A Bayesian VAR with Sign Restrictions

Barcelona GSE. Empirical Time Series Methods for Macroeconomic Analysis.
Instructor: Luca Gambetti.
TA: Nicolò Maffei Faccioli.

1 Introduction:

In this set of notes we review the estimation of Bayesian VARs with sign restrictions and we make an example where we impose sign restrictions to estimate the effects of asset purchases, demand and supply shocks on the macroeconomy.

2 A Bit of Theory:

Consider the following reduced form VAR:

$$Y_t = C + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + u_t \tag{1}$$

where $u_t \sim N(0, \Sigma)$ is the reduced form residual, Y_t is a nx1 vector containing all the n endogenous variables, $A_1, ..., A_p$ are nxn matrices of coefficients and C is a nx1 vector of constants.
Notice that the process represented in (1) can be stacked in a more compact form, abstracting from the subscript t. This will be useful for the OLS estimation of the matrices $A_1, ..., A_p$, for the construction of the likelihood function of the data given the parameters of the model, and the posterior of the parameters given the data. We can rewrite the process in (1) as (SUR representation) follows:

$$\mathbf{Y} = \mathbf{X}B + \mathbf{U} \tag{2}$$

where:
1) $\mathbf{Y} = (Y_{p+1}, ..., Y_T)$ is a $(T-p)$ x n matrix, with $Y_t = (Y_{1t}, ..., Y_{nt})$.
2) $\mathbf{X} = (\mathbf{1}, \mathbf{Y}_{-1}, ..., \mathbf{Y}_{-p})$ is a $(T-p)$ x $(np+1)$ matrix, where $\mathbf{1}$ is a $(T-p)$ x 1 matrix of ones and $\mathbf{Y}_{-k} = (Y_{p+1-k}, ..., Y_{T-k})$ is a $(T-p)$ x n matrix, for $k = 1, ..., p$.
3) $\mathbf{U} = (u_{p+1}, ..., u_T)$ is a $(T-p)$ x n matrix.
4) $B = (C, A_1, ..., A_p)'$ is a $(np+1)$ x n matrix of coefficients.
Vectorizing (2), we obtain:

$$\mathbf{y} = (I_n \otimes \mathbf{X})\beta + \mathbf{u} \tag{3}$$

where $\mathbf{y} = vec(\mathbf{Y})$, $\beta = vec(B)$, $\mathbf{u} = vec(\mathbf{U})$ and $\mathbf{u} \sim N(0, \Sigma \otimes I_{T-p})$.
Given the assumption of normality of errors, we can express the likelihood of the sample, conditional on the parameters of the model and the set of regressors $\mathbf{X}$, as follows:

$$L(\mathbf{y}|\mathbf{X}, \beta, \Sigma) \propto |\Sigma \otimes I_{T-p}|^{-\frac{T-p}{2}} exp\left\{\frac{1}{2}(\mathbf{y} - I_n \otimes \mathbf{X}\beta)'(\Sigma \otimes I_{T-p})^{-1}(\mathbf{y} - I_n \otimes \mathbf{X}\beta)\right\} \tag{4}$$

Denote $\hat{\beta} = vec(\hat{B})$, where $\hat{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the MLE=OLS estimate, and $S = (\mathbf{Y} - \mathbf{X}\hat{B})'(\mathbf{Y} - \mathbf{X}\hat{B})$ is the sum of squared errors. Then we can rewrite (4) as follows:

$$L(\mathbf{y}|\mathbf{X}, \beta, \Sigma) \propto |\Sigma \otimes I_{T-p}|^{-\frac{T-p}{2}} exp\left\{\frac{1}{2}(\beta - \hat{\beta})'(\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X})(\beta - \hat{\beta})\right\} exp\left\{-\frac{1}{2}tr(\Sigma^{-1}S)\right\} \quad (5)$$

By choosing a non-informative (diffuse) prior for B and Σ that is proportional to $|\Sigma|^{-\frac{n+1}{2}}$, namely:

$$p(B|\Sigma) \propto 1$$
$$p(\Sigma) \propto |\Sigma|^{-\frac{n+1}{2}}$$

We can compute the posterior of the parameters given the data at hand using Bayes rule, as follows:

$$\begin{aligned} P(B, \Sigma|\mathbf{y}, \mathbf{X}) &= L(\mathbf{y}|\mathbf{X}, \beta, \Sigma)p(B|\Sigma)p(\Sigma) \\ &= |\Sigma|^{-\frac{T-p+n+1}{2}} exp\left\{\frac{1}{2}(\beta - \hat{\beta})'(\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X})(\beta - \hat{\beta})\right\} exp\left\{-\frac{1}{2}tr(\Sigma^{-1}S)\right\} \end{aligned}$$

Hence:

$$\beta|\Sigma, \mathbf{y}, \mathbf{X} \sim N(\hat{\beta}, \Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1})$$
$$\Sigma|\mathbf{y}, \mathbf{X} \sim IW(S, v)$$

where $v = T - p - (np + 1) - n - 1$.
The marginal posterior distribution of β, $P(\beta|\Sigma, \mathbf{y}, \mathbf{X})$, and the joint distribution of β and Σ,$P(\beta, \Sigma|\mathbf{y}, \mathbf{X})$, and therefore the IRFs of the reduced form VAR, are obtained through the Gibbs Sampler, which conisists of the following steps:

