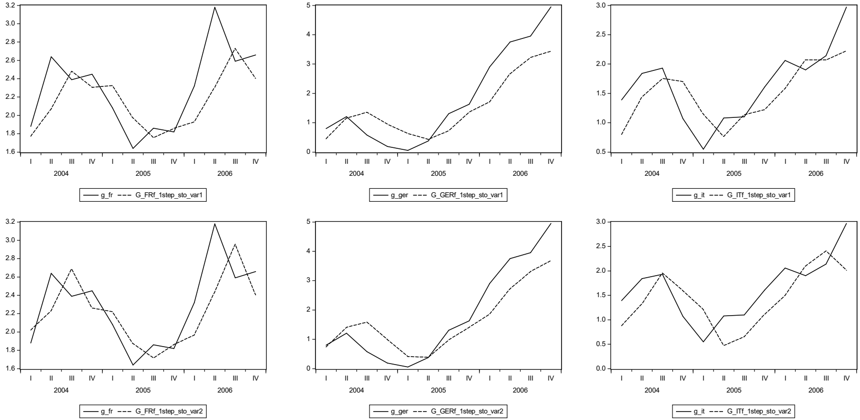

Figure 8: *One step ahead forecasts, stochastic method, from the VAR(1) model (upper panels) and the VAR(2) model (lower panels) in the* ***period 2004 - 2006***

GDP growth in the Euro area

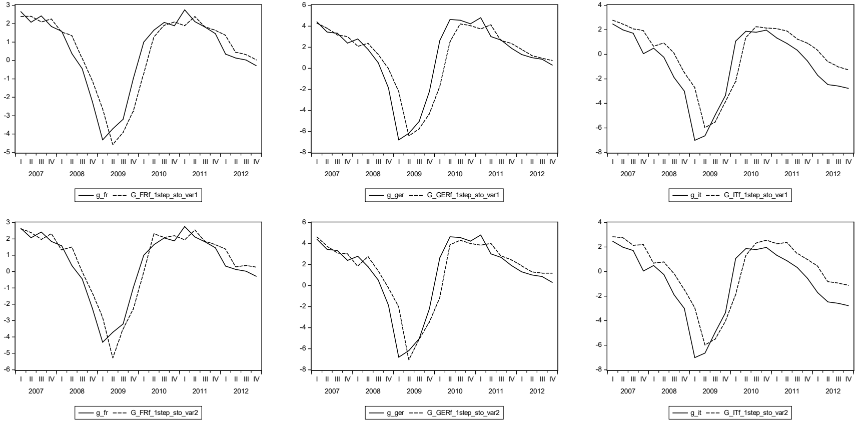

Figure 9: *One-step ahead forecasts, stochastic environment, from the VAR(1) model (lower panels) and the VAR(2) model (upper panels) in the* ***period 2007 - 2012***

GDP growth in the Euro area

- As we expected, re-estimating the two VAR models in the longer sample 1986 - 2006 and producing forecasts for the period 2007 - 2012 proves more challenging.
- In both evaluation samples VAR(2) model yields more accurate predictions than the VAR(1) for all countries but Italy.

Monetary transmission mechanism

- We can now analyze through a VAR model the transmission mechanism of monetary policy, describing how policy-induced changes in the nominal money stock or the short-term nominal interest rate impact real variables.
- We conduct an analysis along the lines of Leeper, Sims, and Zha(1996), and estimate a VAR model with three endogenous variables (ly, lp, i), over the period January 1986 to June 2007.
  - ly: the log of the index of industrial production, as an indicator of real activity.
  - lp: the log of the index of personal consumption expenditures less food and energy, as an indicator of inflationary pressure.
  - i: the federal funds rate, as an indicator of the monetary policy stance.

- We start by estimating a VAR model with a long **lag length** $p = 12$, and then we use standard lag selection criteria and sequential LR tests to reduce the number of lags.
- From Table 15, the most parsimonious specification is suggested by both the Hannan-Quinn and the BIC criterion, with $p = 2$.

