Bayesian Econometrics

Lecture 6: Bayesian Estimation of VAR Models with DSGE Priors

Prof. Dr. Kai Carstensen

Kiel University

Winter Term 2024/25

Idea

This lecture shows how to estimate VAR models in a Bayesian fashion using priors derived from a DSGE model.

Why may this be a good way to proceed?

VAR models are good at capturing the observed dynamic relationships between important macro variables. But they tend to be overparameterized and they are atheoretic and thus difficult to interpret.

DSGE models are derived from theory and can easily be used to study how a model economy reacts to all kinds of shocks. However, they typically depend only on a small set of parameters and are quite restrictive. Hence, they have difficulties to capture the observed dynamics.

Solution: Inform VAR models with DSGE prior. Defining a parameter $0 \leq \lambda < \infty$ of prior strictness, we can move between the polar cases of a pure VAR model ($\lambda = 0$) and a purely theoretical model ($\lambda \to \infty$).

Structural analysis: Del Negor and Schorfheide (2004) show that a VAR model with DSGE prior

- yields sensible "deep" DSGE parameter estimates,
- outperforms a VAR model with Minnesota prior in terms of forecasting,
- generates sensible impulse response functions,
- can be used to "correct" VAR forecasts after a policy shift.

Forecasting: In a data-rich environment, VAR models require shrinkage to be useful for forecasting and/or estimable at all. DSGE priors are a way to achieve theory-based shrinkage.

Outline of this lecture

1. Population moments of DSGE models
2. BVAR based on a DSGE-prior for given θ
3. BVAR based on a DSGE-prior for given distribution of θ
4. Empirical application

Literature

References:

- Del Negro and Schorfheide (2004) Priors from general equilibrium models for VARs, International Economic Review 45(2), 643-673.
- Herbst and Schorfheide (2015), Bayesian Estimation of DSGE Models, Princeton University Press. An earlier version of the book can be downloaded here.

1. Population moments of DSGE models

Introduction

This lecture is not on how to derive, linearize, and solve DSGE models. A good introduction can be found in J. Gali (2015), Monetary Policy, Inflation, and the Business Cycle, Princeton University Press. A nice (but brief) overview over solution techniques can be found in D.N. DeJong and C. Dave (2007), Structural Macroeconometrics, Princeton University Press.

Here we will only state the log-linearized baseline three-equation New Keynesian DSGE model used by Del Negro and Schorfheide (2004), state a solution of the model, and discuss in more detail how this solution can be used to specify a VAR prior.

A baseline New Keynesian DSGE model

The model includes output (GDP), x_t, quarterly gross inflation, π_t, and the quarterly gross nominal interest rate, R_t. A tilde denotes log deviations from steady state. Hence, $\tilde{x}_t$ is interpreted as output gap. The log-linearized model consists of a consumption Euler equation (New Keynesian IS curve), a New Keynesian Phillips curve, and a monetary policy rule:

$$\tilde{x}_t = \mathrm{E}_t\,\tilde{x}_{t+1} - \frac{1}{\tau}(\tilde{R}_t - \mathrm{E}_t\,\tilde{\pi}_{t+1}) + (1-\rho_g)\tilde{g}_t + \frac{\rho_z}{\tau}\tilde{z}_t \tag{1}$$

$$\tilde{\pi}_t = \frac{\gamma}{r^*}\,\mathrm{E}_t\,\tilde{\pi}_{t+1} + \kappa\tilde{x}_t - \kappa\tilde{g}_t \tag{2}$$

$$\tilde{R}_t = \rho_R\tilde{R}_{t-1} + (1-\rho_R)(\psi_1\tilde{\pi}_t + \psi_2\tilde{x}_t) + \tilde{h}_t \tag{3}$$

with shock processes

$$\tilde{z}_t = \rho_z\tilde{z}_{t-1} + \epsilon_{z,t}, \qquad \epsilon_{z,t} \overset{iid}{\sim} \mathcal{N}(0,\sigma_z^2) \tag{4}$$

$$\tilde{g}_t = \rho_g\tilde{g}_{t-1} + \epsilon_{g,t}, \qquad \epsilon_{g,t} \overset{iid}{\sim} \mathcal{N}(0,\sigma_g^2) \tag{5}$$

$$\tilde{h}_t = \epsilon_{R,t}, \qquad \epsilon_{R,t} \overset{iid}{\sim} \mathcal{N}(0,\sigma_R^2). \tag{6}$$

To match the model to observed variables, add the measurement equations

$$\Delta \log x_t = \log \gamma + \Delta \tilde{x}_t + \tilde{z}_t \tag{7}$$

$$\Delta \log P_t = \log \pi^* + \tilde{\pi}_t \tag{8}$$

$$\log R_t^a = 4(\log r^* + \log \pi^*) + 4\tilde{R}_t, \tag{9}$$

where R_t^a is the annualized interest rate, γ is the trend growth rate, π^* is the equilibrium inflation rate, and r^* is the steady state real interest rate.

