Vector Autoregressions I

Empirical Macroeconomics - Lect 1

Dr. Ana Beatriz Galvao

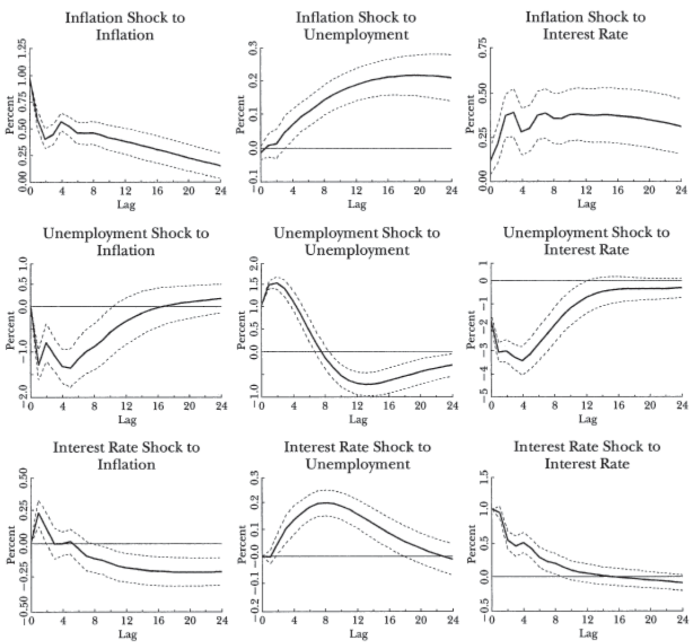
Queen Mary University of London

January 2012

Vector Autoregressions

"This simple framework provides a systematic way to capture rich dynamics in multiple time series, and the statistical toolkit that came with VARs was easy to use and interpret. As Sims (1980) and other argued, VARs held out the promise of providing a coherent and credible approach to ***data description, forecasting, structural inference and policy analysis****" (Stock and Watson, 2001).*

All empirical results of this lecture are from Stock and Watson (2001). US Data, 1960-2000, unemployment, inflation and fed rate.



Vector Autoregressive Model I

- A VAR(p) model represents the process of the $m \times 1$ vector of time series $y_t = (y_{1t}, y_{2t}, ..., y_{mt})$ with autoregressive order p:
$$y_t = c + A_1 y_{t-1} + ... + A_p y_{t-p} + \varepsilon_t$$
where A_i are $m \times m$ coefficients matrices and c is a $m \times 1$ of intercepts.
- ε_t is a $m \times 1$ vector of disturbances that have the following properties:
$$\begin{aligned} E(\varepsilon_t) &= 0 \text{ (mean zero)};\\ E(\varepsilon_t \varepsilon_t') &= \Sigma_\varepsilon \text{ (full variance-covariance matrix)};\\ E(\varepsilon_t \varepsilon_s') &= 0 \text{ for } s \neq t \text{ (no serial correlation)}.\end{aligned}$$

Vector Autoregressive Model II

- For example, for $p = 2$ and $m = 3$:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \begin{bmatrix} a_{2,11} & a_{2,12} & a_{2,13} \\ a_{2,21} & a_{2,22} & a_{2,23} \\ a_{2,31} & a_{2,32} & a_{2,33} \end{bmatrix} \begin{bmatrix} y_{1t-2} \\ y_{2t-2} \\ y_{3t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}$$

The implied dynamic mechanism of a VAR(1) I

- Suppose that $p = 1$, so $y_t = c + A_1 y_{t-1} + \varepsilon_t$, the dynamic mechanism starting at $t = 1$ up to $t = t$ is:

$$\begin{aligned} y_1 &= c + A_1 y_0 + \varepsilon_1 \\ y_2 &= c + A_1 y_1 + \varepsilon_2 = c + A_1(c + A_1 y_0 + \varepsilon_1) + \varepsilon_2 \\ &= (I_m + A_1)c + A_1^2 y_0 + A_1 \varepsilon_1 + \varepsilon_2 \\ &\vdots \\ y_t &= (I_m + A_1 + ... + A_1^{t-1})c + A_1^t y_0 + \textstyle\sum_{i=0}^{t-1} A_1^i \varepsilon_{t-i} \end{aligned}$$

- If all eigenvalues of A_1have module less than 1, then when $t \longrightarrow \infty$,

$$(I_m + A_1 + ... + A_1^{t-1})c \longrightarrow (I_m - A_1)^{-1} c = \mu,$$

$$A_1^t \longrightarrow 0,$$

and we can write the RHS sum as an infinite sum.

