BVAR models "à la Sims" in Dynare

Sébastien Villemot*

July 2007

Dynare incorporates routines for BVAR models estimation, that can be used alone or in parallel with a DSGE estimation. This document describes their implementation and usage.

If you are impatient to try the software and wish to skip mathematical details, jump to section 5.

1 Model setting

Consider the following VAR(p) model:

$$y'_t = y'_{t-1}\beta_1 + y'_{t-2}\beta_2 + \ldots + y'_{t-p}\beta_p + x'_t\alpha + u_t$$

where:

- $t = 1 \ldots T$ is the time index
- y_t is a column vector of ny endogenous variables
- x_t a column vector of nx exogenous variables
- the residuals $u_t \sim \mathcal{N}(0, \Sigma_u)$ are i.i.d. (with Σ a $ny \times ny$ matrix)
- $\beta_1, \beta_2, \ldots, \beta_p$ are $ny \times ny$ matrices
- α is a $nx \times ny$ matrix

Note: in the actual implementation, exogenous variables x_t only include a constant, so that $nx = 1$ and $x'_t = (1, \ldots, 1)$.

The matrix form of the model is:

$$Y = X\Phi + U$$

where:

- Y and U are $T \times ny$
- X is $T \times k$ where $k = ny \cdot p + nx$
- Φ is $k \times ny$

*Paris School of Economics and CEPREMAP. E-mail: `sebastien.villemot@ens.fr`

In other words:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} \; X = \begin{bmatrix} y_0 & \cdots & y_{1-p} & x_1 \\ \vdots & \ddots & \vdots & \vdots \\ y_{T-1} & \cdots & y_{T-p} & x_T \end{bmatrix} \; \Phi = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \\ \alpha \end{bmatrix}$$

2 Constructing the prior

We need a prior distribution over the parameters (Φ, Σ) before moving to Bayesian estimation. This section describes the construction of the prior used in Dynare implementation.

The prior is made of three components, which are described in the following subsections.

2.1 Diffuse prior

The first component of the prior is, by default, Jeffreys' improper prior:

$$p_1(\Phi, \Sigma) \propto |\Sigma|^{-(ny+1)/2}$$

However, it is possible to choose a flat prior by specifying option `bvar_prior_flat`. In, that case:

$$p_1(\Phi, \Sigma) = \text{const}$$

2.2 Dummy observations prior

The second component of the prior is constructed from the likelihood of T^* dummy observations (Y^*, X^*):

$$p_2(\Phi, \Sigma) \propto |\Sigma|^{-T^*/2} \exp\left\{-\frac{1}{2}Tr(\Sigma^{-1}(Y^* - X^*\Phi)'(Y^* - X^*\Phi))\right\}$$

The dummy observations are constructed according to Sims' version of the Minnesota prior[1].

Before constructing the dummy observations, one needs to choose values for the following parameters:

- τ: the overall tightness of the prior. Large values imply a small prior covariance matrix. Controlled by option `bvar_prior_tau`, with a default of 3
- d: the decay factor for scaling down the coefficients of lagged values. Controlled by option `bvar_prior_decay`, with a default of 0.5
- ω controls the tightness for the prior on Σ. Must be an integer. Controlled by option `bvar_prior_omega`, with a default of 1

[1]See Doan, Litterman and Sims (1984).

- λ and μ: additional tuning parameters, respectively controlled by option `bvar_prior_lambda` (with a default of 5) and option `bvar_prior_mu` (with a default of 2)
- based on a short presample Y^0 (in Dynare implementation, this presample consists of the p observations used to initialize the VAR), one also calculates $\sigma = std(Y^0)$ and $\bar{y} = mean(Y^0)$

Below is a description of the different dummy observations. For the sake of simplicity, we should assume that $ny = 2$, $nx = 1$ and $p = 3$. The generalization is straigthforward.