The 13 model parameters can be summarized in the vector

$$\theta = [\log \gamma, \log \pi^*, \log r^*, \kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_z, \sigma_R, \sigma_g, \sigma_z]'.$$

Solution (*)

We may use the method of undetermined coefficients of Uhlig (1999) who shows that the general model

$$0 = \mathrm{E}_t \left[F q_{t+1} + G q_t + H q_{t-1} + M f_t \right] \tag{10}$$

$$f_{t+1} = N f_t + \epsilon_{t+1}, \qquad \mathrm{E}_t(\epsilon_{t+1}) = 0 \tag{11}$$

is solved as

$$f_{t+1} = N f_t + \epsilon_{t+1} \tag{12}$$

$$q_t = P q_{t-1} + Q f_t \tag{13}$$

where P and Q are functions of the model parameters and satisfy $FP^2 + gP + H = 0$ and $(FP + G)Q + M + FQN = 0$, see Uhlig (1999) or DeJong and C. Dave (2007) for a derivation.

To cast the model (1)-(6) in the "solution form", we define the variables

$$q_t = \begin{bmatrix} \tilde{x}_t \\ \tilde{\pi}_t \\ \tilde{R}_t \end{bmatrix}, \quad f_t = \begin{bmatrix} \tilde{z}_t \\ \tilde{g}_t \\ \tilde{h}_t \end{bmatrix}, \quad \text{and} \quad \epsilon_t = \begin{bmatrix} \epsilon_{z,t} \\ \epsilon_{g,t} \\ \epsilon_{R,t} \end{bmatrix}.$$

and the implied coefficient matrices

$$F = \begin{bmatrix} 1 & \frac{1}{\tau} & 0 \\ 0 & \frac{\gamma}{r^*} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} -1 & 0 & -\frac{1}{\tau} \\ \kappa & -1 & 0 \\ (1-\rho_R)\psi_2 & (1-\rho_R)\psi_1 & -1 \end{bmatrix}$$

$$H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \rho_R \end{bmatrix}, \quad M = \begin{bmatrix} \frac{\rho_z}{\tau} & 1-\rho_g & 0 \\ 0 & -\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad N = \begin{bmatrix} \rho_z & 0 & 0 \\ 0 & \rho_g & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then we can compute (using, e.g., Matlab) the numerical solution matrices P and Q.

It is helpful to write the solution (12) and (13) together with the measurement equations (7)-(9) as a state system of the form

$$s_t = Ts_{t-1} + Re_t, \qquad \Sigma_e = \mathrm{E}(e_t e_t') \tag{14}$$

$$y_t = D + Zs_t \tag{15}$$

For the state equation, define $s_t = [f_{t+1}, f_t, q_t, q_{t-1}]'$, $e_t = \epsilon_{t+1}$,

$$T = \begin{bmatrix} N & 0 & 0 & 0 \\ I_3 & 0 & 0 & 0 \\ Q & 0 & P & 0 \\ 0 & 0 & I_3 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} I_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma_e = \begin{bmatrix} \sigma_z^2 & 0 & 0 \\ 0 & \sigma_g^2 & 0 \\ 0 & 0 & \sigma_R^2 \end{bmatrix}.$$

For the measurement equation, define $y_t = [\Delta \log x_t, \Delta \log P_t, \log R_t^a]'$,

$$D = \begin{bmatrix} \log \gamma \\ \log \pi^* \\ 4(\log r^* + \log \pi^*) \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & -4 \end{bmatrix}.$$

Population moments of the DSGE model

Suppose we have found the state space representation of the DSGE model:

$$
\begin{aligned}
s_t &= T s_{t-1} + R e_t, \qquad \Sigma_e = \mathrm{E}(e_t e_t') \\
y_t &= D + Z s_t.
\end{aligned}
$$

The state equation is a VAR(1) with zero mean. Its variance $\Omega_s \equiv \mathrm{E}(s_t s_t')$ can be computed as (see Lütkepohl, 2005, p. 26-27)

$$
\operatorname{vec} \Omega_s = (I - T \otimes T)^{-1} \operatorname{vec}(R \Sigma_e R'),
$$

and the autocovariances of order $h = 0, 1, 2, \ldots$ as

$$
\Gamma_{ss}(h) \equiv \mathrm{Cov}(s_t, s_{t-h}) = \mathrm{E}(s_t s_{t-h}) = T^h \Omega_s.
$$

The mean of the observable variables y_t is $\mu^* \equiv \mathrm{E}(y_t) = D$ and the (uncentered!) autocovariances are

$$
\Gamma^*_{yy}(h) \equiv \mathrm{E}(y_t y_{t-h}) = DD' + Z\,\mathrm{E}(s_t s_{t-h}')Z' = DD' + Z T^h \Omega_s Z'.
$$

2. BVAR based on a DSGE-prior for given θ

How to find DSGE-based prior for a VAR

Following Del Negro and Schorfheide, a DSGE model with given parameter vector θ can be turned into a VAR prior as follows:

- Let us concentrate on a natural conjugate Normal-inverse Wishart prior. Recall that it can be given a fictitious prior sample interpretation.
- Use the DSGE model to generate, for given θ, a fictitious prior sample of size $T^* = \lambda T$, where $0 \leq \lambda < \infty$ governs the strength of the prior.
- Use these fictitious data to estimate a BVAR with improper prior. As shown in the previous chapter, this yields a Normal-inverse Wishart posterior.
- This "fictitious" posterior (it is based on the fictitious sample) is then taken as a prior for the actual sample and yields again a NiW posterior.

$$\boxed{\begin{array}{l}\text{Improper}\\\text{prior}\end{array}} \rightarrow \boxed{\begin{array}{l}\text{fictitious}\\\text{sample}\end{array}} \rightarrow \boxed{\begin{array}{l}\text{fictitious}\\\text{NiW}\\\text{posterior}\end{array}} = \boxed{\begin{array}{l}\text{NiW}\\\text{prior}\end{array}} \rightarrow \boxed{\begin{array}{l}\text{actual}\\\text{sample}\end{array}} \rightarrow \boxed{\begin{array}{l}\text{NiW}\\\text{posterior}\end{array}}$$

