Lecture 8 - Forecasting with Bayesian VAR Models

Pedro Valls[1]

[1]Sao Paulo School of Economics - FGV
CEQEF-FGV

July 26, 2024


FGV EESP
CENTRO DE ESTUDOS
QUANTITATIVOS EM
ECONOMIA E FINANÇAS

Table of Contents I

Table of Contents II

Outline

- Introduction
- A primer on Bayesian econometrics
- Baseline BVAR specification
- Forecasting with BVAR
- Example with simulated data
- Empirical examples

Table of Contents I

Table of Contents II

Introduction I

- Bayesian Vector Autoregressions (BVARs) have a long history in forecasting since the seminal studies of [Doan et al., 1984] and [Litterman, 1986].
- In recent years, the BVAR appear to be used more systematically for policy analysis and forecasting macroeconomic variables (e.g. [Kadiyala and Karlsson, 1997] and [Koop, 2013]).
- The model specification, besides the choice of the prior distribution for the parameters, we need to address issues such as:
  - (i) the choice of the tightness and of the lag length of the BVAR;
  - (ii) the treatment of the error variance and the imposition of cross-variable shrinkage;
  - (iii) whether to transform the variables to get stationarity, and whether to complement this choice with the imposition of priors favoring cointegration and unit roots.

Introduction II

- The model estimation and forecast construction, under some approaches, estimating and forecasting with N-BVAR can be technically and computationally demanding.
- To avoid costly simulation, Litterman´s specification of the Minesota prior treats the error variance matrix as fixed and diagonal.
- [Litterman, 1986] imposes such a strong assumption to allow for equation-by-equation ridge estimation of the system; treating the error variance as random would have required MCMC simulations of the entire system of equations.
- The theory will focus in approaches that make the computation of point and density forecasts from BVARs quick and easy, making specific choices in the priors and using direct rather than iterated forecasts (e.g., [Marcellino et al., 2006]

Table of Contents I

Table of Contents II

A primer on Bayesian econometrics

- Let $\{y_t\} \overset{iid}{\sim} N(\mu, \sigma^2)$ be a sequence of random variables, sampled for $t = 1, \ldots, T$.
- Define $\theta = [\mu, \sigma^2]$ the vector of unknown parameters, and we use the sample realisations $y_1, \ldots, y_T$ to estimate θ.
- In the classical perspective the vector θ is a set of unknown numbers. A point estimator for it is a vector of random variables that should be informative about θ, and have optimality properties (e.g. unbiasedness, consistency, efficiency).
- For example, the sample mean and variance together form as estimator denoted by $\hat{\theta}$.
- In a Bayesian perspective, the parameter θ are viewed as random variables.

Parameters as random variables I

- We have some a priori (initial) beliefs about them, summarised by prior probability distribution.
- The sample $y_1, \ldots, y_T$ provides information about θ, summarised into the likelihood function associated with the econometric model that we specified, for example y_t is a set of i.i.d. random variables distributed as a $N(\mu, \sigma^2)$, so that the likelihood function would be a multivariate normal density, considered as a function of the parameters μ and σ^2.
- The initial beliefs about θ, the prior distribution , should be combined with the information coming from the data and the chosen econometric model, the likelihood function , to obtain a posterior distribution for the parameters, which summarized how our initial beliefs about θ have been updated given the information in the data $y_1, \ldots, y_T$.

From prior to posterior distribution I

- Let the prior distribution of the random parameters θ be denoted as $p(\theta)$, while the likelihood function for a given value of θ is $\ell(\theta \mid y_1, \ldots, y_T)$.
- The joint distribution of the unknown parameters and data is given by

$$f(y_1, \ldots, y_T, \theta) = \ell(\theta \mid y_1, \ldots, y_T) \cdot p(\theta). \tag{1}$$

- The additional object of interest is the marginal likelihood, which can be obtained by integrating θ out of the joint distribution given by (1):

$$\begin{aligned} f(y_1, \ldots, y_T) &= \int_{\theta \in \Theta} f(y_1, \ldots, y_T, \theta) d\theta \\ &= \int_{\theta \in \Theta} \ell(\theta \mid y_1, \ldots, y_T) \cdot p(\theta) d\theta \end{aligned} \tag{2}$$

From prior to posterior distribution II

- The posterior distribution is obtained applying the Bayes rule:

$$\begin{aligned} p(\theta \mid y_1, \ldots, y_T) &= \frac{f(y_1, \ldots, y_T, \theta)}{f(y_1, \ldots, y_T)} \\ &= \frac{\ell(\theta \mid y_1, \ldots, y_T) \cdot p(\theta)}{f(y_1, \ldots, y_T)} \end{aligned} \tag{3}$$

- As the marginal distribution in the denominator is independent of θ, the posterior distribution is proportional to the product of the prior, $p(\theta)$ and of the likelihood $\ell(\theta \mid y_1, \ldots, y_T)$:

$$p(\theta \mid y_1, \ldots, y_T) \propto \ell(\theta \mid y_1, \ldots, y_T) \cdot p(\theta)$$

- (3) can always be applied. Some prior distributions allow for an analytical derivation of the posterior. When the latter are of the same type of the former (for example, normal) we are in the so-called conjugate case .

An example: The posterior of the mean when the variance is known I

- Let $\{y_t\} \overset{iid}{\sim} N(\mu, \sigma^2)$, with $t = 1, \ldots, T$, and assume, for simplicity, that the variance. σ^2, is known while the mean μ is unkown.
- The prior distribution for μ is specify as $\mu \sim N(m, \frac{\sigma^2}{\nu})$, and can be wirtten as:

$$f(\mu) = \frac{1}{\sqrt{\frac{2\pi\sigma^2}{\nu}}} \exp\left\{ -\frac{(\mu - m)^2}{\frac{2\sigma^2}{\nu}} \right\}$$

- The key parameters in the prior distribution are μ and ν.
- The larger ν is, the more our prior beliefs about μ are concentrated around the mean m.