The implied dynamic mechanism of a VAR(1) II

- This means that the VAR(1) process can be represented as:

$$y_t = \mu + \textstyle\sum_{i=0}^{\infty} A_1^i \varepsilon_{t-i}$$

where $\mu = E(y_t)$ is the *unconditional mean* computed as $(I_m - A_1)^{-1} c$.

- y_t can be decomposed between a deterministic term (μ) and a stochastic term. The stochastic term is the weighted sum of all past disturbances.

The MA representation of a VAR(p): I

- Write a VAR(p) as:

$$y_t = c + (A_1 L + A_2 L^2 + ... + A_p L^p) y_t + \varepsilon_t.$$

or:

$$\begin{aligned} A(L) &= I_m - A_1 L - A_2 L^2 - ... - A_p L^p \\ A(L) y_t &= c + \varepsilon_t. \end{aligned}$$

- Now define a operator $\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i$ such that:

$$\Phi(L) A(L) = I_m$$

that is, $\Phi(L) = A(L)^{-1}$. This operator only works if the inverse of $A(L)$ exists that requires that $\det(I_m - A_1 z - ... - A_p z^p) \neq 0$, which also implies that the VAR is *stable*.

The MA representation of a VAR(p): II

- Multiplying the previous VAR(p) representation by $\Phi(L)$:

$$\begin{aligned} y_t &= \Phi(L)c + \Phi(L)\varepsilon_t \\ &= \left(\sum_{i=0}^{\infty}\Phi_i\right)c + \sum_{i=0}^{\infty}\Phi_i\varepsilon_{t-i} \end{aligned}$$

  which is the MA(∞) representation of a VAR(p).
- Algebra can be used to show that how to compute $\Phi(L)$ out of $A(L)$:

$$\begin{aligned} \Phi_0 &= I_m \\ \Phi_i &= \sum_{j=1}^{i}\Phi_{i-j}A_j \\ \mu &= (I_m - A_1 - ... - A_p)^{-1}c. \end{aligned}$$

- And that the MA representation can be simplified to:

$$y_t = \mu + \varepsilon_t + \Phi_1\varepsilon_{t-1} + \Phi_2\varepsilon_{t-2} + ...$$

The MA representation of a VAR(p): III

- For example, suppose a VAR with $p = 2$ and $m = 2$:

$$y_t = c + \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} y_{t-2} + \varepsilon_t.$$

- The coefficients of the MA representation are:

$$\begin{aligned} \Phi_1 &= A_1 = \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} \\ \Phi_2 &= \Phi_1A_1 + A_2 = A_1^2 + A_2 = \begin{bmatrix} .29 & .1 \\ .65 & .29 \end{bmatrix} \\ \Phi_3 &= \Phi_2A_1 + \Phi_1A_2 = A_1^3 + A_2A_1 + A_1A_2 = \begin{bmatrix} .21 & .079 \\ .566 & .21 \end{bmatrix} \\ &\vdots \\ \Phi_i &= \Phi_{i-1}A_1 + \Phi_{i-2}A_2 \end{aligned}$$

Wold Decomposition I

Macroeconomists like to use VARs to represent the empirical joint empirical process of macroeconomic time series. An important support for this is the Wold decomposition.

- The stochastic process of y_t is (covariance) stationary if its mean and variance do not depend on time, that is, they are constant over time.
- A stable VAR(p) process is stationary.