Specifying the prior using a simulated sample

Recall that a VAR(p) model with M variables,

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \Sigma),$$

can be written as

$$y_t' = x_t A + \varepsilon_t', \qquad x_t \equiv [1, y_{t-1}', \ldots, y_{t-p}'], \quad A \equiv [a_0, A_1, \ldots, A_p]'.$$

To estimate it using the fictitious—i.e., simulated—data (which we denote by a *), we stack observations and write the VAR in matrix form:

$$Y^* = X^* A + E^*, \qquad Y^* = [y_1^*, \ldots, y_{T^*}^*]', \quad X^* \equiv [x_1^{*\prime}, \ldots, x_{T^*}^{*\prime}]'.$$

For further use, let us define the fictitious sample moments

$$\hat{\Gamma}_{yy}^* = \frac{1}{T^*} Y^{*\prime} Y^*, \quad \hat{\Gamma}_{xy}^* = \frac{1}{T^*} X^{*\prime} Y^*, \quad \text{and} \quad \hat{\Gamma}_{xx}^* = \frac{1}{T^*} X^{*\prime} X^*.$$

Now suppose we estimate $\alpha = \text{vec}\, A$ and Σ from this fictitious sample by Bayesian methods. Specifically, let us use the improper prior $p(\alpha, \Sigma) \propto |\Sigma|^{-\frac{M+1}{2}}$.

We found in the previous chapter that this yields the (here: "fictitious") posterior

$$\alpha|\Sigma, y^*, \theta \sim \mathcal{N}(\text{vec}(\hat{A}^*), \Sigma \otimes (X^{*\prime}X^*)^{-1})$$

and

$$\Sigma|y^*, \theta \sim iW([Y^* - X^*\hat{A}^*]'[Y^* - X^*\hat{A}^*], T^* - pM - 1)$$

where we condition on θ to emphasize that it holds for a given choice of θ.

We can use this "fictitious" posterior as the natural conjugate prior for the VAR we estimate using the actual sample. This is possible because the natural conjugate prior is Normal-inverse Wishart.

The problem with this simulation approach is that it yields a "random" natural conjugate prior, particularly for small fictitious samples, because the simulated DSGE model is stochastic. To remove this stochastic variation Del Negro and Schrofheide suggest to replace simulated sample moments by the population moments implied by the DSGE model as we show next.

Specifying the prior using population moments of the DSGE model

Consider the parameters of the "fictitious" posterior

$$\alpha|\Sigma, y^*, \theta \sim \mathcal{N}(\text{vec}(\hat{A}^*), \Sigma \otimes (X^{*\prime}X^*)^{-1})$$

and

$$\Sigma|y^*, \theta \sim iW([Y^* - X^*\hat{A}^*]'[Y^* - X^*\hat{A}^*], T^* - pM - 1).$$

Next we show that all sample quantities can be expressed as functions of the sample moments $\hat{\Gamma}^*_{yy}$, $\hat{\Gamma}^*_{xy}$, and $\hat{\Gamma}^*_{xy}$.

The posterior mean of A is

$$\hat{A}^* = \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}Y^* = \left(\frac{X^{*\prime}X^*}{T^*}\right)^{-1} \frac{X^{*\prime}Y^*}{T^*} = \hat{\Gamma}_{xx}^{*-1}\hat{\Gamma}_{xy}^*.$$

The posterior variance of A is

$$\Sigma \otimes (X^{*\prime}X^*)^{-1} = \Sigma \otimes (T^*\hat{\Gamma}_{xx}^*)^{-1}.$$

Using OLS algebra, the first parameter of the iW distribution can be written as

$$\begin{aligned}
[Y^* - X^*\hat{A}^*]'[Y^* - X^*\hat{A}^*] &= Y^{*\prime}(I - X^*(X^{*\prime}X^*)^{-1}X^{*\prime})Y^* \\
&= T^*\left(\frac{Y^{*\prime}Y^*}{T^*} - \frac{Y^{*\prime}X^*}{T^*}\left(\frac{X^{*\prime}X^*}{T^*}\right)^{-1}\frac{X^{*\prime}Y^*}{T^*}\right) \\
&= T^*\left(\hat{\Gamma}_{yy}^* - \hat{\Gamma}_{yx}^*\hat{\Gamma}_{xx}^{*-1}\hat{\Gamma}_{xy}^*\right)
\end{aligned}$$

Taken together, the "fictitious" posterior is written as

$$\alpha|\Sigma, y^*, \theta \sim \mathcal{N}(\text{vec}(\hat{\Gamma}_{xx}^{*-1}\hat{\Gamma}_{xy}^*), \Sigma \otimes (T^*\hat{\Gamma}_{xx}^*)^{-1})$$

$$\Sigma|y^*, \theta \sim iW\left(T^*\left(\hat{\Gamma}_{yy}^* - \hat{\Gamma}_{yx}^*\hat{\Gamma}_{xx}^{*-1}\hat{\Gamma}_{xy}^*\right), T^* - pM - 1\right).$$

Now let us replace the sample moments by the population moments $\Gamma_{yy}^*(\theta)$, $\Gamma_{xy}^*(\theta)$, and $\Gamma_{xy}^*(\theta)$ of the DSGE model which generated the fictitious sample. To emphasize that they depend on a specific choice of θ, we write the moments as functions of θ.