- Dummies for the coefficients on the first lag:

$$\begin{bmatrix} \tau\sigma_1 & 0 \\ 0 & \tau\sigma_2 \end{bmatrix} = \begin{bmatrix} \tau\sigma_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau\sigma_2 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Phi + U$$

- Dummies for the coefficients on the second lag:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \tau\sigma_1 2^d & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tau\sigma_2 2^d & 0 & 0 & 0 \end{bmatrix} \Phi + U$$

- Dummies for the coefficients on the third lag:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & \tau\sigma_1 3^d & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \tau\sigma_2 3^d & 0 \end{bmatrix} \Phi + U$$

- The prior for the covariance matrix is implemented by:

$$\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Phi + U$$

  These observations are replicated ω times.

- Co-persistence prior dummy observation, reflecting the belief that when data on all y's are stable at their initial levels, they will tend to persist at that level:

$$\begin{bmatrix} \lambda\bar{y}_1 & \lambda\bar{y}_2 \end{bmatrix} = \begin{bmatrix} \lambda\bar{y}_1 & \lambda\bar{y}_2 & \lambda\bar{y}_1 & \lambda\bar{y}_2 & \lambda\bar{y}_1 & \lambda\bar{y}_2 & \lambda \end{bmatrix} \Phi + U$$

  Note: in the implementation, if $\lambda < 0$, the exogenous variables will not be included in the dummy. In that case, the dummy observation becomes:

$$\begin{bmatrix} -\lambda\bar{y}_1 & -\lambda\bar{y}_2 \end{bmatrix} = \begin{bmatrix} -\lambda\bar{y}_1 & -\lambda\bar{y}_2 & -\lambda\bar{y}_1 & -\lambda\bar{y}_2 & -\lambda\bar{y}_1 & -\lambda\bar{y}_2 & 0 \end{bmatrix} \Phi + U$$

- Own-persistence prior dummy observations, reflecting the belief that when y_i has been stable at its initial level, it will tend to persist at that level, regardless of the value of other variables:

$$\begin{bmatrix} \mu\bar{y}_1 & 0 \\ 0 & \mu\bar{y}_2 \end{bmatrix} = \begin{bmatrix} \mu\bar{y}_1 & 0 & \mu\bar{y}_1 & 0 & \mu\bar{y}_1 & 0 & 0 \\ 0 & \mu\bar{y}_2 & 0 & \mu\bar{y}_2 & 0 & \mu\bar{y}_2 & 0 \end{bmatrix} \Phi + U$$

This makes a total of $T^* = ny \cdot p + ny \cdot \omega + 1 + ny = ny \cdot (p + \omega + 1) + 1$ dummy observations.

2.3 Training sample prior

The third component of the prior is constructed from the likelihood of T^- observations (Y^-, X^-) extracted from the beginning of the sample:

$$p_3(\Phi, \Sigma) \propto |\Sigma|^{-T^-/2} \exp\left\{-\frac{1}{2}Tr(\Sigma^{-1}(Y^- - X^-\Phi)'(Y^- - X^-\Phi))\right\}$$

In other words, the complete sample is divided in two parts such that $T = T^- + T^+$, $Y = \begin{bmatrix} Y^- \\ Y^+ \end{bmatrix}$ and $X = \begin{bmatrix} X^- \\ X^+ \end{bmatrix}$.

The size of the training sample T^- is controlled by option `bvar_prior_train`. It is null by default.

3 Characterization of the prior and posterior distributions

Notation: in the following, we will use a small "p" as superscript for notations related to the prior, and a capital "P" for notations related to the posterior.

3.1 Prior distribution

We define the following notations:

- $T^p = T^* + T^-$
- $Y^p = \begin{bmatrix} Y^* \\ Y^- \end{bmatrix}$
- $X^p = \begin{bmatrix} X^* \\ X^- \end{bmatrix}$
- $\mathrm{df}^p = T^p - k$ if p_1 is Jeffrey's prior, or $\mathrm{df}^p = T^p - k - ny - 1$ if p_1 is a constant