Any stationary process x_t can be written as a sum of a **deterministic** (perfectly predictable) and a **news component**, that is:

$$y_t = \kappa_t + \sum_{i=0}^{\infty}\Phi_i\varepsilon_{t-i},$$

where ε_t is white noise and it is called news because it is the error in forecasting y_t using all past information up to $t-1$:

$$\varepsilon_t \equiv y_t - E(y_t|y_{t-1}, ...).$$

Wold Decomposition II

- When using a VAR(p), the disturbances can be also interpreted as forecast errors, that is, they can also be interpreted as news.
- This means that the VAR(p) is way of obtaining the Wold decomposition of a time series.
- Note that ε_t has m processes that are *contemporaneous correlated*. This means that to name/identify a specific news (shock), the disturbances need to be transformed such that they are not contemporaneously correlated.

The Impulse-Response Function I

- The coefficients of the MA(∞) representation of the VAR(1) have the interesting interpretation:

$$\frac{\delta y_{t+s}}{\delta \varepsilon_t} = \Phi_s,$$

that is, $\phi_{s,ij}$ is the element of Φ_s that measures the effect of one unit increase in the j^{th} variable's news at date t for the value of the i^{th}variable at date $t+s$ for all $s \geq 0$.

- If $m = 2$, the effect of a shock on ε_{2t} at y_{1t+3} is $\phi_{3,12}$. If one wants to compute the cumulative effect up to $t+3$: $\phi_{0,12} + \phi_{1,12} + \phi_{2,12} + \phi_{3,12}$.

The Impulse-Response Function II

The problem is that ε_t, which are the disturbances of the reduced-form VAR, are contemporaneously correlated. We are normally interested in shocks that are orthogonal to each other. For uncorrelated shocks, we can compute the effect of unexpected changes (*news*) in the variable $y_{j,t}$ on the vector y_{t+s} computed conditional to all information up to t, that is, the effect of changes in $y_{j,t}$ on $E(y_{t+s}|y_{jt}, y_{j-1t}, y_{1t}, y_{t-1}, ...)$ for $s \geq 0$.

- A matrix P can be used to create shocks v_t from ε_t such that:

$$v_t \equiv P^{-1}\varepsilon_t$$

where P is a $(m \times m)$ lower triangular matrix with positive elements in the diagonal.

The Impulse-Response Function III

- Matrix P is computed using the Cholesky decomposition of $\Sigma_\varepsilon = PP'$. The matrix P can be decomposed as:

$$P \equiv AD^{1/2}$$

where D is a diagonal matrix that has the values of the diagonal of P, and A is a lower triangular matrix that has 1s in the diagonal and the remaining elements of P.

- The structural shocks u_t are defined as:

$$u_t \equiv A^{-1}\varepsilon_t.$$

- By definition $\Sigma_\varepsilon = ADA'$, so we can show that:

$$var(u_t) = E(u_t u_t') = D,$$

that is, the shocks u_t are not contemporaneous correlated and their variances are in the diagonal of matrix of matrix D. Note also that $var(v_t) = I_m$.

The Impulse-Response Function IV

- Using $m = 3$, we can explicitly write:

$$\begin{aligned} A &= \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \\ D^{1/2} &= \begin{bmatrix} \sigma_{u_1} & 0 & 0 \\ 0 & \sigma_{u_2} & 0 \\ 0 & 0 & \sigma_{u_3} \end{bmatrix} \\ D &= diag(\sigma^2_{u_1}, \sigma^2_{u2}, \sigma^2_{u3}) \\ P &= \begin{bmatrix} \sigma_{u_1} & 0 & 0 \\ a_{21}\sigma_{u_1} & \sigma_{u_2} & 0 \\ a_{31}\sigma_{u_1} & a_{32}\sigma_{u_2} & \sigma_{u_3} \end{bmatrix} \end{aligned}$$

where P comes from the Cholesky decomposition of

$$\Sigma_\varepsilon = \begin{bmatrix} \sigma^2_{\varepsilon_1} & cov(\varepsilon_1, \varepsilon_2) & cov(\varepsilon_1, \varepsilon_3) \\ cov(\varepsilon_1, \varepsilon_2) & \sigma^2_{\varepsilon_2} & cov(\varepsilon_2, \varepsilon_3) \\ cov(\varepsilon_1, \varepsilon_3) & cov(\varepsilon_2, \varepsilon_3) & \sigma^2_{\varepsilon_3} \end{bmatrix}.$$