We obtain

$$\alpha|\Sigma, \theta \sim \mathcal{N}(\text{vec}(\Gamma_{xx}^{*-1}(\theta)\Gamma_{xy}^*(\theta)), \Sigma \otimes (T^*\Gamma_{xx}^*(\theta))^{-1})$$

$$\Sigma|\theta \sim iW\left(T^*\left(\Gamma_{yy}^*(\theta) - \Gamma_{yx}^*(\theta)\Gamma_{xx}^{*-1}(\theta)\Gamma_{xy}^*(\theta)\right), T^* - pM - 1\right).$$

Discussion

- We show below that the population moments $\Gamma^*_{yy}(\theta)$, $\Gamma^*_{xy}(\theta)$, and $\Gamma^*_{xy}(\theta)$ are functions of the DSGE model's population mean μ^* and its uncentered autocovariances $\Gamma^*_{yy}(h)$, $h = 0, \ldots, p$ alone which we derived on slide 13.
- We also show that the prior mean for A equals the parameter matrix $A^*(\theta)$ of a VAR(p) that approximates the DSGE model in population, i.e., $A^*(\theta) = \Gamma^{*-1}_{xx}(\theta)\Gamma^*_{xy}(\theta)$.
- And we show that the first parameter of the inverse Wishart distribution equals T^* times the error variance matrix $\Sigma^*(\theta)$ of this approximating VAR, i.e., $T^*\Sigma^*(\theta) = T^*\left(\Gamma^*_{yy}(\theta) - \Gamma^*_{yx}(\theta)\Gamma^{*-1}_{xx}(\theta)\Gamma^*_{xy}(\theta)\right)$.
- Hence, in terms of such an approximating VAR, the prior becomes

$$\alpha|\Sigma, y^*, \theta \sim \mathcal{N}(\text{vec}(A^*(\theta)), \Sigma \otimes (T^*\Gamma^*_{xx}(\theta))^{-1})$$

$$\Sigma|y^*, \theta \sim iW\left(T^*\Sigma^*(\theta), T^* - pM - 1\right).$$

- This implies in particular, that the prior mean for A is the population parameter $A^*(\theta)$ of a VAR(p) model that approximates the DSGE model.

The DSGE prior in previous notation

Here is the prior in terms of the notation used in the previous chapter:

$$\alpha|\Sigma,\theta \sim \mathcal{N}(\text{vec}(\bar{A}), \Sigma \otimes \underline{V})$$

$$\Sigma|\theta \sim iW(\underline{S}, \underline{\nu}),$$

where

$$\underline{V} = (T^*\Gamma^*_{xx}(\theta))^{-1} = \frac{1}{\lambda T}\Gamma^{*-1}_{xx}(\theta) \tag{16}$$

$$\underline{A} = \Gamma^{*-1}_{xx}(\theta)\Gamma^*_{xy}(\theta) = A^*(\theta) \tag{17}$$

$$\underline{S} = T^*\left(\Gamma^*_{yy}(\theta) - \Gamma^*_{yx}(\theta)\Gamma^{*-1}_{xx}(\theta)\Gamma^*_{xy}(\theta)\right) = \lambda T\Sigma^*(\theta) \tag{18}$$

$$\underline{\nu} = T^* - pM - 1 = \lambda T - pM - 1 \tag{19}$$

Posterior

Based on the DSGE prior of the NiW type we can now use the actual sample $Y = [y_1, \ldots, y_{T^*}]'$ and $X = [x_1', \ldots, x_{T^*}']'$ with moments $\hat{\Gamma}_{yy} = \frac{1}{T} Y'Y$, $\hat{\Gamma}_{xy} = \frac{1}{T} X'Y$, and $\hat{\Gamma}_{xx} = \frac{1}{T} X'X$ to estimate the VAR, i.e., to find the posterior.

Since the prior is natural conjugate, the posterior is NiW with parameters that can be computed analytically:

$$\alpha | y, \Sigma, \theta \sim \mathcal{N}(\text{vec}(\bar{A}, \Sigma \otimes \bar{V})$$

$$\Sigma | y, \theta \sim iW(\bar{S}, \bar{\nu}),$$

where

$$\bar{V} = (\underline{V}^{-1} + X'X)^{-1} = (\lambda T \Gamma_{xx}^{*-1}(\theta) + T \hat{\Gamma}_{xx})^{-1} \tag{20}$$

$$\bar{A} = \bar{V}(\underline{V}^{-1}\underline{A} + X'X\hat{A}) = \bar{V}(\lambda T \Gamma_{xy}^{*}(\theta) + T\hat{\Gamma}_{xy}) \tag{21}$$

$$\bar{S} = \underline{S} + \underline{A}'\underline{V}^{-1}\underline{A} + Y'Y - \bar{A}'\bar{V}^{-1}\bar{A} = \lambda T \Gamma_{yy}^{*} + T\hat{\Gamma}_{yy} - \bar{A}'\bar{V}^{-1}\bar{A} \tag{22}$$

$$\bar{\nu} = T + \underline{\nu} = T + \lambda T - pM - 1 = (1+\lambda)T - pM - 1 \tag{23}$$

The BVAR estimation procedure

We now state the full BVAR estimation procedure for a given parameter vector θ with DSGE prior:

- Choose the structural DSGE parameters θ and the fictitious sample size $T^* = \lambda T$.
- Solve the DSGE model and compute the population mean and autocovariances of its observables y_t.
- Based on these moments compute a Normal-inverse Wishart prior with parameters (16)-(19).
- Add the data of your sample and obtain the Normal-inverse Wishart posterior with parameters (20)-(23).