With these notations, one can see that the prior is:

$$\begin{aligned} p(\Phi, \Sigma) &= p_1(\Phi, \Sigma) \cdot p_2(\Phi, \Sigma) \cdot p_3(\Phi, \Sigma) \\ &\propto |\Sigma|^{-(\mathrm{df}^p + ny + 1 + k)/2} \exp\left\{-\frac{1}{2}Tr(\Sigma^{-1}(Y^p - X^p\Phi)'(Y^p - X^p\Phi))\right\} \end{aligned}$$

We define the following notations:

- $\hat{\Phi^p} = (X^{p\prime}X^p)^{-1}X^{p\prime}Y^p$ the linear regression of X^p on Y^p
- $S^p = (Y^p - X^p\hat{\Phi^p})'(Y^p - X^p\hat{\Phi^p})$

After some manipulations, one obtains:

$$
\begin{aligned}
p(\Phi,\Sigma) &\propto |\Sigma|^{-(\mathrm{df}^p+ny+1+k)/2}\exp\left\{-\frac{1}{2}Tr(\Sigma^{-1}(S^p+(\Phi-\hat{\Phi^p})'X^{p\prime}X^p(\Phi-\hat{\Phi^p})))\right\}\\
&\propto |\Sigma|^{-(\mathrm{df}^p+ny+1)/2}\exp\left\{-\frac{1}{2}Tr(\Sigma^{-1}S^p)\right\}\times\\
&\quad |\Sigma|^{-k/2}\exp\left\{-\frac{1}{2}Tr(\Sigma^{-1}(\Phi-\hat{\Phi^p})'X^{p\prime}X^p(\Phi-\hat{\Phi^p})))\right\}
\end{aligned}
$$

From the above decomposition, one can see that the prior distribution is such that:

- Σ is distributed according to an inverse-Wishart distribution, with df^p degrees of freedom and parameter S^p
- conditionally to Σ, matrix Φ is distributed according to a matrix-normal distribution, with mean $\hat{\Phi^p}$ and variance-covariance parameters Σ and $(X^{p\prime}X^p)^{-1}$

Remark concerning the degrees of freedom of the inverse-Wishart: the inverse-Wishart distribution requires the number of degrees of freedom to be greater or equal than the number of variables, i.e. $\mathrm{df}^p \geq ny$. When the `bvar_prior_flat` option is not specified, we have:

$$\mathrm{df}^p = T^p - k = ny\cdot(p+\omega+1)+1+T^- - ny\cdot p - nx = ny\cdot(\omega+1)+T^-$$

so that the condition is always fulfilled. When `bvar_prior_flat` option is specified, we have:

$$\mathrm{df}^p = ny\cdot w + T^- - 1$$

so that with the defaults ($\omega = 1$ and $T^- = 0$) the condition is not met. The user needs to increase either `bvar_prior_omega` or `bvar_prior_train`.

3.2 Posterior distribution

Using Bayes formula, the posterior density is given by:

$$p(\Phi,\Sigma|Y^+,X^+) = \frac{p(Y^+|\Phi,\Sigma,X^+)\cdot p(\Phi,\Sigma)}{p(Y^+|X^+)} \tag{1}$$

The posterior kernel is:

$$p(\Phi,\Sigma|Y^+,X^+) \propto p(Y^+|\Phi,\Sigma,X^+)\cdot p(\Phi,\Sigma)$$

Since the likelihood is given by:

$$p(Y^+|\Phi,\Sigma,X^+) = (2\pi)^{-\frac{T^+\cdot ny}{2}}|\Sigma|^{\frac{T^+}{2}}\exp\left\{-\frac{1}{2}Tr(\Sigma^{-1}(Y^+-X^+\Phi)'(Y^+-X^+\Phi))\right\}$$

We obtain the following posterior kernel, when combining with the prior:

$$p(\Phi,\Sigma|Y^+,X^+) \propto |\Sigma|^{-(\mathrm{df}^P+ny+1+k)/2}\exp\left\{-\frac{1}{2}Tr(\Sigma^{-1}(Y^P-X^P\Phi)'(Y^P-X^P\Phi))\right\}$$

where:

- $T^P = T^+ + T^p = T^+ + T^- + T^*$
- $Y^P = \begin{bmatrix} Y^p \\ Y^+ \end{bmatrix} = \begin{bmatrix} Y^* \\ Y^- \\ Y^+ \end{bmatrix}$
- $X^P = \begin{bmatrix} X^p \\ X^+ \end{bmatrix} = \begin{bmatrix} X^* \\ X^- \\ X^+ \end{bmatrix}$
- $\mathrm{df}^P = \mathrm{df}^p + T^+$. If p_1 is Jeffrey's prior, then $\mathrm{df}^P = T^P - k$. If p_1 is a constant, $\mathrm{df}^P = T^P - k - ny - 1$.

Using the same manipulations than for the prior, the posterior density can be rewritten as:

$$\begin{aligned} p(\Phi, \Sigma | Y^+, X^+) \quad \propto \quad & |\Sigma|^{-(\mathrm{df}^P + ny + 1)/2} \exp\left\{ -\frac{1}{2} Tr(\Sigma^{-1} S^P) \right\} \times \\ & |\Sigma|^{-k/2} \exp\left\{ -\frac{1}{2} Tr(\Sigma^{-1} (\Phi - \hat{\Phi^P})' X^{P\prime} X^P (\Phi - \hat{\Phi^P}))) \right\} \end{aligned}$$

where:

- $\hat{\Phi^P} = (X^{P\prime} X^P)^{-1} X^{P\prime} Y^P$ the linear regression of X^P on Y^P
- $S^P = (Y^P - X^P \hat{\Phi^P})'(Y^P - X^P \hat{\Phi^P})$

From the above decomposition, one can see that the posterior distribution is such that:

- Σ is distributed according to an inverse-Wishart distribution, with df^P degrees of freedom and parameter S^P
- conditionally to Σ, matrix Φ is distributed according to a matrix-normal distribution, with mean $\hat{\Phi^P}$ and variance-covariance parameters Σ and $(X^{P\prime} X^P)^{-1}$

Remark concerning the degrees of freedom of the inverse-Wishart: in theory, the condition over the degrees of freedom of the inverse-Wishart may not be satisfied. In practice, it is not a problem, since T^+ is great.

4 Marginal density

By integrating equation (1) over (Φ, Σ), one gets:

$$p(Y^+ | X^+) = \int p(Y^+ | \Phi, \Sigma, X^+) \cdot p(\Phi, \Sigma) d\Phi d\Sigma$$

We define the following notation for the unnormalized density of a matrix-normal-inverse-Wishart:

$$\begin{aligned} f(\Phi, \Sigma | \mathrm{df}, S, \hat{\Phi}, \Omega) \quad = \quad & |\Sigma|^{-(\mathrm{df} + ny + 1)/2} \exp\left\{ -\frac{1}{2} Tr(\Sigma^{-1} S) \right\} \times \\ & |\Sigma|^{-k/2} \exp\left\{ -\frac{1}{2} Tr(\Sigma^{-1} (\Phi - \hat{\Phi})' \Omega^{-1} (\Phi - \hat{\Phi}))) \right\} \end{aligned}$$

We also note:

$$F(\mathrm{df}, S, \hat{\Phi}, \Omega) = \int f(\Phi, \Sigma | \mathrm{df}, S, \hat{\Phi}, \Omega) d\Phi d\Sigma$$

The function F has an analytical form, which is given by the normalization constants of matrix-normal and inverse-Wishart densities[2]:

$$F(\mathrm{df}, S, \hat{\Phi}, \Omega) = (2\pi)^{\frac{ny \cdot k}{2}} |\Omega|^{\frac{ny}{2}} \cdot 2^{\frac{ny \cdot \mathrm{df}}{2}} \pi^{\frac{ny(ny-1)}{4}} |S|^{-\frac{\mathrm{df}}{2}} \prod_{i=1}^{ny} \Gamma\left(\frac{\mathrm{df} + 1 - i}{2}\right)$$

The prior density is:

$$p(\Phi, \Sigma) = c^p \cdot f(\Phi, \Sigma | \mathrm{df}^p, S^p, \hat{\Phi^p}, (X^{p\prime} X^p)^{-1})$$

where the normalization constant is $c^p = F(\mathrm{df}^p, S^p, \hat{\Phi^p}, (X^{p\prime} X^p)^{-1})$.