An example: The posterior of the mean when the variance is known II

- The likelihood is given by:

$$\ell(\theta \mid y_1, \ldots, y_T) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \mu)^2\right\}$$

- The posterior distribution $f(\mu \mid y_1, \ldots, y_T, \sigma^2)$ is given by:

$$f(\mu \mid y_1, \ldots, y_T, \sigma^2) = \frac{1}{[2\pi\sigma^2/(\nu + T)]^{1/2}} \exp\left\{-\frac{(\mu - m^*)^2}{2\sigma^2/(\nu + T)}\right\}$$

where

$$m^* = \frac{\nu}{\nu + T}m + \frac{T}{\nu + T}\overline{y}, \quad \overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

- The posterior distribution is also normal, $N(m^*, \sigma^2/(\nu + T))$.

An example: The posterior of the mean when the variance is known III

- The posterior mean, m^*, is a linear combination of the prior mean, m, and of the sample mean, $\overline{y}$.
- The weights depend on the "tightness" parameter, ν, and on the sample size, T.
- For fixed T, the larger ν the closer is m^* to m.
- When ν is large, i.e., the prior distribution is highly concentrated around m, then the prior plays a relatively larger role than the likelihood in the determination of the posterior distribution.
- When ν is close to zero ("diffuse" prior), we are very uncertain about μ, and we put more weight in the sample information.
- When T grows, the weight on $\overline{y}$ increases and that on m dectreases, for fixed ν.
- Both ν and T have the same effect on the variance of the posterior distribution.

Bayesian linear regression model I

- Consider the linear regression whichg can be written as:

$$y_t = X_{1t}\beta_1 + X_{2t}\beta_2 + \cdots + X_{kt}\beta_k + \epsilon_t \tag{4}$$

  and ϵ_t is $\overset{iid}{\sim} N(0, \sigma^2)$ by construction.

- Therefore y_t is $\overset{iid}{\sim} N(X_t\beta, \sigma^2)$.
- The likelihood function can then be written as:

$$\ell(\theta \mid y_1, \ldots, y_T, X) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left\{-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}\right\} \tag{5}$$

Known variance I

- We assume that the prior distribution of the vector β is multivariate normal, $N(m, \sigma^2 M)$, i.e.:

$$f(\beta \mid \sigma^2) = \frac{1}{(2\pi\sigma^2)^{k/2}} |M|^{-1/2} \exp\left\{ -\frac{(\beta - m)' M^{-1} (\beta - m)}{2\sigma^2} \right\} \tag{6}$$

- The posterior is given by:

$$f(\beta \mid y_1, \ldots, y_T, X, \sigma^2) = \frac{\ell(\theta \mid y_1, \ldots, y_T, X) \cdot f(\beta \mid \sigma^2)}{f(y_1, \ldots, y_T \mid X, \sigma^2)} \tag{7}$$

- The likelihood and the prior in the numerator are those in, respectively, (5) and (6).
- The marginal in the denominator can be obtained by integrating β out of $f(y_1, \ldots, y_T, \beta \mid X, \sigma^2)$

Known variance II

- Combining we get the posterior distribution:

$$f(\beta \mid y_1, \ldots, y_T, X, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{k/2}} |M^{-1} + X'X|^{1/2} \cdot \qquad ($$
$$\cdot \exp\left\{ - \frac{(\beta - m^*)'(M^{-1} + X'X)(\beta - m^*)}{2\sigma^2} \right.$$

- The posterior distribution is $N(m^*, \sigma^2 M^*)$, with

$$\begin{aligned} m^* &= (M^{-1} + X'X)^{-1}(M^{-1}m + X'y) \\ M^* &= (M^{-1} + X'X)^{-1} \end{aligned}$$

- The prior distribution impose a set of stochastic constraints on the parameter θ(see, e.g. [Theil and Goldberger, 1961]).

Known variance III

- The prior implies that

$$\beta = m + \epsilon_\beta, \quad \epsilon_\beta \sim N(0, \sigma^2 M),$$

which we can also write as

$$m = \beta + \epsilon_\beta, \quad \epsilon_\beta \sim N(0, \sigma^2 M) \tag{9}$$

since the sign of the error term does not matter.

- We can group these constraints with the linear regression model as:

Known variance IV

$$\begin{cases} m = \beta + \epsilon_\beta, & E(\epsilon_\beta \epsilon_\beta') = \sigma^2 M \\ y = X\beta + \epsilon, & E(\epsilon\epsilon') = \sigma^2 I_T \end{cases}$$

or

$$y^* = X^*\beta + \epsilon^* \tag{10}$$

where

$$\begin{aligned} y^* &= \begin{bmatrix} m \\ y \end{bmatrix}, \quad X^* = \begin{bmatrix} I_k \\ X \end{bmatrix} \\ E(\epsilon^*\epsilon^{*\prime}) &= \sigma^2 \begin{bmatrix} M & 0 \\ 0 & I_T \end{bmatrix} \end{aligned}$$

Known variance V

- (10) is a linear regression with heteroscedastic errors, so we need to estimate β by GLS, namely:

$$\begin{aligned}\widehat{\beta}^{GLS} &= \left[X^{*\prime} \begin{bmatrix} \sigma^2 M & 0 \\ 0 & \sigma^2 I_T \end{bmatrix} X^* \right]^{-1} \left[X^{*\prime} \begin{bmatrix} \sigma^2 M & 0 \\ 0 & \sigma^2 I_T \end{bmatrix}^{-1} y^* \right] \\ &= \left(M^{-1} + X'X\right)^{-1} \left(M^{-1} m + X'y\right) = m^*\end{aligned}$$

Moroever

$$\begin{aligned}Var\left[\widehat{\beta}^{GLS}\right] &= \left[X^{*\prime} \begin{bmatrix} \sigma^2 M & 0 \\ 0 & \sigma^2 I_T \end{bmatrix} X^* \right]^{-1} \\ &= \sigma^2 (M^{-1} + X'X)^{-1} = \sigma^2 M^*\end{aligned}$$

- The Bayesian posterior distribution for the regression parameters β is equivalent to the classical GLS estimator of the regression parameters in the model where the prior is interpreted as a set of stochastic constraints .