Note that λ is assumed fixed here. However, Del Negro and Schorfheide (2004) show how to estimate it by maximizing the data density.

Discussion

- While this is a perfectly fine BVAR estimation procedure, there is an important problem: Most people are not so sure about the parameters of their DSGE model that they would come up with just one particular choice of θ.
- Even if they would be sure, two researchers would probably choose two different sets of values for θ. Which results should policy then use?
- From a Bayesian perspective, it is straightforward to solve this problem: specify prior distributions for θ instead of using fixed values.
- The drawback of this approach is, however, that it departs from simple natural conjugate BVAR analysis and requires simulation techniques to obtain the posterior. In the next section we briefly sketch this approach.

Proof 1 (*)

$\Gamma^*_{yy}(\theta)$, $\Gamma^*_{xy}(\theta)$, and $\Gamma^*_{xy}(\theta)$ are functions of μ^* and $\Gamma^*_{yy}(h)$, $h = 0, \ldots, p$ alone

Part A: $\Gamma^*_{yy}(\theta)$ is the population counterpart of $\hat{\Gamma}^*_{yy} = \frac{1}{T^*} Y^{*\prime} Y^* = \frac{1}{T^*} \sum_{t=1}^{T^*} y^*_t y^{*\prime}_t$.
Hence, $\Gamma^*_{yy}(\theta) = \mathrm{E}(y^*_t y^{*\prime}_t)$ and equals $\Gamma^*_{yy}(1)$ derived on slide 13.

Part B: $\Gamma^*_{xy}(\theta)$ is the population counterpart of $\hat{\Gamma}^*_{xy} = \frac{1}{T^*} X^{*\prime} Y^* = \frac{1}{T^*} \sum_{t=1}^{T^*} x^{*\prime}_t y^{*\prime}_t$.
Hence, $\Gamma^*_{xy}(\theta) = \mathrm{E}(x^{*\prime}_t y^{*\prime}_t)$. It relates to μ^* and $\Gamma^*_{yy}(h)$ as follows:

$$\Gamma^*_{xy}(\theta) = \mathrm{E}\begin{bmatrix} 1 \\ y^*_{t-1} \\ \vdots \\ y^*_{t-p} \end{bmatrix} y^{*\prime}_t = \begin{bmatrix} \mathrm{E}(y^*_t)' \\ \mathrm{E}(y^*_{t-1} y^{*\prime}_t) \\ \vdots \\ \mathrm{E}(y^*_{t-p} y^{*\prime}_t) \end{bmatrix} = \begin{bmatrix} \mu' \\ \Gamma^*_{yy}(-1) \\ \vdots \\ \Gamma^*_{yy}(-p) \end{bmatrix} = \begin{bmatrix} \mu' \\ \Gamma^*_{yy}(1)' \\ \vdots \\ \Gamma^*_{yy}(p)' \end{bmatrix}$$

where we used $\Gamma^*_{yy}(-h) = \mathrm{E}(y^*_{t-h} y^{*\prime}_t) = [\mathrm{E}(y^*_t y^{*\prime}_{t-h})]' = \Gamma^*_{yy}(h)'$.

Part C: $\Gamma^*_{xx}(\theta)$ is the population counterpart of $\hat{\Gamma}^*_{xx} = \frac{1}{T^*}X^{*\prime}X^* = \frac{1}{T^*}\sum_{t=1}^{T^*} x_t^{*\prime}x_t^*$. Hence, $\Gamma^*_{xx}(\theta) = \mathrm{E}(x_t^{*\prime}x_t^*)$. It relates to μ^* and $\Gamma^*_{yy}(h)$ as follows:

$$
\begin{aligned}
\Gamma^*_{xx}(\theta) &= \mathrm{E}\begin{bmatrix} 1 \\ y^*_{t-1} \\ \vdots \\ y^*_{t-p} \end{bmatrix} [1, y^*_{t-1}, \ldots, y^*_{t-p}] \\
&= \mathrm{E}\begin{bmatrix} 1 & y^{*\prime}_{t-1} & y^{*\prime}_{t-2} & \cdots & y^{*\prime}_{t-p} \\ y^*_{t-1} & y^*_{t-1}y^{*\prime}_{t-1} & y^*_{t-1}y^{*\prime}_{t-2} & \cdots & y^*_{t-1}y^{*\prime}_{t-p} \\ y^*_{t-2} & y^*_{t-2}y^{*\prime}_{t-1} & y^*_{t-2}y^{*\prime}_{t-2} & \cdots & y^*_{t-2}y^{*\prime}_{t-p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y^*_{t-p} & y^*_{t-p}y^{*\prime}_{t-1} & y^*_{t-p}y^{*\prime}_{t-2} & \cdots & y^*_{t-p}y^{*\prime}_{t-p} \end{bmatrix} \\
&= \begin{bmatrix} 1 & \mu^{*\prime} & \mu^{*\prime} & \cdots & \mu^{*\prime} \\ \mu^* & \Gamma^*_{yy}(0) & \Gamma^*_{yy}(1) & \cdots & \Gamma^*_{yy}(p-1) \\ \mu^* & \Gamma^*_{yy}(1)' & \Gamma^*_{yy}(0) & \cdots & \Gamma^*_{yy}(p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu^* & \Gamma^*_{yy}(p-1)' & \Gamma^*_{yy}(p-2)' & \cdots & \Gamma^*_{yy}(0) \end{bmatrix}
\end{aligned}
$$