The Impulse-Response Function V

Recursive VAR I

- To understand better the difference between the structural shocks u_t and the reduced form shocks ε_t, write ($m = 3$):

$$\begin{aligned} Au_t &= \varepsilon_t \\ \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} &= \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}. \end{aligned}$$

- The implied equations (that allow the computation of are $u_t s$) are:

$$\begin{aligned} u_{1t} &= \varepsilon_{1t} \\ u_{2t} &= \varepsilon_{2t} - a_{21}u_{1t} \\ u_{3t} &= \varepsilon_{3t} - a_{31}u_{1t} - a_{32}u_{2t}. \end{aligned}$$

Recursive VAR II

- These equations show that this recursive identification implies that y_{1t}, the variable of the first equation, is not affected by shocks at other equations at time t (recall $\Phi_0 = I_m$). The variable of the third equation is affected by shocks on y_{1t} at time t (impact of a_{31} on y_{3t}) and shocks on y_{2t} (impact of a_{32}). While the second variable is affected by shocks at y_{1t}.

Impulse responses of the recursive VAR I

- The response of one-unit shock on u_{jt} on y_{t+s} (the response measures changes in the predictive values of y_t at horizon s) is:

$$\Phi_s \mathbf{a}_j$$

where $\mathbf{a}_j$ is a vector taking from column j of matrix A.
- The effect of a shock on y_{1t} on y_{2t} is:

$$\begin{aligned} t &= 0; a_{21} \\ t &= 1; \phi_{1,21}a_{21} \\ t &= 2; \phi_{2,21}a_{21} \\ &\vdots \end{aligned}$$

Impulse responses of the recursive VAR II

- The problem with unit responses is that depending on the scale of the variable, one-unit (1%, 1 pound) represents a different sized shock. A solution is the use of one-standard deviation shock:

$$\Phi_s \mathbf{a}_j \sqrt{var(u_j)} = \Phi_s \mathbf{p}_j$$

where $\mathbf{p}_j$ is a column taking from the matrix P (Cholesky decomposition).

B. Variance Decompositions from the Recursive VAR Ordered as π, u, R

B.i. Variance Decomposition of π

Forecast Horizon	Forecast Standard Error	Variance Decomposition (Percentage Points)		
		π	u	R
1	0.96	100	0	0
4	1.34	88	10	2
8	1.75	82	17	1
12	1.97	82	16	2

B.ii. Variance Decomposition of u

Forecast Horizon	Forecast Standard Error	Variance Decomposition (Percentage Points)		
		π	u	R
1	0.23	1	99	0
4	0.64	0	98	2
8	0.79	7	82	11
12	0.92	16	66	18

B.iii. Variance Decomposition of R

Forecast Horizon	Forecast Standard Error	Variance Decomposition (Percentage Points)		
		π	u	R
1	0.85	2	19	79
4	1.84	9	50	41
8	2.44	12	60	28
12	2.63	16	59	25

Forecast errors I

- VAR(p) forecasts can be computed as:

$$\begin{aligned}
\hat{y}_{t+1} &= c + A_1 y_t + \ldots + A_p y_{t-p-1} \\
\hat{y}_{t+2} &= c + A_1 \hat{y}_{t+1} + \ldots + A_p y_{t-p-2} \\
&\vdots \\
\hat{y}_{t+s} &= c + A_1 \hat{y}_{t+s-1} + \ldots + A_p \hat{y}_{t-s-p}.
\end{aligned}$$

- The sum of the forecasts errors for each step from 1 to s are equivalent to the MA specification of the VAR, that is:

$$y_{t+s} - \hat{y}_{t+s} = \varepsilon_{t+s} + \Phi_1 \varepsilon_{t+s-1} + \Phi_2 \varepsilon_{t+s-2} + \ldots + \Phi_{s-1} \varepsilon_{t+1}.$$