Unkown variance I

- When the error variance is unkwown, the posterior distribution becomes:

$$f(\beta, \sigma^2 \mid y_1, \ldots, y_T, X) = \frac{\ell(\theta \mid y_1, \ldots, y_T, X) \cdot f(\beta \mid \sigma^2) \cdot f(\sigma^2)}{f(y_1, \ldots, y_T \mid X)} \tag{11}$$

- In (11) the likelihood remains as in (5).
- The prior distribution for β and σ^2 is written as the product of the conditional distribution, $f(\beta \mid \sigma^2)$ and the marginal distribution of $f(\sigma^2)$.
- The conditional distribution, $f(\beta \mid \sigma^2)$ is the same as that in (6).
- We need to specify the marginal distribution of $f(\sigma^2)$.
- We choose an inverse-gamma distribution . $\sigma^2 \sim IG(\lambda, \nu)$, which guarantee positivity of the variance.

Unkown variance II

- Equivalently, one could use a gamma distribution for σ^{-2}, $\sigma^{-2} \sim G(\lambda, \nu)$,
- λ may be interpreted as the prior degrees of freedom. The hyper-parameter ν may be interpreted as a prior variance or initial prior estimate for σ^2.
- In particular, it is

$$G(\lambda, \nu) \propto (\sigma^{-2})^{\frac{\nu-2}{2}} exp(-\frac{\sigma^{-2}\nu}{2\lambda})$$

  and $E(\sigma^{-2}) = \lambda$, $var(\sigma^{-2}) = \frac{2\lambda^2}{\nu}$.
- The posterior distribution for $f(\beta \mid \sigma^2, y, X)$ remains the same as (8).
- The posterior for σ^2, $f(\sigma^2 \mid y, X)$, is instead $IG(\lambda^*, \nu^*)$, see for example [Koop, 2003].

Unkown variance III

- [Koop, 2003] also show that the marginal posterior distribution for β, i.e., without conditioning on σ^2, $f(\beta \mid y, X)$ is a multivariate Student t distribution.

Table of Contents I

Table of Contents II

Baseline Bayesian VAR case

- In this section we will present the Bayesian VAR case.
- First the specification of the model.
- Next the Prior parameterizations.

Baseline BVAR specification I

- The baseline specification is a BVAR with normal-inverted Wishart (NIW) conjugate prior, which is a multivariate generalization of the inverse-gamma distribution discussed in the Bayesian linear regression model.
- Given the vector $N \times 1$, $\mathbf{y_t} = (y_{1t}, y_{2t}, \cdots, y_{Nt})$, we consider the following VAR:

$$\mathbf{y_t} = \Phi_c + \Phi_1 \mathbf{y_{t-1}} + \Phi_2 \mathbf{y_{t-2}} + \cdots + \Phi_p \mathbf{y_{t-p}} + \boldsymbol{\epsilon_t}; \quad \boldsymbol{\epsilon_t} \overset{iid}{\sim} N(0, \Sigma) \tag{12}$$

  where $t = 1, \cdots, T$.
- Each equation has $M = Np + 1$ regressors.
- Grouping the coefficients matrices in the $N \times M$ matrix $\mathbf{\Phi} = [\Phi_c, \Phi_1, \Phi_2, \cdots, \Phi_p]$, and defining, $\mathbf{x}_t = (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$, the VAR can be written as:

$$\mathbf{y_t} = \mathbf{\Phi}' \mathbf{x}_t + \boldsymbol{\epsilon_t} \tag{13}$$

Baseline BVAR specification II

- Or

$$\mathbf{Y} = \mathbf{X}\mathbf{\Phi}' + \boldsymbol{E} \tag{14}$$

where $\mathbf{Y} = [\mathbf{y_1}, \cdots, \mathbf{y_T}]'$, $\mathbf{X} = [\mathbf{x_1}, \cdots, \mathbf{x_T}]'$, and $\mathbf{E} = [\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon_T}]'$ are, respectively $T \times N$, $T \times M$ and $T \times T$ matrices.

- Apply the *vec* operator to both sides of (14) we have:

$$vec(\mathbf{Y}) = (\mathbf{I} \otimes \mathbf{X})vec(\mathbf{\Phi}) + vec(\mathbf{E})$$

which is a linear regression.

