Bayesian Estimation of Structural Models

I: Bayesian VARs and SVARs

Gianni Amisano

University of Brescia

European Central Bank

amisano@eco.unibs.it

Vienna, 23/04/08
Contents

1 The multivariate LRM 6

1.1 The Wishart distribution 7

1.2 A Prior for B and H (I) 8

1.3 A more convenient prior for B and H 8

2 VARs 10

3 ML estimation of a VAR model 13
4 VMA representation

4.1 Impulse Response Functions ........................................... 19

4.2 How to choose $C$? .......................................................... 24

5 Structural VAR models .................................................. 30

5.1 Examples of S-VARs ...................................................... 35

5.1.1 Cholesky factorisation .................................................. 35

5.1.2 Long-run neutrality of demand shocks .......................... 38

5.1.3 AS-AD model ............................................................ 43
6 Bayesian estimation of VARs  

7 Bayesian estimation of exactly identified SVARs  

8 Example I: a trivariate model for identifying monetary shocks  

9 Example II: long run restrictions on New Keynesian model of labour market  

9.1 Long run coefficients  

9.2 Impact coefficients  

9.3 Identification via long run restrictions
10 "Identification" via sign restrictions

10.1 Bayesian estimation algorithm ....................... 60
10.2 How to find Q .................................................. 60
10.3 Example: Dedola and Neri (2007) .................... 61
1 The multivariate LRM

Model is:

\[
Y_{(T \times n)} = X_{(T \times k)}B_{(k \times n)} + E_{(T \times n)} \tag{1}
\]

\[
\Rightarrow y = (I_n \otimes X) \beta + \varepsilon \tag{2}
\]

\[
y = \text{vec}(Y), \beta = \text{vec}(B), \varepsilon = \text{vec}(E)
\]

\[
\varepsilon \sim N(0, H \otimes I_T)
\]

Log-Likelihood reads

\[
\ln L(Y|\theta) = -\frac{Tn}{2} \ln(2\pi) + \frac{T}{2} \ln |H| - \frac{1}{2} tr \left[ E'EH \right] \tag{3}
\]
1.1 The Wishart distribution

(see Zellner, 1971, appendix B)

A Wishart distribution for $\mathbf{H}$, an $(n \times n)$ symmetric and positive definite random matrix, is:

$$ p(\mathbf{H}|\nu, S) = W_n(\nu, S) = c|\mathbf{H}|^{\nu-(n+1)/2} \cdot \exp[-\text{tr}(\mathbf{HS}^{-1})] \quad (4) $$

$$ c = \frac{1}{|S|^{\nu/2} \Gamma_n(\nu)} \cdot \Gamma_n(\nu) = \pi^{n(n-1)/4} \prod_{i=1}^{n} \Gamma\left(\frac{2\nu + 1 - i}{2}\right) $$

$$ E(h_{ij}) = \nu s_{ij}, \quad V(h_{ij}) = \nu \left(s_{ij}^2 + s_{ii}s_{jj}\right) $$

can use this distribution on $\Sigma = \mathbf{H}^{-1}$
1.2 A Prior for B and H (I)

1.3 A more convenient prior for B and H

(natural conjugate)

Use

\[ p(\beta, H) = p(\beta | H) \cdot p(H) \]  \hspace{1cm} (5)

with

\[ p(H) = W_n(\nu, S) \]  \hspace{1cm} (6)

\[ p(\beta | H) = N(m_B, H^{-1} \otimes H_B^{-1}) \]  \hspace{1cm} (7)

\[ m_B = \text{vec}(M_B) \]  \hspace{1cm} (k \times n)  

(8)
With some algebra it is possible to show that:

\[ p(\beta | H, Y) = N(\overline{m}_B, H^{-1} \otimes (X'X + H_B)^{-1}) \]  \hspace{1cm} (9)

\[ \overline{m}_B = \text{vec}(\overline{M}_B), \]  \hspace{1cm} (10)

\[ \overline{M}_B = (X'X + H_B)^{-1} (X'Y + H_B M_B) \]  \hspace{1cm} (11)

and

\[ p(H|\beta, Y) = W_n(\overline{\nu}, \overline{S}) \], \hspace{1cm} (12)

\[ \overline{\nu} = T + \nu, \]  \hspace{1cm} (13)

\[ \overline{S} = [S^{-1} + (B - \overline{M}_B)' H_B (B - \overline{M}_B) + E'E]^{-1} \]  \hspace{1cm} (14)
2 VARs

Particularly important in macro applications for:

- forecasting tools

- answer question: what is the dynamic response of macro variables to exogenous shocks?

- answer question: what is the relative importance of different shocks in determining the behaviour of variables at different horizons (supply/demand shocks?)