Proof 2 (*)

The prior mean for A equals the parameter matrix $A^*(\theta)$ of an approximating VAR(p)

Recall that any VAR(p) model can be written as

$$y_t' = x_t A + \varepsilon_t'.$$

Like in a simple regression model, the relationship between population parameters and population moments can be obtained by multiplying with x_t' from the left and taking expectations:

$$\mathrm{E}(x_t' y_t') = \mathrm{E}(x_t' x_t) A \qquad \Rightarrow \qquad A = \mathrm{E}(x_t' x_t)^{-1} \, \mathrm{E}(x_t' y_t') = \Gamma_{xx}^{-1} \Gamma_{xy}.$$

The error variance can be obtained by by multiplying with y_t from the left and taking expectations:

$$\mathrm{E}(y_t y_t') = \mathrm{E}(y_t x_t) A + \Sigma \qquad \Rightarrow \qquad \begin{aligned} \Sigma &= \mathrm{E}(y_t y_t') - \mathrm{E}(y_t x_t) \, \mathrm{E}(x_t' x_t)^{-1} \, \mathrm{E}(x_t' y_t') \\ &= \Gamma_{yy} - \Gamma_{yx} \Gamma_{xx}^{-1} \Gamma_{xy}. \end{aligned}$$

Hence, the VAR parameters A and Σ are fully determined by Γ_{yy}, Γ_{xy} and Γ_{xx}. (Recall: in Econometrics I we called this property identification.)

Now consider a DSGE model with population moments $E(y_t^*) = \mu^*$ and $E(y_t^* y_{t-h}^{*\prime}) = \Gamma_{yy}^*(h)$, $h = 0, \ldots, p$, from which we compute the population moments $\Gamma_{yy}^*(\theta)$, $\Gamma_{xy}^*(\theta)$, and $\Gamma_{xy}^*(\theta)$.

By construction, the VAR(p) model with parameters

$$A^*(\theta) = \Gamma_{xx}^{*-1}(\theta)\Gamma_{xy}^*(\theta)$$

and

$$\Sigma^*(\theta) = \Gamma_{yy}^*(\theta) - \Gamma_{yx}^*(\theta)\Gamma_{xx}^{*-1}(\theta)\Gamma_{xy}^*(\theta)$$

has the same mean and the same autocovariances of order 0 to h as the DSGE model. In this sense, it approximates the DSGE model. Of course, approximating VAR and DSGE model may differ in terms of higher order autocovariances and other moments.

The definitions of $A^*(\theta)$ and $\Sigma^*(\theta)$ are the same as in the DSGE prior. In particular, $A^*(\theta)$ is the prior mean for A.

3. BVAR based on a DSGE-prior derived from a prior for θ

Hierarchical priors

On a general level, a hierarchical prior is a prior that can be written in two (or more) conditioning steps. This is convenient in many situations. For example, think of the bivariate regression model with known variance, σ^2,

$$y_i = x_i\beta_i + \varepsilon_i, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, N.$$

Trying to estimate β_i, $i = 1, \ldots, N$ as N independent and unrestricted parameters would exhaust the degrees of freedom of the sample. So we may use the prior that each β_i is independently drawn from the same univariate distribution

$$\beta_i | \alpha \sim \mathcal{N}(\alpha, 1)$$

which only depends on the hyperparameter α for which we may assume another normal prior, say,

$$\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{v}^2).$$

Formally, we have the prior (defining $\beta = [\beta_1, \ldots, \beta_N]'$)

$$p(\beta, \alpha) = p(\beta|\alpha)p(\alpha).$$

Note that we do not necessarily need such a hierarchical prior because the data directly only depend on β only. We could just integrate α out:

$$p(\beta) = \int p(\beta, \alpha) d\alpha = \int p(\beta|\alpha)p(\alpha)d\alpha.$$

Using the rules of the normal distribution, the marginal prior for β is again normal. By the laws of iterated expectations, it has mean $\mathrm{E}(\beta_i) = \mathrm{E}[\mathrm{E}(\beta_i|\alpha)] = \mathrm{E}(\alpha) = \underline{\alpha}$ and covariance structure $\mathrm{Cov}(\beta_i, \beta_j) = \mathrm{E}[\mathrm{Cov}(\beta_i, \beta_j|\alpha)] + \mathrm{Cov}[\mathrm{E}(\beta_i|\alpha), \mathrm{E}(\beta_j|\alpha)] = 1_{(i=j)} + \underline{v}^2$. So we could have used

$$\beta \sim \mathcal{N}(\alpha\mathbf{1}, I + \mathbf{1}\mathbf{1}'\underline{v}^2), \qquad \mathbf{1} = [1, \ldots, 1]',$$

right away but would probably not come up with this specific covariance matrix.

Hence, a hierarchical prior can be an efficient tool to specify plausible priors.