Baseline BVAR specification III

- For multi-step forecasts it is better to use the companion form, that is

$$\mathbf{x}_{t+1} = \mathbf{\Phi}^{+}\mathbf{x}_t + \widetilde{\boldsymbol{\epsilon}}_t \tag{15}$$

where $\widetilde{\boldsymbol{\epsilon}}_t$ is a $M \times 1$ vector containing $\boldsymbol{\epsilon}_t$ and $0's$ elsewhere and $\mathbf{\Phi}^{+}$ is a $M \times M$ matrix defined as:

$$\mathbf{\Phi}^{+} = \begin{bmatrix} 1 & \mathbf{0}_{1\times N} & \mathbf{0}_{1\times N} & \cdots & \mathbf{0}_{1\times N} \\ \mathbf{\Phi}_c & \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \cdots & \mathbf{\Phi}_p \\ \mathbf{0}_{N\times 1} & \mathbf{I}_N & \mathbf{0}_{N\times N} & \cdots & \mathbf{0}_{N\times N} \\ \vdots & \vdots & \ddots & & \vdots \\ \mathbf{0}_{N\times 1} & \mathbf{0}_{N\times N} & \cdots & \mathbf{I}_N & \mathbf{0}_{N\times N} \end{bmatrix} \tag{16}$$

- In this notation $\mathbf{y}_t$ corresponds to rows $2, \cdots, N+1$ of $\mathbf{x}_{t+1}$, so we can write $\mathbf{y}_t = \mathbf{s}\mathbf{x}_{t+1}$, where $\mathbf{s}$ is a selection matrix.]

Baseline BVAR specification IV

- With this representation the multi-steps forecasts can be obtained as

$$\widehat{\mathbf{x}}_{t+h} = \left(\mathbf{\Phi}^{+}\right)^{h}\mathbf{x}_{t}$$

Prior parametrization I

- The prior density for the VAR parameters, we use the conjugate N-IW prior:

$$vec(\mathbf{\Phi}) \mid \Sigma \sim N(\mathbf{\Phi}_0, \Sigma \otimes \Omega_0), \quad \Sigma \sim IW(S_0, \nu_0) \tag{17}$$

- The conditional posterior distribution of this model is also N-IW:

$$vec(\mathbf{\Phi} \mid \Sigma, \mathbf{Y}) \sim \mathbf{N}(\overline{\mathbf{\Phi}}, \Sigma \otimes \overline{\Omega}), \quad \Sigma \mid \mathbf{Y} \sim IW(\overline{S}, \overline{\nu}) \tag{18}$$

- Let $\hat{\mathbf{\Phi}}$ and $\hat{\mathbf{E}}$ be the OLS estimates of $\mathbf{\Phi}$ and $\mathbf{E}$, respectively.
- Therefore we have similar formulas, the multivariate counterpart of those derived before, that is:

$$\begin{aligned}
\overline{\mathbf{\Phi}} &= (\Omega_0^{-1} + X'X)^{-1}(\Omega_0^{-1}\mathbf{\Phi}_0 + X'Y), \\
\overline{\Omega} &= (\Omega_0^{-1} + X'X)^{-1}, \\
\overline{\nu} &= \nu_0 + T, \\
\overline{S} &= \mathbf{\Phi}_0 + \hat{\mathbf{E}}'\hat{\mathbf{E}} + \hat{\mathbf{\Phi}}'X'X\hat{\mathbf{\Phi}} + \mathbf{\Phi}_0'\Omega_0^{-1}\mathbf{\Phi}_0 - \overline{\mathbf{\Phi}}'\overline{\Omega}^{-1}\overline{\mathbf{\Phi}}.
\end{aligned}$$

Prior parametrization II

- In the case of the natural conjugate N-IW prior, the marginal posterior distribution of $\mathbf{\Phi}$ is matrix-variate-t with expected value $\overline{\mathbf{\Phi}}$.
- We assume that the prior expectation and standard deviation of the coefficient matrices to be:

$$E\left[\mathbf{\Phi}_k^{(i,j)}\right] = \begin{cases} \mathbf{\Phi}^* & \text{if } i = j, k = 1 \\ 0 & \text{otherwise} \end{cases},$$

$$\text{st.dev.}\left[\mathbf{\Phi}_k^{(i,j)}\right] = \begin{cases} \frac{\lambda_1 \lambda_2}{k} \frac{\sigma_i}{\sigma_j} & k = 1, \ldots p \\ \lambda_0 \sigma_i & \text{k} = 0 \end{cases} \tag{19}$$

  where $\mathbf{\Phi}_k^{(i,j)}$ denotes the element in position (i, j) in the matrix $\mathbf{\Phi}_k$
- The prior mean $\mathbf{\Phi}^*$ is set to 1 in the VAR in levels specifications and to 0 in the VAR in growth rates specification.
- For the intercept we assume an informative prior with mean 0 and standard deviation $\lambda_0 \sigma_i$.

Prior parametrization III

- The shrinkage parameter λ_1 measures the overall tightness of the prior: when $\lambda_1 \to 0$, the prior is imposed exactly and the data do not influence the estimates, while as $\lambda_1 \to \infty$ the prior becomes loose and the prior information does not influence the estimates, which will approach the standard OLS estimates.
- The parameter λ_2 implements additional shrinkage on lags of the other variables than for lags of the dependent variable. It is refered as the cross-shrinkage parameter, and in the baseline specification it is set it to $\lambda_2 = 1$, which means that the cross-shrinkage take place.
- The scale parameter $_i$ the common practice (see e.g. [Litterman, 1986] and [Sims and Zha, 1998]) is to set it equal to the standard deviation of the residuals from the univariate AR model.
- The use the following parametrization for the prior:

$$\lambda_0 = 1; \lambda_1 = 0.2; \text{and } \lambda_2 = 1, \tag{20}$$

Prior parametrization IV

- The prior beliefs in (19), defining the traditional Minnesota prior , only include the prior mean and variance of the coefficients , and do not elicit any prior beliefs about the correlation among the coefficients.
- [Doan et al., 1984] and [Sims, 1993] proposed to complement the prior beliefs in (17) with additional priors that favor unit roots and cointegration.