Basic references on VARs:
A \textit{VAR}(p) model:

\begin{align*}
y_t &= A_1y_{t-1} + A_2y_{t-2} + \ldots + A_py_{t-p} + \varepsilon_t, \quad (15) \\
\varepsilon_t &\sim VWN(0, \Sigma) \quad (16)
\end{align*}
or:

\[ A(L)y_t = \varepsilon_t \]  

\[ A(L) = I_n - A_1 L - A_2 L^2 - ... - A_p L^p \]  

\( A_h, h = 1, 2, .., p \), are \((n \times n)\) coefficient matrices, in which \( a_{ij,h} \), is coefficient on \( y_{jt-h} \) in the i-th equation (in which \( y_{it} \) is dep.var.)

Stationarity condition: roots of:

\[ |A(L)| = 0 \]  

must lie outside unit circle. (Multivariate generalisation of univariate stationarity condition).
3 ML estimation of a VAR model

If errors are Gaussian (beside VWN), then we can write the (conditional upon initial observations) log-likelihood as:

\[
\ln L = -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} |\Sigma| \\
- \frac{1}{2} \text{tr} \left[ E'E\Sigma^{-1} \right] \\
= -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} |\Sigma| \\
- \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t'\Sigma^{-1}\varepsilon_t
\]

(20)
which can be maximised analytically yielding:

\[ \hat{\beta}_{ML} = \hat{\beta}_{OLS} \]  
\[ \hat{\Sigma}_{ML} = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}' \hat{\epsilon} = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t' \]  

\section{VMA representation}

If model is stationary, we can obtain stationary \( VMA(\infty) \) representation:

\[ y_t = B(L) \varepsilon_t, \]  
\[ B(L) = I_n + B_1 L + B_2 L^2 + ... = [A(L)]^{-1} \]

How can we obtain these VMA coefficient matrices? Two alternative ways:
1. By solving

\[ A(L) \cdot B(L) = I_n \]  \hspace{1cm} (26)

2. Use state space representation (companion form representation) of VAR model:

\[ z_t = M z_{t-1} + \eta_t \]  \hspace{1cm} (27)

\[ z_t = \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \end{bmatrix} (np \times 1), \hspace{1cm} M = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_n & [0] & \cdots & [0] & [0] \\ & I_n & \cdots & [0] & [0] \\ & & \cdots & \cdots & \cdots \\ & & & [0] & [0] \\ & & & (n \times n) & (n \times n) \end{bmatrix} \]  \hspace{1cm} (28)
\[ \eta_t = \begin{bmatrix} \varepsilon_t \\ [0] \\ (n \times 1) \\ [0] \\ (n \times 1) \\ \vdots \\ [0] \\ (n \times 1) \end{bmatrix} = \mathbf{J} \varepsilon_t, \quad \mathbf{J} = \begin{bmatrix} \mathbf{I}_n \\ [0] \\ (n \times n) \\ [0] \\ (n \times 1) \\ \vdots \\ [0]' \\ (1 \times n) \end{bmatrix} \] (29)

\[ y_t = \mathbf{J}' \mathbf{z}_t \] (30)

By recursive substitution we obtain the VMA representation as follows:

\[ y_t = \sum_{i=0}^{\infty} \mathbf{J}' \mathbf{M}^i \mathbf{J} \varepsilon_{t-i} = \mathbf{B}(L) \varepsilon_t, \] (31)

\[ \mathbf{B}(L) = \sum_{i=0}^{\infty} \mathbf{B}_i L^i, \mathbf{B}_i = \mathbf{J}' \mathbf{M}^i \mathbf{J}, i = 0, 1, 2, \ldots \] (32)
whereas (theoretical) forecasts and forecast errors are respectively:

\[
\begin{align*}
    y_{t+h|t} &= J' z_{t+h|t} = J' M^h z_t = \\
    &= \sum_{i=0}^{\infty} J' M^{h+i} J \epsilon_{t-i} = \sum_{i=h}^{\infty} B_i \epsilon_{t+h-i} \\
    e_{t+h|t} &= y_{t+h} - y_{t+h|t} = J' \left( z_{t+h} - z_{t+h|t} \right) = \\
    &= \sum_{i=0}^{h-1} J' M^i J \epsilon_{t+h-i} = \sum_{i=0}^{h-1} B_i \epsilon_{t+h-i}
\end{align*}
\]
Hence:

\[
E \left( e_{t+h|t} \right) = [0], \forall t, h, \quad (35)
\]

\[
V \left( e_{t+h|t} \right) = \sum_{i=0}^{\infty} B_i \Sigma B_i', \quad (36)
\]

\[
Cov \left( e_{t+h|t}, e_{t+h-j|t-j} \right) = E \left( e_{t+h|t} e_{t+h-j|t-j}' \right) = \\
= \sum_{i=j}^{h-1} B_i \Sigma B_{i-j}', \forall t, |j| \leq h - 1 \quad (37)
\]

\[
= [0] , \forall t, |j| > h - 1 \quad (38)
\]

\( (n \times n) \)

Same as in univariate case.