1. Generate $\Sigma^{(0)}$ by drawing from $IW(S, v)$, where S is obtained by estimating (2) by OLS=MLE.
2. Compute $\Sigma^{(0)} \otimes (\mathbf{X}'\mathbf{X})^{-1}$ and draw $\beta^{(0)}$ from $N(\hat{\beta}, \Sigma^{(0)} \otimes (\mathbf{X}'\mathbf{X})^{-1})$, where $\hat{\beta} = vec(\hat{B})$ and $\hat{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.
3. Repeat 1. and 2., say, $M = 3000$ times.
4. Discard, say, $P = 1000$ "initial" draws and pick $H = 1000$ of the draws.

The MATLAB function $BVAR(y, p, c)$ computes the first two steps, given the data y, the number of lags p and the constant term c.

```
%%% BVAR - Diffuse/Uninformative Prior (Uniform -inf, +inf)
%%% Author: Nicolo' Maffei Faccioli

function [PI,BigA,Q,errornorm,fittednorm]=BVAR(y,p,c)

    [Traw,K]=size(y);

    T=Traw-p;

    [pi_hat,Y,X,¬,¬,err]=VAR(y,p,c); % ML = OLS estimate
```

```
    sigma=err'*err; % SSE

    % Draw Q from IW ¬ (S,v), where v = T-(K*p+1)-K-1 :

    Q=iwishrnd(sigma,T-size(X,2)-K-1);

    % Compute the Kronecker product Q x inv(X'X) and vectorize the matrix
    % pi_hat in order to obtain vec(pi_hat):

    XX=kron(Q,inv(X'*X));
    s=size(pi_hat)';
    vec_pi_hat=reshape(pi_hat,s(1)*s(2),1);

    % Draw PI from a multivariate normal distribution with mean vec(pi_hat)
    % and variance Q x inv(X'X):

    PI=mvnrnd1(vec_pi_hat,XX,1);
    PI=PI';
    PI=reshape(PI,[K*p+c,K]); % reshape PI such that Y=X*PI+e, i.e. PI is ...
        (K*p+c)x(K).

    % Create the companion form representation matrix A:

    BigA=[PI(1+c:end,:)'; eye(K*p-K) zeros(K*p-K,K)]; % (K*p)x(K*p) matrix

    % Store errors and fitted values:

    errornorm=Y-X*PI;
    fittednorm=X*PI;

end
```

Steps 3 and 4 are then implemented within the script:

```
%% Estimating a Bayesian VAR(2) with Gibbs Sampler:

y=[log(GDPMA(1:end-24))*100, log(CPI(1:end-24))*100, ...
    APWW20081(1:end)./143.839, GS10(1:end-24), log(SP500(1:end-24))*100];

[T,n]=size(y);
p=2; % # of lags
c=1; % include constant | if c=0, no constant

% Draws set up:

sel=2;
maxdraws=3000;
b_sel=1001:sel:maxdraws; % exclude the first 1000 draws
drawfin=size(b_sel,2); % number of stored draws

% Set up the loop for each draw :

PI=zeros(n*p+c,n,maxdraws+1);
BigA=zeros(n*p,n*p,maxdraws+1);
Q=zeros(n,n,maxdraws+1);
errornorm=zeros(T-p,n,maxdraws+1);
fittednorm=zeros(T-p,n,maxdraws+1);

for i=1:maxdraws+1

```

```
[PI(:,:,i),BigA(:,:,i),Q(:,:,i),errornorm(:,:,i),fittednorm(:,:,i)]=BVAR(y,p,c);

end
```

How do we obtain the reduced form IRFs? First of all, notice that any $VAR(p)$ can be rewritten in terms of a $VAR(1)$ process using what is called the companion form representation. Define $\mathbf{Y_t} = (Y_t, Y_{t-1}, ..., Y_{t-p+1})'$, $\mathbf{u_t} = (u_t, 0, ..., 0)'$, $\mathbf{C} = (c_1, ..., c_n, 0, ..., 0)'$ and construct the following np x np companion matrix:

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & ... & ... & A_p \\ I_n & 0 & ... & ... & 0 \\ 0 & I_n & ... & ... & 0 \\ 0 & 0 & I_n & ... & 0 \\ . & . & . & ... & 0 \\ . & . & . & ... & 0 \\ . & . & . & ... & 0 \\ 0 & 0 & ... & I_n & 0 \end{bmatrix} \tag{6}$$

Then the process (1) can be rewritten as follows

$$\mathbf{Y_t} = \mathbf{C} + \mathbf{A}\mathbf{Y_{t-1}} + \mathbf{u_t} \tag{7}$$