Combining with the likelihood, one can see that the density is:

$$\begin{aligned} p(Y^+ | X^+) &= \frac{\int (2\pi)^{-\frac{T^+ \cdot ny}{2}} f(\Phi, \Sigma | \mathrm{df}^P, S^P, \hat{\Phi^P}, (X^{P\prime} X^P)^{-1}) d\Phi d\Sigma}{F(\mathrm{df}^p, S^p, \hat{\Phi^p}, (X^{p\prime} X^p)^{-1})} \\ &= \frac{(2\pi)^{-\frac{T^+ \cdot ny}{2}} F(\mathrm{df}^P, S^P, \hat{\Phi^P}, (X^{P\prime} X^P)^{-1})}{F(\mathrm{df}^p, S^p, \hat{\Phi^p}, (X^{p\prime} X^p)^{-1})} \end{aligned}$$

5 Dynare commands

Dynare incorporates two commands related to BVAR models à la Sims:

- `bvar_density` for computing marginal density,
- `bvar_forecast` for forecasting (and RMSE computation).

5.1 Common options

The two commands share a set of common options, which can be divided in two groups. They are described in the following subsections.

An important remark concerning options: in Dynare, all options are global. This means that, if you have set an option in a given command, Dynare will remember this setting for subsequent commands (unless you change it again). For example, if you call `bvar_density` with option `bvar_prior_tau = 2`, then all subsequent `bvar_density` and `bvar_forecast` commands will assume a value of 2 for `bvar_prior_tau`, unless you redeclare it. This remark also applies to `datafile` and similar options, which means that you can run a BVAR estimation after a Dynare estimation without having to respecify the datafile.

[2]Function `matricint` of file `bvar_density.m` implements the calculation of the log of F.

5.1.1 Options related to the prior specification

They are:

- `bvar_prior_tau` (default: 3)
- `bvar_prior_decay` (default: 0.5)
- `bvar_prior_lambda` (default: 5)
- `bvar_prior_mu` (default: 2)
- `bvar_prior_omega` (default: 1)
- `bvar_prior_flat` (not enabled by default)
- `bvar_prior_train` (default: 0)

Please refer to section 2 for the discussion of their meaning.

Remark: when option `bvar_prior_flat` is specified, the condition over the degrees of freedom of the inverse-Wishart distribution is not necessarily verified (see section 3.1). One needs to increase either `bvar_prior_omega` or `bvar_prior_train` in that case.

5.1.2 Options related to the estimated dataset

The options related to the estimated dataset are the same than for the `estimation` command (please refer to the Dynare reference manual for more details):

- `datafile`
- `first_obs`
- `presample`
- `nobs`
- `prefilter` (not yet implemented)
- `xls_sheet`
- `xls_range`

The (endogenous) variables of the BVAR model must be declared through a `varobs` statement (see Dynare reference manual).

Restrictions related to the initialization of lags: in DSGE estimation routines, the likelihood (and therefore the marginal density) are evaluated starting from the observation numbered `first_obs + presample` in the datafile[3]. The BVAR estimation routines use the same convention (i.e. the first observation of Y^+ will be `first_obs + presample`). Since we need p observations to initialize the lags, and since we may also use a training sample, the user must ensure that the following condition holds (estimation will fail otherwise):

$$\texttt{first\_obs} + \texttt{presample} > \texttt{bvar\_prior\_train} + \text{number\_of\_lags}$$

[3] `first_obs` points to the first observation to be used in the datafile (defaults to 1), and `presample` indicates how many observations after `first_obs` will be used to initialize the Kalman filter (defaults to 0).

5.2 Marginal density

The syntax for computing the marginal density is:

`bvar_density(`*options_list*`)` *max_number_of_lags*`;`

The options are those described above.