The big problem is parameter uncertainty
4.1 Impulse Response Functions

From stationary $VMA(\infty)$ representation:

$$y_t = B(L)\varepsilon_t, \quad B(L) = I_n + B_1L + B_2L^2 + ... = [A(L)]^{-1}$$ (39) (40)

we can obtain dynamic response of $y_t$ wrt shocks (IRF=impulse response functions):

$$\frac{\partial y_{it+h}}{\partial \varepsilon_{jt}} = b_{ij}^{(h)} \quad (41)$$

We know that under stationarity:

$$\lim_{h \to \infty} \frac{\partial y_{it+h}}{\partial \varepsilon_{jt}} = 0, \forall i, j \quad (42)$$

no long term effects
But:

- what are these shocks $\varepsilon_{jt}$?

- does it make sense to compute isolated effects of contemporaneously correlated shocks? ($\Sigma$ not diagonal)

If we could define $(n \times n)$ matrix $C$ such that:

$$C e_t = \varepsilon_t \quad (43)$$

$$e_t \sim VWN(0, I_n),$$

hence:

$$CC' = \Sigma \quad (44)$$
Then we could obtain **orthogonalised VMA representation**:

\[
y_t = B(L)\varepsilon_t = B(L)Ce_t = C(L)e_t \tag{45}
\]

\[
C(L) = C_0 + C_1L + C_2L^2 + ... \tag{46}
\]

\[
C_i = B_i \cdot C, i = 0, 1, 2, ... \tag{47}
\]

with orthogonalised IRFs:

\[
\frac{\partial y_{it+k}}{\partial e_{jt}} = c_{ij,h} = (i, j) \text{ element of } di \quad C_h = B_h \cdot C \tag{48}
\]

Another very important tool: **FEVD (Forecast Error Variance Decomposition)** or **IA (Innovation Accounting)** coefficients at different horizons. This tool allows us to gauge the relative importance of each element of \( e_t \) in determining the behaviour of the single elements of \( y_{t+h} (h = 1, 2, 3, ...) \).

Start from orthogonalised VMA representation:

\[
y_{t+h}|_t = E(y_{t+h} \mid I_t) = \\
= E \left[ \left( \sum_{k=0}^{\infty} C_k e_{t+h-k} \right) | I_t \right] = \sum_{k=h}^{\infty} C_k e_{t+h-k} \tag{49}
\]
where $I_t$ is information set available at $t$; the h-step ahead forecast error is:

$$e_{t+h|t} = y_{t+h} - y_{t+h|t} = \sum_{k=0}^{h-1} C_k e_{t+h-k}$$

with covariance matrix:

$$V(e_{t+h|t}) = \sum_{k=0}^{h-1} C_k C_k' = \Omega_h$$

Diagonal elements are variances:

$$\omega_{ii,h} = u_i^n' \Omega_h u_i^n = \sum_{k=0}^{h-1} \sum_{l=1}^n c^2_{il,k}$$

$u_i^n =$ i-th column of $I_n$.

Spse all shocks different from $e_{jt}$ are identically zero over the forecasting horizon:

$$e_{lt+1} = e_{lt+2} = \ldots = e_{lt+k} = 0, \forall l \neq j$$
i.e.: assume (counterfactually):

$$E(e_t'e_t) = \begin{bmatrix}
0 & \ldots & 0 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 1 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 0 & \ldots & 0 \\
\end{bmatrix}$$

(all elements equal to zero but the i-th diagonal element). In this case:

$$\omega_{ii,h}^* = \sum_{k=0}^{h-1} c_{ij,k}^2$$  \hspace{1cm} (54)

Construct relative importance index (of shocks $e_j$) in determining the behaviour of $y_{jt+h}$:
\[ F E V D(i, j, k) = \frac{\omega_{ii,k}^*}{\omega_{ii,k}} = \frac{\sum_{k=0}^{h-1} c_{ij,k}^2}{\sum_{k=0}^{n} \sum_{l=1}^{n} c_{il,k}^2} \leq 1 \] (55)

These coefficients are bound to lie within the [0,1] interval. They are non-linear functions of VAR parameters (like VMA and orthogonalised VMA parameters).