VAR Lag Order Selection Criteria

Lag	*LogL*	*LR*	*FPE*	*AIC*	*SC*	*HQ*
0	28.608	NA	0.000	-0.199	-0.157	-0.182
1	2431.464	4731.205	0.000	-18.756	-18.590	-18.689
2	2480.569	95.546	0.000	-19.066	-18.777*	-18.950*
3	2488.235	14.738	0.000	-19.056	-18.643	-18.890
4	2500.958	24.164	0.000	-19.085	-18.548	-18.869
5	2514.379	25.178	0.000*	-19.119*	-18.458	-18.853
6	2522.058	14.227	0.000	-19.109	-18.324	-18.793
7	2531.022	16.399	0.000	-19.109	-18.200	-18.743
8	2533.830	5.070	0.000	-19.061	-18.028	-18.645
9	2536.803	5.301	0.000	-19.014	-17.857	-18.549
10	2548.146	19.961*	0.000	-19.032	-17.751	-18.517
11	2550.463	4.023	0.000	-18.980	-17.576	-18.416
12	2555.384	8.432	0.000	-18.949	-17.420	-18.334

* Indicates lag order selected by the criterion. LR: sequential test (each at 5% level), FPE: final prediction error, AIC: Akaike information criterion, SC: Schwarz, HQ: Hannan-Quinn

Table 15: *Lag length criteria for VAR for (ly, lp, i)*

Table 16 shows the resulting estimated coefficients for the VAR(2).

Vector Autoregression Estimates

	LP	LY	I
LP(-1)	1.011	0.106	17.505
	(0.062)	(0.289)	(10.318)
	[16.387]	[0.369]	[1.697]
LP(-2)	-0.013	-0.100	-17.779
	(0.062)	(0.288)	(10.303)
	[-0.212]	[-0.348]	[-1.726]
LY(-1)	-0.018	0.962	8.604
	(0.013)	(0.062)	(2.217)
	[-1.322]	[15.514]	[3.880]
LY(-2)	0.016	0.029	-8.347
	(0.013)	(0.062)	(2.207)
	[1.223]	[0.478]	[-3.783]

Table16: *VAR(2) model estimated for (ly, lp, i)*

Vector Autoregression Estimates			
	LP	LY	I
I(-1)	0.000	0.006	1.413
	(0.000)	(0.002)	(0.054)
	[1.389]	[3.803]	[25.978]
I(-2)	-0.000	-0.006	-0.425
	(0.000)	(0.002)	(0.054)
	[-1.189]	[-4.039]	[-7.807]
C	0.016	0.013	0.103
	(0.003)	(0.016)	(0.572)
	[4.647]	[0.813]	[0.180]

Table16: *VAR(2) model estimated for (ly, lp, i)*

Vector Autoregression Estimates			
Standard errors in () and t-statistics in []			
R-squared	1.000	0.999	0.993
Adj. R-squared	1.000	0.999	0.993
Sum sq. resids	0.000	0.007	8.679
S.E.equation	0.001	0.005	0.182
F-statistic	918307.100	65977.590	6367.156
Log likelihood	1451.289	1037.938	79.353
Akaike AIC	-10.778	-7.694	-0.540
Schwarz BIC	-10.684	-7.600	-0.446
Mean dep	4.412	4.312	5.081
S.D. dep	0.157	0.196	2.189
Determinant resid covariance(dof adj.)			0.000
Determinant resid covariance			0.000
Log likelihood			2573.232
Akaike information criterion			-19.047
Schwarz criterion			-18.765

Table 16: *VAR(2) model estimated for (ly, lp, i)*

- The model diagnostics in Figure 10 shows that the residuals display a few **outliers** and therefore some **non-normality problems.**
- More in depth tests on these residuals (not reported here) confirm the need to correct this non-normality problem through the use of **dummy variables.**

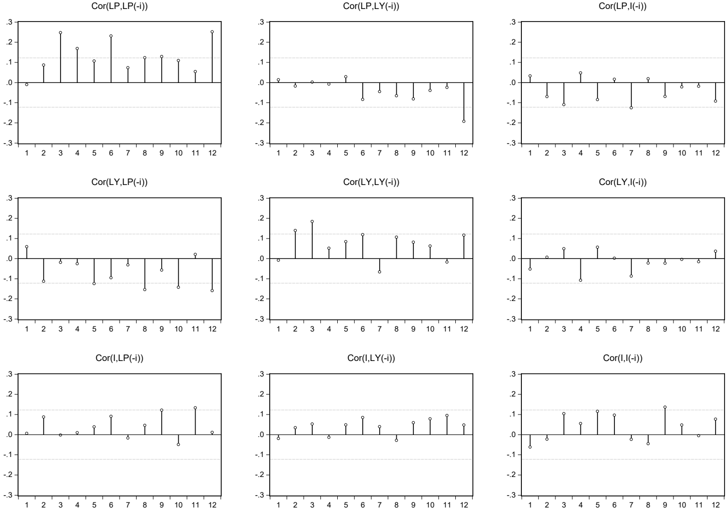

Figure 10: *Empirical correlograms of the residuals of the VAR(2) model*

Monetary transmission mechanism

- In particular, dummy variables need to be created for two months, September and October 2001, in correspondence of the terrorist attack at the World Trade Center in New York.
- Table 17 shows the resulting estimated coefficients for the VAR(2) with dummy variables.