Table of Contents I

Table of Contents II

Forecasting with the BVAR

- We discuss forecasting with the BVAR model.
- First the proper method, which is computationally demanding in several cases for multi-step forecasting.
- Second we discuss two alternative approaches for multi-step forecasting, that tend to perform well for point forecast.

The proper method I

- Under standard N-IW prior, the full distribution of the one-step ahead forecasts is given by:

$$y'_{T+1} \mid x'_{T+1} \sim MT(x'_{T+1}\overline{\mathbf{\Phi}}, (x'_{T+1}\overline{\Omega}_{T+1} + 1)^{-1}, \overline{S}, \overline{\nu}) \tag{21}$$

  where MT denotes the matrix-variate-t distribution.
- Multi-step ahead forecasts is obtained by iteration and are not available in close form, but can be simulated using MCMC, that draws a sequence of Σ and parameters $\mathbf{\Phi}$ from (18) and shocks amd at each draw j computes the implied path of $\hat{y}^{(j)}_{t+h}$.
- An intuitive way to draw $\mathbf{\Phi}$, conditionally on a draw of the error variance Σ, is to vectorize and draw from a multivariate normal.

The proper method II

- A draw of $\mathbf{\Phi}$ from (18) is obtained as follows:

$$vec(\mathbf{\Phi}) = vec() + chol(\mathbf{\Sigma} \otimes \overline{\mathbf{\Omega}}) \times \nu \tag{22}$$

  where μ is a $MN \times 1$ standard Gaussian vector process, and $chol$ is the Cholesky decomposition that require $(MN)^3$ elementary operations.

- Organizing the elements of ν in a $M \times N$ matrix V, such that $\nu = vec(V)$, one could draw the matrix $\mathbf{\Phi}$ as follows:

$$\mathbf{\Phi} = \overline{\mathbf{\Phi}} + chol(\overline{\mathbf{\Omega}}) \times V \times chol(\mathbf{\Sigma})' \tag{23}$$

The pseudo-iterated approach

- Can choose to approximate the results by just integrating out the uncertainty in the coefficients and then using the posterior mean of the coefficients to porduce posterior means of the multi-steps forecasts.
- The multi-steps point forecasts is computed as:

$$\hat{y}_{t+h} = \mathbf{s} \cdot (\overline{\mathbf{\Phi}}^{+})^{h} x_{t+1} \tag{24}$$

- Empirical results in [Carriero et al., 2009] indicated that the gains for point forecasts from the proper simulation-based approach with respect to the pseudo-iterated method are negligible.

The direct forecasting approach

- To overcome the problem of non-linearity in the multi-steps forecasts is to use the direct approach.
- Cosider the following VAR:

$$\begin{aligned} y_t &= \Phi_{c,h} + \Phi_{1,h} y_{t-h} + \Phi_{2,h} y_{t-(h-1)} + \\ &+ \cdots + \Phi_{p,h} y_{t-(h-p+1)} + \epsilon_t \end{aligned} \tag{25}$$

- The iterated approach is more efficient under correct specification.
- But in the presence of mis-specification it is very dangerous, because the mis-specification will inflate with the forecasts horizon when the forecasts are computed recursively.
- [Carriero et al., 2016] show that in their empirical example and for horizons shorter than six-steps ahead, there is little loss in using the direct approach rather than the iterated one.

Table of Contents I

Table of Contents II

Example with simulated data I

- The DGP for the bivariate VAR(1) model is given by:

$$\begin{aligned} y_t &= 1 + 0.8y_{t-1} + \epsilon_{y,t} \\ x_t &= 1 + 0.6y_{t-1} + 0.5x_{t-1} + \epsilon_{x,t} \end{aligned} \tag{26}$$

- with

$$\upsilon_t = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} \epsilon_t \quad \text{and} \quad \upsilon_t \overset{i.i.d.}{\sim} N(0, I) \tag{27}$$

- Then $\epsilon_t \overset{i.i.d.}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} : \begin{bmatrix} 1.25 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \right)$.
- Also note that both variables are stationary because the eigenvalues of the matrix $\begin{bmatrix} 0.8 & 0 \\ 0.6 & 0.5 \end{bmatrix}$ are $\lambda_1 = 0.8$ and $\lambda_2 = 0.5$.
- The sample size is $T = 600$ but the estimation sample will be $T = 101, \cdots, 500$ and the forecast sample from $501, \cdots, 600$.

Example with simulated data II

- We estimate model (26) using classical and Bayesian methods for the correct model, i.e. $VAR(1)$, and the mis-specified model $VAR(4)$.
- The estimates for the $VAR(1)$ are given below:

Example with simulated data III

Vector Autoregression Estimates
Sample: $101, \ldots, 500$

	Y_t	X_t
Y_{t-1}	0.7900	0.5968
	(0.0348)	(0.0308)
	[22.6752]	[19.3587]
X_{t-1}	−0.0462	0.5236
	(0.0305)	(0.0270)
	[−1.5162]	[19.4250]
C	1.4118	0.8555
	(0.2428)	(0.2149)
	[5.8140]	[3.9813]

R-squared	0.5940	0.7586
Adj. R-squared	0.5919	0.7574
Sum sq. resids	484.5398	379.4251
S.E. equation	1.1048	0.9776
F-statistic	290.4038	623.8913
Log likelihood	−607.2576	−557.0139
Akaike AIC	3.0446	2.8001
Schwarz SC	3.0746	2.8300

Example with simulated data IV

Mean dependent	4.9706	8.0149
S.D. dependent	1.7295	1.9849

Determinant resid covariance (dof adj.)	0.9319
Determinant resid covariance	0.9180
Log likelihood	−1118.0350
Akaike information criterion	5.6202
Schwarz criterion	5.6800
Number of coefficients	6