### 4.2 How to choose \( C \)?

Possible and simple choice:

\[ C = P \] (56)
where $P$ is Cholesky factor of $\Sigma$. In this case:

$$C_0 = B_0 \cdot P = P = \begin{bmatrix}
p_{11} & 0 & 0 & 0 \\
p_{21} & p_{22} & 0 & 0 \\
... & ... & ... & 0 \\
p_{n1} & p_{n2} & ... & p_{nn}
\end{bmatrix}$$

so we have a triangular ordering for instantaneous responses: $y_{1t}$ instantaneously depend only on $e_{1t}$, $y_{2t}$ only on $e_{1t}$ and $e_{2t}$, ..., etc...

Given the ordering of the variables, this representation is unique.
Example:

\[
y_t = \begin{bmatrix} \Delta \ln y_t \\ \Delta \ln m_t \end{bmatrix}
\]  \hspace{1cm} (58)

\[
A(L)y_t = \varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \varepsilon_t \sim VWN(0, \Sigma),
\]  \hspace{1cm} (59)

\[
\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \text{, PP'}=\Sigma,
\]  \hspace{1cm} (60)

\[
P = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ \frac{1}{\sqrt{\sigma_{11}}} \sigma_{12} & \sqrt{\sigma_{22} - \frac{1}{\sigma_{11}} \sigma_{12}^2} \end{bmatrix}
\]  \hspace{1cm} (61)

Now, permute order of elements of \(y_t\):

\[
y_t^* = \begin{bmatrix} \Delta \ln m_t \\ \Delta \ln y_t \end{bmatrix},
\]  \hspace{1cm} (62)

with obviously:

\[
\Sigma^* = \begin{bmatrix} \sigma_{22} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix}
\]  \hspace{1cm} (63)
We have:

$$P^*P^* = \Sigma^*$$  \hspace{1cm} (64)

$$P^* = \begin{bmatrix} \sqrt{\sigma_{22}} & 0 \\ \frac{\sigma_{12}}{\sqrt{\sigma_{22}}} & \sqrt{\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}} \end{bmatrix}$$  \hspace{1cm} (65)

So we obtain two conceptually different orthogonal VMA representations:

1)  $$y_t = C(L)e_t \text{ con } e_t = P\varepsilon_t$$  \hspace{1cm} (66)

2)  $$y_t^* = C^*(L)e^*_t \text{ con } e^*_t = P^*\varepsilon^*_t$$  \hspace{1cm} (67)

which stem from two different ways of mapping $\varepsilon_t$ onto orthogonal shocks.

There exist $n!$ possible orderings.
It is possible to see how the Cholesky-based orthogonalised VMA is determined from a different viewpoint.

In fact, write the VAR model as follows:

$$y_t = \sum_{i=1}^{p} A_i y_{t-i} + \mathbf{A}_0^{-1} \mathbf{F} t$$

where $\mathbf{F}$ is diagonal and $\mathbf{A}_0$ is lower triangular with unit diagonal elements. Thus:

$$\mathbf{A}_0 y_t = \sum_{i=1}^{p} \mathbf{A}_i^* y_{t-i} + \mathbf{F} t$$

$$\mathbf{A}_i^* = \mathbf{A}_0 \mathbf{A}_i$$

i.e. a SEM with constraints on the coefficients on endogenous variables, and on the variance-covariance matrix of SF errors which are supposed to be orthogonal. NO CONSTRAINTS ARE IMPOSED ON LAGGED ENDOGENOUS VARIABLES ("incredible restrictions", in Sims’ (1980) wording).
How many constraints?

\[ \frac{n(n-1)}{2} \text{ on } \Xi, \text{ SF error vcv matrix} \]

\[ \frac{n(n+1)}{2} \text{ on } A_0 \]

Hence \( n^2 \) constraints: order conditions are satisfied. In fact: exactly indentified structure. It is impossible to check its validity on the grounds of statistical tests.
5 Structural VAR models

Remember:

\[ C \varepsilon_t = \varepsilon_t \] (71)
\[ e_t \sim VWN(0, I_n) \] (72)
\[ CC' = \Sigma \] (73)

Simplest (but is it sensible?) choice:

\[ C = P \]

⇒ Wold causal chain (triangular or recursive system).

Sometime different structures (Structural VAR or SVAR literature).
Consider a VAR\((p)\) model:

\[
A(L)y_t = \varepsilon_t, \quad (74)
\]

\[
A(L) = I_n - A_1L - A_2L^2 - \ldots - A_pL^p \quad (75)
\]

\[
\varepsilon_t \sim WN(0, \Sigma) \quad (76)
\]

This is clearly a reduced form (RF) \(\Rightarrow\) simultaneity linkages are solved out and relegated in \(\Sigma\).