A hierarchical DSGE-prior for the VAR

In the previous section, we introduced—for given θ—the NiW prior

$$p(\alpha, \Sigma|\theta) = p(\alpha|\Sigma, \theta)p(\Sigma|\theta).$$

Now let us specify a prior for the DSGE parameters θ and call it $p(\theta)$, which yields the prior

$$p(\alpha, \Sigma, \theta) = p(\alpha, \Sigma|\theta)p(\theta) = p(\alpha|\Sigma, \theta)p(\Sigma|\theta)p(\theta).$$

This changes a lot. While the prior for α and Σ are still NiW conditional on θ, their unconditional prior

$$p(\alpha, \Sigma) = \int \underbrace{p(\alpha|\Sigma, \theta)p(\Sigma|\theta)}_{\text{Normal-inverse Wishart}} p(\theta)d\theta$$

can be interpreted as a mixture of all NiW distributions indexed by θ and has in general a different (unknown) distribution.

The posterior corresponding to the hierarchical prior becomes

$$p(\alpha, \Sigma, \theta|y) \propto p(y|\alpha, \Sigma)p(\alpha, \Sigma, \theta) = \underbrace{p(y|\alpha, \Sigma)}_{\text{Normal}}\underbrace{p(\alpha|\Sigma, \theta)p(\Sigma|\theta)}_{\text{Normal-inverse Wishart}}p(\theta).$$

This implies that the nice natural conjugate property is lost. In particular, the marginal posterior for α and Σ is

$$p(\alpha, \Sigma|y) \propto \int \underbrace{p(y|\alpha, \Sigma)}_{\text{Normal}}\underbrace{p(\alpha|\Sigma, \theta)p(\Sigma|\theta)}_{\text{Normal-inverse Wishart}}p(\theta)d\theta$$

$$= \underbrace{p(y|\alpha, \Sigma)}_{\text{Normal}}\underbrace{\int p(\alpha|\Sigma, \theta)p(\Sigma|\theta)p(\theta)d\theta}_{\text{NOT Normal-inverse Wishart}}.$$

Hence, we have to evaluate the posterior with simulation techniques.

Simulating the hierarchical posterior

A possible way to simulate the posterior starts from the observation that conditional on θ, we know the posterior distribution of α and Σ is Normal-inverse Wishart with parameters (20)-(23). In contrast, both the conditional and unconditional posterior for θ is unknown. This suggests to use a Metropolis-within-Gibbs approach that switches between θ on the one hand and α and Σ on the other hand.

Metropolis-within-Gibbs simulation

- Initialization: draw $\theta^{(0)}$ from some distribution, e.g., from the prior $p(\theta)$.
- Recursive replication $s = 1, \ldots, S$:
  - **(a)** Conditional on $\theta^{(s-1)}$, draw $\alpha^{(s)}$ and $\Sigma^{(s)}$ from the Normal-inverse Wishart with parameters (20)-(23).
  - **(b)** Conditional on $\alpha^{(s)}$ and $\Sigma^{(s)}$, use the MH algorithm to draw $\theta^{(s)}$ from the distribution $p(\theta|y, \alpha^{(s)}, \Sigma^{(s)})$.

The conditional posterior for θ

To run the Metropolis-within-Gibbs simulation, we have to compute $p(\theta|y, \alpha, \Sigma)$. This is possible as all we need for the MH algorithm is its kernel (since it appears in both the numerator and denominator of the formula for the acceptance probability). Using the results of the previous pages, we obtain

$$p(\theta|y, \alpha, \Sigma) = \frac{p(\alpha, \Sigma, \theta|y)}{\underbrace{p(\alpha, \Sigma|y)}_{\text{no } \theta \text{ here}}} \propto p(\alpha, \Sigma, \theta|y) \propto \underbrace{p(y|\alpha, \Sigma)}_{\text{no } \theta \text{ here}} p(\alpha, \Sigma, \theta) \propto p(\alpha, \Sigma, \theta)$$

where

$$p(\alpha, \Sigma, \theta) = \underbrace{p(\alpha, \Sigma|\theta)}_{\text{conditional NiW prior}} \times \underbrace{p(\theta)}_{\text{DSGE parameter prior}} .$$

Hence, the kernel of the conditional posterior of θ is proportional to the prior,

$$p(\theta|y, \alpha, \Sigma) \propto p(\alpha, \Sigma|\theta)p(\theta).$$

For a Metropolis step, we need to actually compute the conditional posterior pdf (given $\alpha^{(s)}$ and $\Sigma^{(s)}$) for any proposal θ^*, i.e.,

$$p(\theta^*|y, \alpha^{(s)}, \Sigma^{(s)}) \propto p(\alpha^{(s)}, \Sigma^{(s)}|\theta^*)p(\theta^*).$$

How can this be done?

To compute $p(\theta^*)$, simply substitute θ^* into the DSGE prior.

To compute $p(\alpha^{(s)}, \Sigma^{(s)}|\theta^*)$, use the NiW prior stated in (16)-(19). This is achieved by going through the following steps:

- Given a proposal θ^*, solve the DSGE model and compute the population mean and autocovariances of its observables y_t, i.e., $\Gamma^*_{yy}(\theta^*)$, $\Gamma^*_{yx}(\theta^*)$, $\Gamma^*_{xy}(\theta^*)$, and $\Gamma^*_{xx}(\theta^*)$.
- Compute the NiW parameters $\underline{V}^* = \frac{1}{\lambda T}\Gamma^{*-1}_{xx}(\theta^*)$, $\underline{A}^* = \Gamma^{*-1}_{xx}(\theta^*)\Gamma^*_{xy}(\theta^*)$, $\underline{S}^* = \lambda T[\Gamma^*_{yy}(\theta^*) - \Gamma^*_{yx}(\theta^*)\Gamma^{*-1}_{xx}(\theta^*)\Gamma^*_{xy}(\theta^*)]$, and $\underline{\nu}^* = \lambda T - pM - 1$.
- Evaluate the $NiW(\text{vec}(\underline{A}^*), \underline{V}^*, \underline{S}^*, \underline{\nu}^*)$ pdf for $\alpha^{(s)}$ and $\Sigma^{(s)}$.