- The estimates for the $VAR(4)$ are given below:

Example with simulated data V

Vector Autoregression Estimates
Date: 11/21/22 Time: 16:08
Sample: 101 500
Included observations: 400
Standard errors in ( ) & t-statistics in [ ]

	Y	X
Y(-1)	0.835636	0.646022
	(0.05642)	(0.04973)
	[14.8120]	[12.9905]
Y(-2)	−0.089510	−0.101776
	(0.09441)	(0.08322)
	[-0.94808]	[-1.22292]
Y(-3)	0.095518	−0.060413
	(0.09422)	(0.08305)
	[1.01378]	[-0.72739]
Y(-4)	−0.150502	0.164286
	(0.07862)	(0.06931)
	[-1.91424]	[2.37046]

Example with simulated data VI

X(-1)	0.007471	0.576246
	(0.06412)	(0.05652)
	[0.11653]	[10.1959]
X(-2)	−0.044995	−0.048726
	(0.07403)	(0.06526)
	[-0.60776]	[-0.74663]
X(-3)	0.115892	−0.069054
	(0.07374)	(0.06500)
	[1.57156]	[-1.06229]
X(-4)	−0.037314	0.023130
	(0.04437)	(0.03911)
	[-0.84100]	[0.59138]
C	1.203275	0.937936
	(0.30920)	(0.27256)
	[3.89151]	[3.44117]

Example with simulated data VII

R-squared	0.598969	0.763425
Adj. R-squared	0.590764	0.758585
Sum sq. resids	478.5969	371.8880
S.E. equation	1.106360	0.975254
F-statistic	72.99847	157.7194
Log likelihood	−603.4543	−553.0011
Akaike AIC	3.062271	2.810005
Schwarz SC	3.152079	2.899813
Mean dependent	4.970602	8.014926
S.D. dependent	1.729457	1.984886

Determinant resid covariance (dof adj.)	0.930783
Determinant resid covariance	0.889369
Log likelihood	−1111.702
Akaike information criterion	5.648511
Schwarz criterion	5.828127
Number of coefficients	18

Table of Contents I

Table of Contents II

BVAR(1) - Minnesota prior and others I

- Minnesota prior setup λ has a Gamma hyperprior and is handed upper and lower bounds for its Gaussian proposal distribution in the Metropolis-Hasting step α parameter is treated as fixed via the mode argument.
- The prior variance on the constant term of the model (var) is dealt a large value, for a diffuse prior.
- The prior mean for the var-cov matrix Σ, is Φ is left to be set automatically - i,e., to the square root of the innovation variance, after fitting AR(p) models to each of the variables.
- The command in R is:
  - mn $<-$ bv_minnesota(lambda $=$ bv_lambda(mode=0.2, sd =0.4, min $=$ 0.0001, max =5), alpha $=$ bv_alpha(mode=2), var $=$ 1e07)
- Could use instead the sum-of-coefficients ([Doan et al., 1984]) or a single-unit-roots priors ([Sims and Zha, 1998].

BVAR(1) - Minnesota prior and others II

- The hyperprior of their key hyperparameters are assigned Gamma distributions, with specification working in the same way as for λ.
- The command in R is:
  - soc $<-$ bv_soc(mode $=$ 1, sd $=$ 1, min $=$ 1e-04, max $=$ 50)
  - sur $<-$ bv_sur(mode $=$ 1, sd $=$ 1, min $=$ 1e-04, max $=$ 50)
- Via hyper we choose which hyperparameters should be treated hierarchically.
- The default is "auto" which includes λ and the key hyperparameters of all provided dummy-observation priors.
- This is equivalent to providing the character vector c("lambda", "soc","sur") α are treated as fixed and set equal to their mode.
- The command in R is:
  - priors $<-$ bv_priors(hyper="auto", mn $=$ mn, soc $=$ soc, sur $=$ sur)

Adjust the MH algorithm via bv_metropolis()