Spse the structural form (SF) is:

\[
AA(L)y_t = F\varepsilon_t \quad (77)
\]

\[
\varepsilon_t \sim WN(0, I_n) \quad (78)
\]

where \(A\) and \(F\) are \((n \times n)\) invertible matrices.

Hence:

\[
C = A^{-1}F \quad (79)
\]
Remember the SEM:

\[
\Gamma' y_t + \Phi' x_t = v_t, \quad (80)
\]
\[
\Gamma' = A, \quad (81)
\]
\[
\Phi' = \begin{bmatrix} -A_1^* & -A_2^* & \ldots & -A_p^* \end{bmatrix}, \quad (82)
\]
\[
A_i^* = A \cdot A_i \quad (83)
\]
\[
x_t = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix} \quad (84)
\]
\[
v_t = Fe_t, \quad \Omega = FF' \quad (85)
\]

The key relationship is:

\[
A \varepsilon_t = F e_t \quad (n \times n) \quad (n \times n)
\]

Remember \(e_t\) are orthogonal (and unit variance).
Identification is achieved by imposing constraints on SF parameters $\Gamma(A)$ and $F$ (and not on $\Phi \Rightarrow "\text{incredible}"$ restrictions).

Note that in RF we have $nk + n(n + 1)/2$ free parameters ($A_i, i = 1, 2, \ldots p$, $\Sigma$) while in SF we have $nk + 2n^2$ free parameters ($\Phi, A$ and $F$).

$\Rightarrow$ need at least $2n^2 - n(n + 1)/2$ constraints. Key relationship is:

$$A\Sigma A' = FF' \quad (86)$$

Consider linear non-homogeneous constraints on $A$ and $F$:

$$\begin{cases} 
R_A \quad \text{vec}(A) = r_A \\
(q_A \times n^2) (n^2 \times 1) \quad (q_A \times 1) \\
R_F \quad \text{vec}(F) = r_F \\
(q_F \times n^2) (n^2 \times 1) \quad (q_F \times 1)
\end{cases} \quad \text{(implicit form)} \quad (87)$$
or:

\[
\begin{align*}
vec(A) &= \begin{bmatrix} H_A & \varphi_A \\ (n^2 \times 1) & (n^2 \times s_A)(s_A \times 1) & (n^2 \times 1) \end{bmatrix} + h_A, s_A = n^2 - q_A \\
vec(F) &= \begin{bmatrix} H_F & \varphi_F \\ (n^2 \times 1) & (n^2 \times s_F)(s_F \times 1) & (n^2 \times 1) \end{bmatrix} + h_F, s_F = n^2 - q_F
\end{align*}
\]

(88) 

where:

\[
\begin{align*}
R_A H_A &= [0], R_A h_A = r_A \\
R_F H_F &= [0], R_F h_F = r_F
\end{align*}
\]

(89) (90)

Now let us see some examples.
5.1 Examples of S-VARs

5.1.1 Cholesky factorisation

\[
\varepsilon_t = P e_t \\
PP' = \Sigma
\]

(91)  
(92)

To achieve (77), we can define:

- \( A \) = lower triangular with unit diagonal elements;
- \( F \) = diagonal;
Constraints:

\[ a_{ij} = 0 \ \forall j > i, \ (n(n - 1)/2 \text{ constraints}) \]  (93)
\[ a_{ii} = 1 \ \forall i, \ (n \text{ constraints}) \]  (94)
\[ f_{ij} = 0 \ \forall j \neq i \ (n(n - 1) \text{ constraints}) \]  (95)

Total # of constraints:

\[ n + \frac{n(n-1)}{2} + n(n-1) = 2n^2 - \frac{n(n+1)}{2} \]  (96)

⇒ Order conditions suggest exact identification.

Example with \( n = 2 \):

\[ A = \begin{bmatrix} 1 & 0 \\ a_{21} & 1 \end{bmatrix}, \ F = \begin{bmatrix} f_{11} & 0 \\ 0 & f_{22} \end{bmatrix} \]  (97)

constraints can be written as:
\[ R_A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad r_A = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \]  
\[ R_F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad r_F = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \]  

or:

\[ H_A = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad h_A = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \]

\[ H_F = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad h_F = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \]

In practice, (exact identification) use RF to estimate \( \Sigma (\hat{\Sigma}) \), and obtain \( \hat{P} \)
and obtain:

\[
\hat{A} = \begin{bmatrix}
q_{11} & 0 & \ldots & 0 \\
0 & q_{22} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & q_{nn}
\end{bmatrix} \cdot Q \quad (102)
\]

\[
\hat{F} = \begin{bmatrix}
q_{11} & 0 & \ldots & 0 \\
0 & q_{22} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & q_{nn}
\end{bmatrix}
\]

\[
Q = \hat{P}^{-1} \quad (103)
\]

5.1.2 Long-run neutrality of demand shocks

(Blanchard and Quah, 1989)

\[
y_t = \begin{bmatrix}
\Delta y_t \\
n_t
\end{bmatrix} \quad (105)
\]
\( \Delta y_t = \text{GDP growth (} \Delta \ln GDP \text{)} \)

\( n_t \) log employment.