Vector Autoregression Estimates

	LP	LY	I
LP(-1)	1.223	0.087	11.502
	(0.056)	(0.300)	(10.584)
	[21.665]	[0.289]	[1.087]
LP(-2)	-0.219	-0.076	-11.732
	(0.057)	(0.302)	(10.632)
	[-3.865]	[-0.250]	[-1.104]
LY(-1)	-0.016	0.960	8.246
	(0.012)	(0.062)	(2.188)
	[-1.404]	[15.455]	[3.769]
LY(-2)	0.013	0.029	-8.009
	(0.012)	(0.062)	(2.171)
	[1.091]	[0.471]	[-3.690]

Table17: *Estimated coefficient for the VAR(2) model for (ly,lp, i) including also dummy variables for 9/11 as exogenous*

Vector Autoregression Estimates

	LP	LY	I
I(-1)	0.001	0.006	1.400
	(0.000)	(0.002)	(0.054)
	[2.379]	[3.647]	[25.847]
I(-2)	-0.001	-0.006	-0.411
	(0.000)	(0.002)	(0.054)
	[-1.760]	[-3.809]	[-7.602]
D0109	-0.007	-0.005	-0.495
	(0.001)	(0.005)	(0.181)
	[-7.171]	[-0.936]	[-2.729]
D0110	0.007	-0.003	-0.236
	(0.001)	(0.006)	(0.199)
	[7.046]	[-0.612]	[-1.189]

Table17: *Estimated coefficient for the VAR(2) model for (ly,lp, i) including also dummy variables for 9/11 as exogenous*

Vector Autoregression Estimates			
R-squared	1.000	0.999	0.993
Adj.R-squared	1.000	0.999	0.993
Sum sq. resids	0.000	0.007	8.402
S.E.equation	0.001	0.005	0.180
F-statistic	1018404.000	56453.350	5616.953
Log likelihood	1486.321	1038.217	83.696
Akaike AIC	-11.032	-7.688	-0.565
Schwarz BIC	-10.925	-7.581	-0.458
Mean dep	4.412	4.312	5.081
S.D. dep	0.157	0.196	2.189
Determinant resid covariance (dof adj.)			0.000
Determinant resid covariance			0.000
Log likelihood			2611.999
Akaike information criterion			-19.313
Schwarz criterion			-18.992

Table 17: *Estimated coefficient for the VAR(2) model for (ly,lp, i) including also dummy variables for 9/11 as exogenous*

Monetary transmission mechanism

- The outliers related to the 9/11 period have been corrected, however, the normality test confirms that the null has to be rejected, and also the heteroskedasticity test signals some problems, see Tables 18 and 19.
- On top of that, some of the roots of the characteristic polynomial lie on the unit circle, and this is something that we expected given that we did not account for the non-stationarity of at least two of our indicators.

VAR Residual normality Tests

Component	Skewness	Chi-sq	df	Prob.
1	0.259	2.985	1	0.084
2	-0.112	0.565	1	0.452
3	-0.176	1.389	1	0.239
Joint		4.939	3	0.176
Component	**Kurtosis**	**Chi-sq**	**df**	**Prob.**
1	3.121	0.165	1	0.685
2	4.660	30.764	1	0.000
3	7.078	185.688	1	0.000
Joint		216.618	3	0.000
Component	**Jarque-Bera**	**df**	**Prob.**	
1	3.150	2	0.207	
2	31.329	2	0.000	
3	187.077	2	0.000	
Joint	221.556	6	0.000	

Table 18: *Diagnostic test on residuals of the VAR(2) model with dummy variables*

VAR Heteroskedasticity Tests:

Joint test:

Chi-sq	df	Prob.
144.113	84	0.000

Individual components:

Dependent	R-squared	F(14,253)	Prob.	Chi-sq(14)	Prob.
res1*res1	0.069	1.343	0.182	18.540	0.183
res2*res2	0.047	0.887	0.574	12.538	0.563
res3*res3	0.188	4.174	0.000	50.288	0.000
res2*res1	0.048	0.913	0.546	12.883	0.536
res3*res1	0.126	2.602	0.002	33.731	0.002
res3*res2	0.069	1.344	0.182	18.556	0.183

Table 19: *VAR(2) residuals heteroskedasticity tests*

- Let us have a look at the **impulse response functions** of the model up to 48 months (four years), obtained using the **Cholesky structural factorization.**

- The variables are ordered as lp, ly and i, so that:

  (1)price (demand) shocks can have a contemporaneous effect on output and interest rate,

  (2)output (supply) shocks can affect contemporaneously the interest rate, and

  (3)interest rate (monetary) shocks have only a delayed effect on output and prices, since it takes some time for both variables to adjust.