Forecast errors II

- The mean squared error of using $\hat{y}_{t+s}$ as forecast is:

$$\begin{aligned}
MSE(\hat{y}_{t+s}) &= E[(y_{t+s} - \hat{y}_{t+s})(y_{t+s} - \hat{y}_{t+s})'] \\
&= E[(\varepsilon_{t+s} + \Phi_1 \varepsilon_{t+s-1} + \ldots + \Phi_{s-1} \varepsilon_{t+1}) \\
&\quad (\varepsilon_{t+s} + \Phi_1 \varepsilon_{t+s-1} + \ldots + \Phi_{s-1} \varepsilon_{t+1})'] \\
&= \Sigma_\varepsilon + \Phi_1 \Sigma_\varepsilon \Phi_1' + \Phi_2 \Sigma_\varepsilon \Phi_2' + \ldots + \Phi_{s-1} \Sigma_\varepsilon \Phi_{s-1}'
\end{aligned}$$

Variance Decomposition I

- Similarly to impulse responses, the variance decomposition is the effect of a structural shock on the total variance of y_{t+s}, or the contribution of u_j for the variation of y_{t+s} for values of $s \geq 0$
- We can write the $MSE(\hat{y}_{t+s})$ using u_t instead of ε_t.
- Recall that:

$$\varepsilon_t = Au_t = \mathbf{a}_1 u_{1t} + \mathbf{a}_2 u_{1t} + ... + \mathbf{a}_m u_{mt}$$

  where, as before, $\mathbf{a}_j$ is a vector taking from column j of matrix A.
- Using the previous expression:

$$\Sigma_\varepsilon = \mathbf{a}_1\mathbf{a}_1' var(u_{1t}) + \mathbf{a}_2\mathbf{a}_2' var(u_{2t}) + ... + \mathbf{a}_m\mathbf{a}_m' var(u_{mt}).$$

Variance Decomposition II

- Using this representation, we can compute the $MSE(\hat{y}_{t+s})$ using u_{jt} :

$$\begin{aligned} MSE(\hat{y}_{t+s}) &= \textstyle\sum_{j=1}^{m}\{var(u_{jt})[\mathbf{a}_j\mathbf{a}_j' + \Phi_1\mathbf{a}_j\mathbf{a}_j'\Phi_1' + \\ &\quad \Phi_2\mathbf{a}_j\mathbf{a}_j'\Phi_2' + ... + \Phi_{s-1}\mathbf{a}_j\mathbf{a}_j'\Phi_{s-1}']\}. \end{aligned}$$

- If we want to see the effect of a specific shock j on the total variance of y_{t+s}, we can use:

$$\begin{aligned} &var(u_{jt})[\mathbf{a}_j\mathbf{a}_j' + \Phi_1\mathbf{a}_j\mathbf{a}_j'\Phi_1' + \Phi_2\mathbf{a}_j\mathbf{a}_j'\Phi_2' \\ &+ ... + \Phi_{s-1}\mathbf{a}_j\mathbf{a}_j'\Phi_{s-1}']. \end{aligned}$$

- Econometric software normally computes the proportion of the forecast error variance $MSE(\hat{y}_{t+s})$ that is explained by each of the structural shocks (u_{jt} for $j = 1, ..., m$) for each horizon s.

Variance Decomposition III

- For example, if $m = 3$, the proportion of the variation of y_{2t} explained by u_{1t} is:

$$\begin{aligned} t &= 0; a_{21}^2(var(u_{1t}))/MSE(\hat{y}_{1t}) \\ t &= 1; (\phi_{1,21}a_{21})^2(var(u_{1t}))/MSE(\hat{y}_{1t+1}) \\ t &= 2; (\phi_{2,21}a_{21})^2(var(u_{1t}))/MSE(\hat{y}_{1t+2}) \\ &\vdots \end{aligned}$$