Both series are assumed stationary.

Spse \( e_t \) contains a supply and a demand shock and that demand shocks have no effect on the LEVEL of GDP (cumulated \( \Delta y_t \)).

What are the effects of shocks on \( y_t \)? (I(1) series)

Must look at cumulated IRFs:

\[
\frac{\partial y_{t+k}}{\partial e_{jt}} \equiv \sum_{h=0}^{k} \frac{\partial \Delta y_{t+h}}{\partial e_{jt}}, \quad (106)
\]

\[
\lim_{k \to \infty} \frac{\partial y_{t+k}}{\partial e_{jt}} = \sum_{h=0}^{\infty} \frac{\partial \Delta y_{t+h}}{\partial e_{jt}} \quad (107)
\]
Hence we need to look at:

\[ C(1) = \sum_{i=0}^{\infty} C_i = \begin{bmatrix} c_{11}(1) & c_{12}(1) \\ c_{21}(1) & c_{22}(1) \end{bmatrix} \]  

(108)

and we impose:

\[ C(1) = \begin{bmatrix} c_{11}(1) & 0 \\ c_{21}(1) & c_{22}(1) \end{bmatrix} \]  

(109)

i.e. \( C(1) \) is lower triangular. Most convenient way of achieving this is to assume \( A = I_2 \) and \( F \) such that:

\[ c_{12}(1) = 0 \]  

(110)

To find the constraint(s?) on \( F \), remember that:

\[ C(1) = \sum_{i=0}^{\infty} C_i = \sum_{i=0}^{\infty} B_i C = B(1)F \]  

(111)

since:

\[ C = A^{-1}F = F \]  

(112)

\[ A = I_2 \]  

(113)
so we require that the first row of $B(1)$ and the second column of $F$ be orthogonal:

$$b_{11}(1) \cdot f_{12} + b_{12}(1) \cdot f_{22} = 0 \quad (114)$$

Constraints:

$$R_A = I_4, \quad r_A = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad (115)$$

$$R_F = \begin{bmatrix} 0 & 0 & b_{11}(1) & b_{12}(1) \end{bmatrix}, \quad r_F = [0] \quad (116)$$

Note that:

- $b_{11}(1), b_{12}(1)$ need to be estimated

- By simple contraints accounting we can easily see that we are in an exact identification framework (4 constraints on $A$ and 1 on $F$)
Practically, to obtain a consistent estimate of $F$ we can consider:

$$C(1)C(1)' = B(1)FF'B(1)' = B(1)\Sigma B(1)'$$

$$\Rightarrow F = [B(1)]^{-1}C(1)$$

(117)

(118)

where $C(1)$ is lower triangular $\Rightarrow$ Cholesky factor of $B(1)\Sigma B(1)'$.

So, in order to estimate this SVAR model we can use the following procedure:

1. Estimate RF VAR and obtain:

$$\hat{B}(1) = \sum_{i=0}^{\infty} \hat{B}_i,$$

$$\hat{B}(L) = [\hat{A}(L)]^{-1}$$

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

(119)

(120)

(121)
2. Estimate \( C(1) \) (\( \hat{C}(1) \)) as Cholesky factor of \( \hat{B}(1)\hat{\Sigma}\hat{B}(1)' \);

3. estimate \( F \) (\( \hat{F} \)) as:
\[
\hat{F} = [\hat{B}(1)]^{-1}\hat{C}(1) \tag{122}
\]

5.1.3 AS-AD model

(simplified version of Gali, 1992)

\[
y_t = \begin{bmatrix} y_t \\ i_t \\ m_t \\ p_t \end{bmatrix} \tag{123}
\]
• $y_t = \log \text{income}$;

• $i_t = \text{short run interest rate}$;

• $m_t = \log \text{money}$;

• $p_t = \log \text{price level}$.

Spse we have 4 "structural relationships involving current endogenous variables (ISLM model + AS)

1. IS curve:

$$y_t = \alpha + \gamma e_{st} - c(i_t - \Delta p) + \delta e_{IS_t}$$  \hspace{1cm} (124)

$e_{st} = \text{supply shock, } e_{IS_t} = \text{AD shock}$. 
2. M-demand:
\[ m_t - p_t = \phi y_t - \lambda i_t + \eta e_{md} \]  
\( e_{md} = \) M-dem shock.