Monetary transmission mechanism

- Figure 11 displays the responses of lp, ly, and i to a unitary shock to i at time 0 (first three panels) and the response of the policy variable i to a unitary shock to lp and ly (last two panels).
- Coherently with the major findings of the MTM literature, we see that i increases after demand and supply shocks, though the reaction to higher prices is lower than to higher output and not statistically significant.

Monetary transmission mechanism

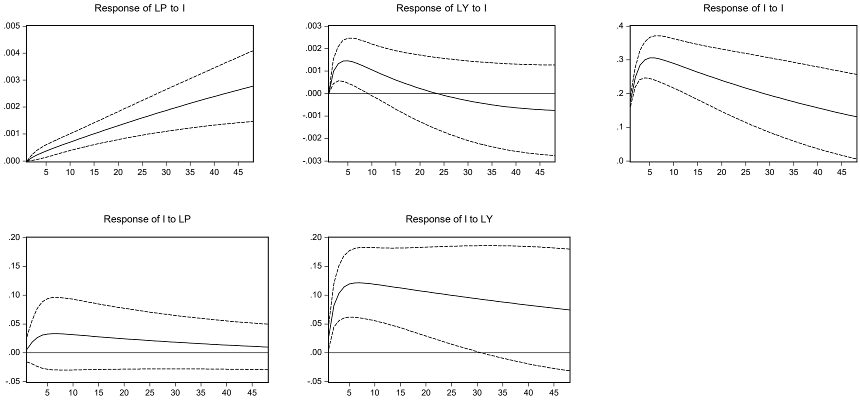

Figure 11: *MTM in the pre-crisis period: Impulse response function from the estimated VAR(2) model*

- However, in the first panels we see a **puzzling result**: after an increase in the interest rate (contradictory monetary shock), both prices and output increase.
- These puzzling results still persist even when the number of lags is increased as suggested by other information criteria.

The MTM literature in general has tried to solve puzzles like the ones we obtained by **including other indicators** in the VAR model, so as to eliminate a possible omitted variable bias.

- For example, the price puzzle we obtained might be due to the fact that **the model lacks a leading indicator for inflation**: for this reason we can try and add the producer price index including commodities $lpcm$.

- On the other hand, the fact that ly was increasing with higher interest rates could be due to the fact **the federal funds rate is not enough to capture the policy actions implemented by the Fed**: the literature has then suggested to include in MTM models also some reserve aggregates, which the Fed can control directly. We add to our indicators the borrowed ($lsmtr$) and non-borrowed reserves ($lsmnbr$).
- Besides the dummies we previously employed, the analysis of the residuals underlined the need to add also a **dummy** for November 2001 and one for January 1991.
- The chosen **lag length** this time is $p = 1$, and the results of the estimation are reported in Table 20.

Vector Autoregression Estimates

	LPCM	LP	LY	I	LSMTR	LSMNBR
LPCM(-1)	0.969	0.001	0.004	0.523	0.188	0.033
	(0.014)	(0.002)	(0.009)	(0.371)	(0.123)	(0.025)
	[70.946]	[0.911]	[0.385]	[1.410]	[1.529]	[1.302]
LP(-1)	0.040	1.000	-0.033	-2.633	-0.325	0.021
	(0.026)	(0.003)	(0.018)	(0.719)	(0.239)	(0.050)
	[1.514]	[325.470]	[-1.829]	[-3.661]	[-1.362]	[0.421]
LY(-1)	0.013	0.000	1.009	1.380	0.050	-0.030
	(0.012)	(0.001)	(0.008)	(0.324)	(0.107)	(0.022)
	[1.114]	[0.201]	[123.070]	[4.265]	[0.469]	[-1.347]
I(-1)	-3.300	7.250	-0.001	0.964	-0.004	-0.001
	(0.000)	(3.900)	(0.000)	(0.009)	(0.003)	(0.001)
	[-0.098]	[1.849]	[-2.219]	[104.960]	[-1.440]	[-1.952]

Table20: *Estimated coefficients for the VAR(1) model for the enlarged set of indicators*