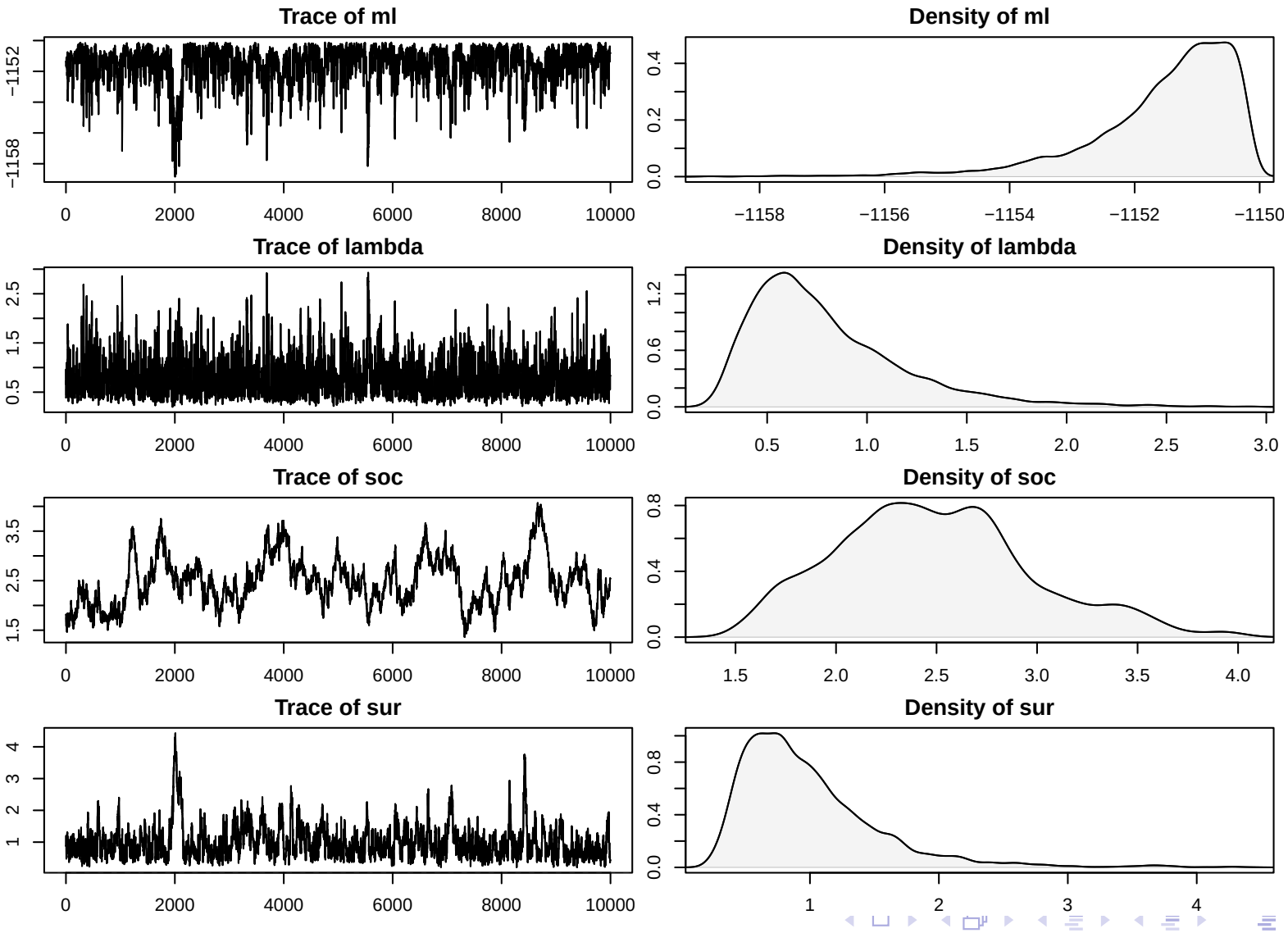
- The primary argument is scale_hess, which allows scaling the inverse Hessian, which is used as VCOV matrix of the Gaussian proposal distribution.
- The scalling can use adjust_cc = TRUE which enables automatic scale adjustment.
- This is done in the initial share of the burn-in period via adjust_burn.
- Automatic adjustment is performed iteratively by acc_change percent, until an acceptance rate between acc_lower and acc_upper is reached.
- The command in R is:
  - mh $<-$ bv_metropolis(scale= C90,05, 0.001, 0.0001), adjust_acc = TRUE, acc_lower = 0.25, acc_upper = 0.45)

Now bvar() is the BAVR specification I

- The BVAR specification uses the command bvar().
- The order of the VAR is defined in **lags**.
- The total number of iterations with **n_draw**.
- The number of initial iterations to discard with **n_burns**, and a fraction of drawns to be store via **n_thin**.
- When estimating the model, **verbose = TRUE** prompts printing of intermediate results.
- The command in R is:
  - sim.est$bvar1 < − bvar(sim.data[1:400]. lags =1, n_draw = 15000, n_burn = 5000, n_thin = 1, priors = pirors, mh = mh, verbose = TRUE)
- To assess convergence of the MCMC algorithm

Now bvar() is the BAVR specification II

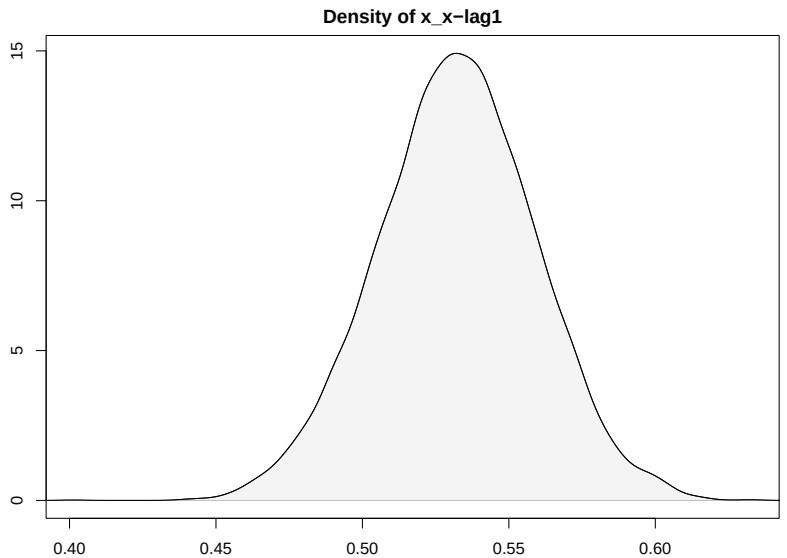
Figure: Convergence of the MCMC algorithm