The command will actually compute the marginal density for several models: first for the model with one lag, then with two lags, and so on up to *max_number_of_lags* lags.

5.3 Forecasting

The syntax for computing forecasts is:

`bvar_density(`*options_list*`)` *max_number_of_lags*`;`

The options are those describe above, plus a few ones:

- `forecast`: the number of periods over which to compute forecasts after the end of the sample (no default)
- `bvar_replic`: the number of replications for Monte-Carlo simulations (default: 2000)
- `conf_sig`: confidence interval for graphs (default: 0.9)

The `forecast` option is mandatory.

The command will draw `bvar_replic` random samples from the posterior distribution. For each draw, it will simulate one path without shocks, and one path with shocks.

It will produce one graph per observed variable. Each graph displays:

- a blue line for the mean forecast (equal to the mean of the simulated paths by linearity),
- two green lines giving the confidence interval for the forecasts without shocks,
- two red lines giving the confidence interval for the forecasts with shocks.

Morever, if option `nobs` is specified, the command will also compute root mean squared error (RMSE) for all variables between end of sample and end of datafile.

6 Examples

This section presents two short examples of BVAR estimations. These examples and the associated datafile (`test.xls`) can be found in the `tests/bvar_a_la_sims` directory of the Dynare v4 subversion tree.

6.1 Standalone BVAR estimation

Here is a simple `mod` file example for a standalone BVAR estimation:

```
var dx dy;
varobs dx dy;

bvar_density(datafile = test, first_obs = 20, bvar_prior_flat,
             bvar_prior_train = 10) 8;

bvar_forecast(forecast = 10, bvar_replic = 10000, nobs = 200) 8;
```

Note that you must declare twice the variables used in the estimation: first with a `var` statement, then with a `varobs` statement. This is necessary to have a syntactically correct `mod` file.

The first component of the prior is flat. The prior also incorporates a training sample. Note that the `bvar_prior_*` options also apply to the second command since all options are global.

The `bvar_density` command will compute marginal density for all models from 1 up to 8 lags.

The `bvar_forecast` command will compute forecasts for a BVAR model with 8 lags, for 10 periods in the future, and with 10000 replications. Since `nobs` is specified and is such that `first_obs + nobs - 1` is strictly less than the number of observations in the datafile, the command will also compute RMSE.

6.2 In parallel with a DSGE estimation

Here follows an example `mod` file, which performs both a DSGE and a BVAR estimation:

```
var dx dy;
varexo e_x e_y;
parameters rho_x rho_y;

rho_x = 0.5;
rho_y = -0.3;

model;
dx = rho_x*dx(-1)+e_x;
dy = rho_y*dy(-1)+e_y;
end;

estimated_params;
rho_x,NORMAL_PDF,0.5,0.1;
rho_y,NORMAL_PDF,-0.3,0.1;
stderr e_x,INV_GAMMA_PDF,0.01,inf;
stderr e_y,INV_GAMMA_PDF,0.01,inf;
end;

varobs dx dy;
```

```
check;

estimation(datafile = test, mh_replic = 1200, mh_jscale = 1.3,
           first_obs = 9);

bvar_density 8;

bvar_forecast(forecast = 10, bvar_replic = 2000, nobs = 200) 8;
```

Note that the BVAR commands use the defaults for the prior, and take their `datafile` and `first_obs` options from the `estimation` command.

References

Doan, Thomas, Robert Litterman, and Christopher Sims (1984), "*Forecasting and Conditional Projections Using Realistic Prior Distributions*", Econometric Reviews, **3**, 1-100

Schorfheide, Frank (2004), "*Notes on Model Evaluation*", Department of Economics, University of Pennsylvania

Sims, Christopher (2003), "*Matlab Procedures to Compute Marginal Data Densities for VARs with Minnesota and Training Sample Priors*", Department of Economics, Princeton University

Acknowledgements

Many thanks to Christopher Sims for his BVAR Matlab routines, and to Stéphane Adjemian and Michel Juillard for their helpful support.

License