3. Money supply:
\[ \Delta m_t = \kappa + \theta e_{mS_t} \]  
\( e_{mS_t} = \) M-supply shock.

4. Phillips curve (AS):
\[ \Delta p_t = \Delta p_{t-1} + \beta (y_t - \xi e_{S_t}) \]  

Primitive exogenous shocks:

\[
e_t = \begin{bmatrix} e_{IS} \\ e_{md} \\ e_{mS} \\ e_s \end{bmatrix}
\]  \tag{128}

\[
e_t \sim VWN(0, I_4)
\]  \tag{129}

where shocks are normalised to unit variance and they are orthogonal.

Take S-VAR :

\[
A \ A(L) \ y_t = F \ e_t
\]

need to specify \( A \) and \( F \) according to structure. Leave dynamics coefficients unrestricted.

\[
A = \begin{bmatrix} 1 & a_{12} & 0 & -a_{12} \\ a_{21} & 1 & a_{23} & -a_{23} \\ 0 & 0 & 1 & 0 \\ a_{41} & 0 & 0 & 1 \end{bmatrix}, \ F = \begin{bmatrix} f_{11} & 0 & 0 & f_{14} \\ 0 & f_{22} & 0 & 0 \\ 0 & 0 & f_{33} & 0 \\ 0 & 0 & 0 & f_{44} \end{bmatrix}
\]  \tag{130}
12 constraints on $A$ and 11 on $F$, while for exact identification we should need:

$$2n^2 - \frac{n(n + 1)}{2} = 22$$

(131)

Overidentification?

6 Bayesian estimation of VARs

Use prior, for instance Use

$$p(\beta, H) = p(\beta|H) \cdot p(H)$$

(132)

with

$$p(H) = W_n(\nu, S)$$

(133)

$$p(\beta|H) = N(m_B, H^{-1} \otimes H_B^{-1})$$

(134)

$$m_B = vec(M_B)$$

(135)
With some algebra it is possible to show that:

\[
p(\beta|H, Y) = N(\overline{m}_B, H^{-1} \otimes (X'X + H_B)^{-1}) \tag{136}
\]

\[
\overline{m}_B = \text{vec}(\overline{M}_B), \tag{137}
\]

\[
\overline{M}_B = (X'X + H_B)^{-1}(X'Y + H_B \overline{M}_B) \tag{138}
\]

and

\[
p(H|\beta, Y) = W_n(\overline{\nu}, \overline{S}), \tag{139}
\]

\[
\overline{\nu} = T + \nu, \tag{140}
\]

\[
\overline{S} = \left[ S^{-1} + (B - \overline{M}_B)' H_B (B - \overline{M}_B) + E'E \right]^{-1} \tag{141}
\]

Then can use a 2-step Gibbs sampling approach
1. Sample $H$ from its conditional posterior distribution (Wishart)

2. Sample $\beta$ from its conditional posterior distribution (Gaussian)

This will generate sample from joint posterior distribution of the parameters

7 Bayesian estimation of exactly identified SVARs

Suppose the mapping

$$CC' = \Sigma$$

is exactly identified. For overidentified cases, see Waggoner and Zha (2004)

We can therefore apply the following Bayesian estimation algorithm
• simulate $H$ and $\beta$ from posterior distribution (using Gibbs sampling)

• for each draw $H$ compute $C$ and use it for evaluating IRFs and FEVDs.

• store them

then we have a sample from posterior distribution of those features (IRFs and FEVDs)

Remember

$$E(f(\theta)|Y) = \int f(\theta)p(\theta|Y)d\theta$$  \hspace{1cm} (142)

$$\approx \frac{1}{M} \sum_{i=1}^{M} f(\theta^{(i)})$$  \hspace{1cm} (143)

$$\theta^{(i)} \sim p(\theta|Y)$$  \hspace{1cm} (144)
8 Example I: a trivariate model for identifying monetary shocks


We have trivariate system

\[ y_t = \begin{bmatrix} u_t \\ \pi_t \\ r_t \end{bmatrix} \] (145)

monthly data for the US, 1949-2005

estimate VAR for the trivariate system
To identify the monetary shock, assume $r_t$ does not influence instantaneously $u_t$ and $\pi_t$:

$$A_0y_t = A_0 \sum_{i=1}^{p} A_i y_{t-i} + F e_t$$

$$A_0 = \begin{bmatrix} 1 & * & 0 \\ * & 1 & 0 \\ * & * & 1 \end{bmatrix}, F = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}$$

In practice, if interested only in identifying monetary policy shock, then impose that $A_0$ be lower triangular.