Posterior Distribution for own lag I

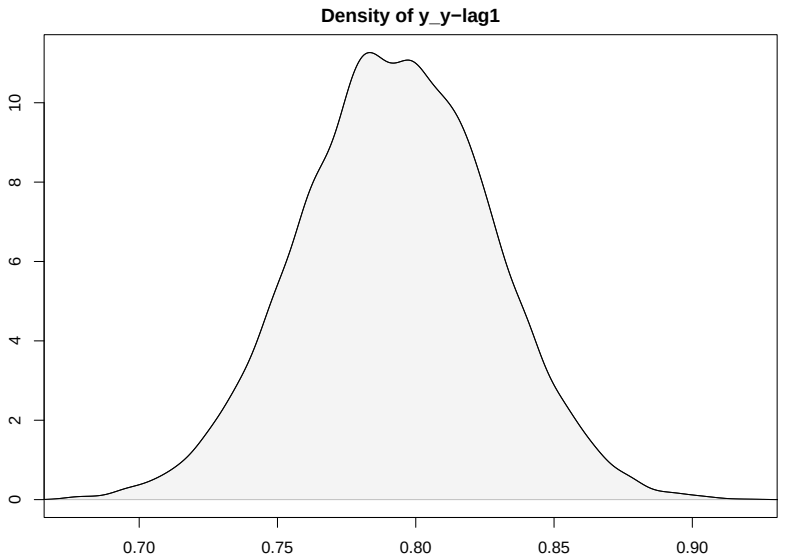
- Plot the posterior density of the first lag of each variable in each equation

Figure: Posterior For x_{t-1}



Posterior Distribution for own lag II

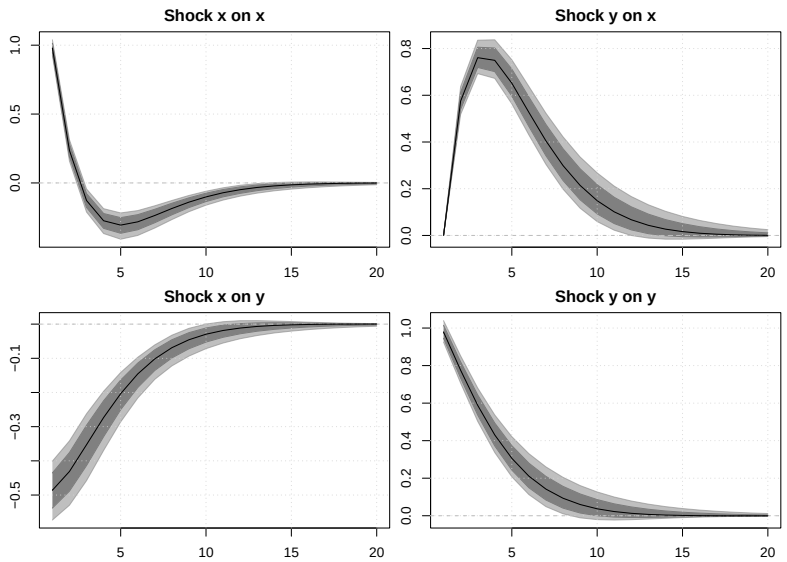
Figure: Posterior For y_{t-1}



IRF

- Can construct and plot the IRF
- The command in R is:
  - sim_opt_irf $<-$ bv_irf(horizon $=$ 20, identification $=$ TRUE)
  - irf(sim.est$bvar1) $<-$ irf(sim.est$bvar1, sim_opt_irf, conf_bands =c(0.05,0.16))
  - plot(irf(sim.est$bvar1), area $=$ TRUE, var_impulse $=$ c("y","x"))

Figure: IRF for BVAR(1)



Forecasting with BVAR(1)

- Can construct forecast using **predict()**
- The command in R is:
  - sim.fore$bvar1 <− predict(sim.est$bvar1, horizon = 100, conf_bands = c(0.05, 0.16))
  - Retain the median forecast
  - sim.fore$bvar1 <− data.table(sim.fore$bvar1$quant[3,1:100,1:2])
  - setnames(sim.fore$bvar1, c('x', 'y'))

Evaluation of VAR(1), VAR(4), BVAR(1) and BVAR(4)

Table: Comparison for the x variable

	Model	RMSE	MAE
1	var1	2.47703	2.06357
2	var4	2.47802	2.06732
3	bvar1	**2.47454**	**2.06066**
4	bvar4	2.48000	2.06923

Table: Comparison for the y variable

	Model	RMSE	MAE
1	var1	1.94030	1.63363
2	var4	1.93879	1.63278
3	bvar1	1.94293	1.63604
4	bvar4	**1.93661**	**1.62850**

Table of Contents I

Table of Contents II

Reference I


Carriero, A., Clark Todd, E., and Marcellino, M. (2016).
Large vector autoregressions with stochastic volatility and flexible priors.
Technical report, Federal Reserve Bank of Cleveland.


Carriero, A., Kapetanios, G., and Marcellino, M. (2009).
Forecasting exchange rates with a large bayesian VAR.
International Journal of Forecasting, 25(2):400–417.


Doan, T., Litterman, R., and Sims, C. (1984).
Forecasting and conditional projection using realistic prior distributions.
Econometric Reviews, 3(1):1–100.

Reference II


Kadiyala, K. R. and Karlsson, S. (1997).
Numerical methods for estimation and inference in bayesian var-models.
Journal of Applied Econometrics, 12(2):99–132.


Koop, G. (2003).
Bayesian Econometrics.
Wiley-Interscience.


Koop, G. M. (2013).
Forecasting with medium and large bayesian VARS.
28(2):177–203.


Litterman, R. B. (1986).
Forecasting with bayesian vector autoregressions: Five years of experience.
Journal of Business &amp Economic Statistics, 4(1):25.

Reference III


Marcellino, M. G., Stock, J. H., and Watson, M. W. (2006).
A compariosn of direct and iterated multistep ar methos for forecasting macroeconomic time series.
Journal of Econometrics, 18:427–443.


Sims, C. A. (1993).
A nine-variable probabilistic macroeconomic forecast model, pages 179 – 212.


Sims, C. A. and Zha, T. (1998).
Bayesian methods for dynamic multivariate models.
International Economic Review, 39(4):949.


Theil, H. and Goldberger, A. S. (1961).
On pure and mixed statistical estimation in economics.
International Economic Review, 2(1):65.