Vector Autoregression Estimates

	LPCM	LP	LY	I	LSMTR	LSMNBR
LSMTR(-1)	0.007	0.001	-0.003	-0.026	0.985	0.010
	(0.002)	(0.000)	(0.001)	(0.057)	(0.019)	(0.004)
	[3.284]	[5.604]	[-1.994]	[-0.445]	[51.72]0	[2.501]
LSMNBR(-1)	-0.014	-0.002	0.012	0.334	0.044	0.980
	(0.005)	(0.001)	(0.004)	(0.141)	(0.047)	(0.010)
	[-2.666]	[-3.263]	[3.273]	[2.362]	[0.938]	[100.710]

Table20: *Estimated coefficients for the VAR(1) model for the enlarged set of indicators*

Vector Autoregression Estimates

	LPCM	LP	LY	I	LSMTR	LSMNBR
D0109	-0.004	-0.007	-0.005	-0.566	1.227	0.318
	(0.007)	(0.001)	(0.005)	(0.202)	(0.067)	(0.014)
	[-0.519]	[-8.038]	[-1.023]	[-2.807]	[18.351]	[22.938]
D0110	-0.030	0.005	-0.007	-0.671	-0.745	-0.189
	(0.007)	(0.001)	(0.005)	(0.205)	(0.068)	(0.014)
	[-3.942]	[5.501]	[-1.444]	[-3.277]	[-10.959]	[-13.380]
D9101	0.001	0.003	-0.005	-0.263	-0.311	-0.000
	(0.007)	(0.001)	(0.005)	(0.203)	(0.067)	(0.014)
	[0.117]	[2.883]	[-0.958]	[-1.294]	[-4.604]	[-0.029]
D0111	-0.009	0.001	-0.008	-0.432	-0.305	-0.104
	(0.007)	(0.001)	(0.005)	(0.203)	(0.067)	(0.014)
	[-1.220]	[0.587]	[-1.548]	[-2.134]	[-4.542]	[-7.462]

Table20: *Estimated coefficients for the VAR(1) model for the enlarged set of indicators*

Vector Autoregression Estimates						
Standard errors in () and t-statistics in []						
R-squared	0.997	1.000	0.999	0.992	0.987	0.994
Adj.R-squared	0.997	1.000	0.999	0.992	0.987	0.994
Sum sq. resids	0.014	0.000	0.007	10.416	1.147	0.049
S.E.equation	0.007	0.001	0.005	0.201	0.067	0.014
F-statistic	9823.380	100859.000	44771.000	3537.160	2188.280	4706.770
Log likelihood	943.714	1523.120	1044.470	55.611	352.355	775.555
Akaike AIC	-6.942	-11.250	-7.691	-0.339	-2.545	-5.692
Schwarz BIC	-6.808	-11.116	-7.558	-0.205	-2.412	-5.558
Mean dep	4.828	4.410	4.311	5.093	9.709	10.679
S.D. dep	0.134	0.158	0.197	2.195	0.574	0.174
Determinant resid covariance			0.000			
Determinant resid covariance			0.000			
Log likelihood			4730.441			
Akaike information criterion			-34.724			
Schwarz criterion			-33.923			

Table 20: *Estimated coefficients for the VAR(1) model for the enlarged set of indicators*

- Now after a unitary increase in the federal funds rate at time 0, the reaction of both price measures is close to zero and not significant, while there is a negative and significant effect on output.
- Furthermore, both types of reserves decline as we expect.
- We also note that none of the responses eventually goes to zero, a further confirmation of the presence of stochastic trends in the system, which makes the effects of the shocks persistent.

Representation

Specification of the model

Estimation

Diagnostic checks

Forecasting

Impulse response functions

Forecast error variance decomposition

Structural VARs with long-run restrictions

VAR models with simulated data

Empirical examples

Concluding remarks

Concluding remarks

VAR models are a powerful tool for forecasting using time series methods.

- From a statistical theoretical point of view, they can be considered as approximations to the infinite order multivariate MA representation of weakly stationary processes implied by the Wold theorem.
- From an economic point of view, they can capture the dynamic inter-relationship across the variables, without imposing (too many) restrictions, as often done in more structural models.
- In addition, estimation, testing, and forecasting are rather easily considered theoretically and implemented empirically.

Concluding remarks

However, three main caveats should also be kept in mind.

- First, typically there are many parameters in VARs, whose precise estimation in finite samples is problematic. Hence, just a few variables are typically modeled, which can create an omitted variable problem.
- Second, the relationships across variables can be unstable, as well as the variance of the shocks hitting them.
- Variables are not always weakly stationary as stochastic trends are important drivers of several economic and financial variables.