In terms of $VMA(\infty)$ representation:

$$\mathbf{Y_t} = \sum_{j=0}^{\infty} \mathbf{A^j}\mathbf{u_{t-j}} = C(L)\mathbf{u_t} \tag{8}$$

where $C_0 = \mathbf{A^0}, C_1 = \mathbf{A^1}, ..., C_j = \mathbf{A^j}$. Notice that the matrix of coefficients of the Wold representation C_j has the following interpretation:

$$\frac{\delta Y_{t+j}}{\delta u_t'} = C_{j\,(1:n,1:n)}$$

That is, the row i, column k element of $C_{j\,(1:n,1:n)}$ identifies the consequences of a unit increase in the kth variable's innovation at date t for the value of the ith variable at time $t+j$ holding all other innovation at all dates constant. After having computed the Gibbs sampler and stored a set of estimates $(C^{(h)}, A_1^{(h)}, ..., A_p^{(h)})$ for $h = 1, ..., H$, we can compute the IRFs by simply substituting $(A_1, ..., A_p)$ in (6) with $(A_1^{(h)}, ..., A_p^{(h)})$, for $h = 1, ..., H$ and computing $C_{j\,(1:n,1:n)}^{(h)}$. The 68% bands are constructed taking the 16th and 84th percentile of $C_{j\,(1:n,1:n)}^{(h)}$, for $h = 1, ..., H$.
Now, how do we identify the shocks and corresponding IRFs based on sign restrictions? Let $u_t = A\epsilon_t$, where $\epsilon_t \sim N(0, I_n)$ and A is such that $AA' = \Sigma$. In what follows, I will assume that A is a Cholesky decomposition of Σ (in principle any different decomposition such that $AA' = \Sigma$ will do the work). In order to identify all the shocks in the system we need additional $\frac{n(n-1)}{2}$ conditions. The additional (sign) restrictions are imposed using the QR decomposition algorithm proposed by Rubio-Ramirez, Waggoner and Zha (2010):

1. Make a draw from a $MN(0_n, I_n)$ and perform a QR decomposition of the matrix with the diagonal or R normalized to be positive, where $QQ' = I_n$.

2. Compute $IRF_j = C_j AQ'$, for $j = 0, ..., J$. If the set of IRFs satisfy the sign restriction store them. If not, discard them.

3. Repeat 1. and 2. until you stored H impulse responses.

3 The Effects of Asset Purchases in the U.S.:

In what follows, we use sign restrictions to assess the effects of asset purchase announcements on the U.S. economy. To do so, we follow the approach of Weale and Wieladek (2016) and we estimate the VAR with two lags, a flat prior and the following restrictions on impact:

Sign Restrictions			
Variables	*Supply* ↑	*Demand* ↑	*AP* ↑
Real GDP	+	+	+
CPI	-	+	+
AP	NA	NA	+
10-yrs Rate	+	+	-
SP500	+	+	+

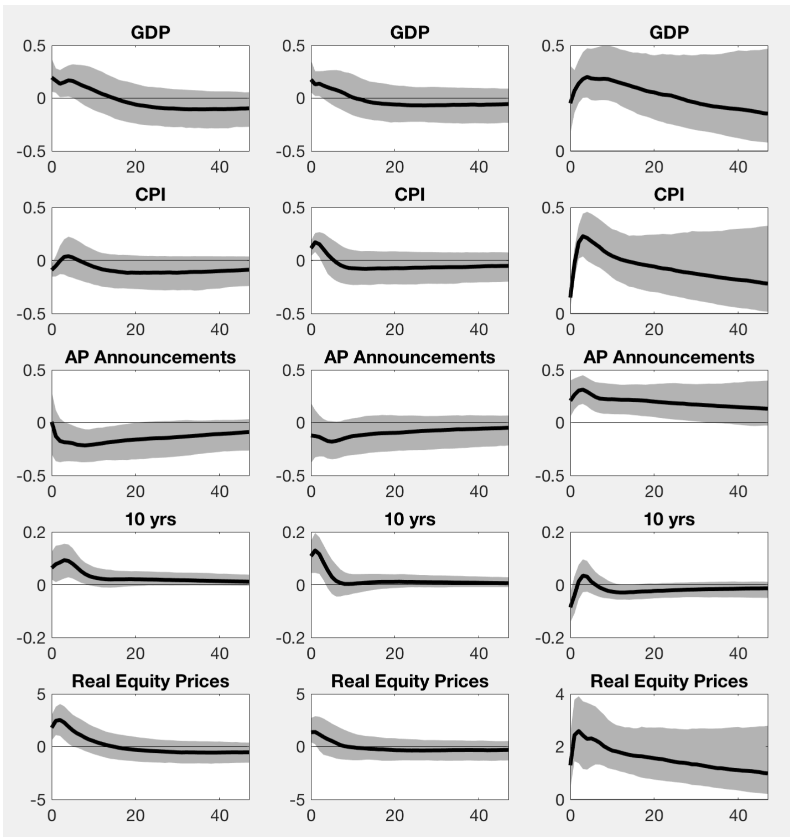

Figure 1: Impact Only - IRFs

Figure 1 presents the responses of, respectively, supply, demand and asset purchase announcements shocks. In line with the results with Weale and Wieladek, a positive AP shock increases significantly GDP, inflation and stock prices. In this sense, it seems that asset purchases have been beneficial in stimulating the economy.