Then we use

$$C = A_0^{-1} F$$

the Cholesky factor of $\Sigma$.

See Matlab code main_bvar_trivariate.m
9 Example II: long run restrictions on New Keynesian model of labour market


Modified version of the one contained in Balmaseda et al. (2000) (henceforth BDLS).
Model is:

\[ AD \quad : \quad y_t = \phi(d_t - p_t) + a\theta_t \quad (147) \]

\[ AS \quad : \quad y_t = n_t + \theta_t \quad (148) \]

\[ PS \quad : \quad p_t = w_t - \theta_t \quad (149) \]

\[ WS \quad : \quad w_t = \{w_t : n_t^e = \bar{n}_t\} \quad (150) \]

\[ \bar{n}_t = \lambda l_{t-1} + (1 - \lambda) n_{t-1}, \quad (151) \]

\[ PE \quad : \quad l_t = \alpha(w_t - p_t^e) - bu_{t-1} + \tau_t \quad (152) \]

\[ \Delta \theta_t = \varepsilon_t^s \quad (153) \]

\[ \Delta d_t = \varepsilon_t^d \quad (154) \]

\[ \Delta \tau_t = \varepsilon_t^l \quad (155) \]
9.1 Long run coefficients

The system can be written in final form:

\[
y_t = B(L)Ce_t, \quad (156)
\]

\[
y_t = \begin{bmatrix}
(w - p)_t \\
y_t \\
\sum_{\tau=1} u_\tau
\end{bmatrix}, \quad e_t = \begin{bmatrix}
e^s_t \\
e^l_t \\
e^d_t
\end{bmatrix}
\]

\[
B(1)C = \begin{bmatrix}
\sigma_s & 0 & 0 \\
(1 + \alpha) \sigma_s & \sigma_l & 0 \\
-(-1 - \alpha + \phi + a) \frac{\sigma_s}{\lambda} & \frac{1}{\lambda} \sigma_l & -\frac{\phi}{\lambda} \sigma_d
\end{bmatrix} \quad (157)
\]
9.2 Impact coefficients

\[ B_0C = \begin{bmatrix}
\sigma_s & 0 & 0 \\
(\phi + \alpha)\sigma_s & 0 & \phi\sigma_d \\
(1 - \phi - \alpha)\sigma_s & \sigma_l & -\phi\sigma_d
\end{bmatrix} \]  

(158)

9.3 Identification via long run restrictions

Compute \( B(1) \) and impose \( B(1)C \) is triangular, ie use

\[ B(1)C = P \]  

(159)

where \( P \) is Cholesky factor of \( B(1)\Sigma B(1)' \)

Then obtain

\[ C = (B(1))^{-1}P \]  

(160)
See Matlab code main_bvar_AS.m
10 "Identification" via sign restrictions

- Faust (1998)
- Canova and De Nicolò (2002)
- Uhlig (2005)
- Paustian (2007)
- Dedola and Neri (2007)

- Zero impact restrictions are "unbelievable", especially in a general equilibrium framework.
• Use sign restrictions: demand shock has positive effects on prices and quantities, supply shock negative on prices and positive on quantities

• Simple idea: rotate an identified structure to achieve responses which satisfy constraints

• Mathematically

\[ \Sigma = CC' \]

\[ = CQQ'C' \]

\[ = C^*C'^* \]

• search for possible \( Q \) stochastically or deterministically

• Problem: \( Q \) is not unique: not point identification but "interval identification": for each value of \( \Sigma \) there is a continuum of orthonormal \( Q \) that satisfy the constraints.
10.1 Bayesian estimation algorithm

• draw from reduced form parameters $\beta$ and $H$

• compute $C$ as Cholesky factor of $\Sigma$

• find a rotation $Q$ which generates IRFs satisfying the constraints

• repeat $M$ times

10.2 How to find $Q$

Several algorithm. Most efficient in Rubio, Waggoner and Zha (2007):
• Draw each element of $X (n \times n)$ from NID(0,1)

• Compute QR decomposition of $X$, the $Q$ factor is a candidate rotation (from Haar distribution)

• compute $C^* = CQ$, the associated IRFs and see whether constraints are satisfied

• if constraints satisfied, stop, otherwise continue.

10.3 Example: Dedola and Neri (2007)

• What is the effect of a tech shock on hours worked
• Neoclassical model: positive!

• New Keynesian model with frictions: at the outset hours go down

• Take large(ish) information set VAR and impose acceptable constraints to identify tech shock:
  – increases per capita labour productivity
  – increases real wage
  – increases per capita consumption,
  – increases per capita investment
  – increases per capita output
• Constraints can regard impact effects and or delayed effects

• See Matlab file main_bvar_DN.m