4. Empirical application

Data

In the following, some of the results of Del Negro and Schorfheide (2004), henceforth DNS, are presented.

They specify a trivariate VAR(4) in quarterly data that run from 1955Q3 to 2001Q3:

- Quarterly real output growth (GDP SAAR, in Billions chained 1996 US$)
- Quarterly inflation rate (CPI-U, all items, SA)
- Average Fed funds rate during the first month of a quarter.

DSGE priors (DNS p. 656)

Name	Range	Density	Mean	SD
$\ln \gamma$	$I\!R$	Normal	0.500	0.250
$\ln \pi^*$	$I\!R$	Normal	1.000	0.500
$\ln r^*$	$I\!R^+$	Gamma	0.500	0.250
κ	$I\!R^+$	Gamma	0.300	0.150
τ	$I\!R^+$	Gamma	2.000	0.500
ψ_1	$I\!R^+$	Gamma	1.500	0.250
ψ_2	$I\!R^+$	Gamma	0.125	0.100
ρ_R	[0,1)	Beta	0.500	0.200
ρ_g	[0.1)	Beta	0.800	0.100
ρ_z	[0,1)	Beta	0.300	0.100
σ_R	$I\!R^+$	Inv. Gamma	0.251	0.139
σ_g	$I\!R^+$	Inv. Gamma	0.630	0.323
σ_z	$I\!R^+$	Inv. Gamma	0.875	0.430

NOTES: The model parameters $\ln \gamma$, $\ln \pi^*$, $\ln r^*$, σ_R, σ_g, and σ_z are scaled by 100 to convert them into percentages. The Inverse Gamma priors are of the form $p(\sigma \mid \nu, s) \propto \sigma^{-\nu-1} e^{-\nu s^2/2\sigma^2}$, where $\nu = 4$ and s equals 0.2, 0.5, and 0.7, respectively. Approximately 1.5% of the prior mass lies in the indeterminacy region of the parameter space. The prior is truncated in order to restrict it to the determinacy region of the DSGE model (SD is standard deviation).

DSGE posteriors (DNS p. 657), sample 1959Q3-1979Q2

	Prior		Posterior, $\lambda = 1$		Posterior, $\lambda = 10$	
Name	CI (Low)	CI (High)	CI (Low)	CI (High)	CI (Low)	CI (High)
$\ln \gamma$	0.101	0.922	0.473	1.021	0.616	1.045
$\ln \pi^*$	0.219	1.863	0.433	1.613	0.553	1.678
$\ln r^*$	0.132	0.880	0.113	0.463	0.126	0.384
κ	0.063	0.513	0.101	0.516	0.081	0.416
τ	1.197	2.788	1.336	2.816	1.684	3.225
ψ_1	1.121	1.910	1.011	1.559	1.009	1.512
ψ_2	0.001	0.260	0.120	0.497	0.150	0.545
ρ_R	0.157	0.812	0.530	0.756	0.550	0.747

NOTES: We report 90% confidence intervals (CI) based on the output of the Metropolis–Hastings Algorithm. The model parameters $\ln \gamma$, $\ln \pi^*$, and $\ln r^*$ are scaled by 100 to convert them into percentages.

Pseudo out-of-sample forecasting experiment

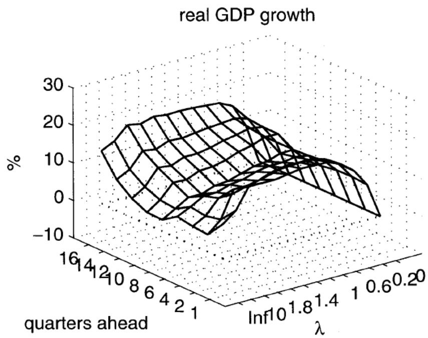
Rolling sample of 80 observations with sample endpoints 1975Q2-1997Q3 to estimate VAR and produce 1- to 16-step forecasts.

- Sample 1: 1955Q3-1975Q2, forecasts of 1975Q3-1979Q2
- ...
- Sample 90: 1977Q4-1997Q3, forecasts of 1997Q4-2001Q3

This yields 90 *h*-step forecasts that are compared to the true values. Metric: RMSE.

Note that data revisions are neglected: “pseudo” real time.

Fixed λ: percentage gain in RMSE relative to unrestricted VAR



Recursively estimated λ: percentage gain in RMSE relative to unrestricted VAR and VAR with Minnesota prior

	rGDP Growth		Inflation		Fed Funds		Multivariate	
Horizon	V-unr	V-Minn	V-unr	V-Minn	V-unr	V-Minn	V-unr	V-Minn
1	17.335	1.072	8.389	1.653	7.250	−7.593	12.842	0.943
2	16.977	6.965	7.247	1.339	5.024	−4.895	10.993	2.884
4	15.057	5.803	8.761	4.767	5.008	−1.878	9.630	3.959
6	14.116	3.452	10.460	7.240	6.648	−0.713	10.388	4.290
8	12.387	4.230	11.481	7.794	8.420	−0.204	11.023	5.187
10	14.418	7.986	12.261	8.351	8.242	−0.639	12.864	6.463
12	15.078	12.512	12.626	9.011	6.404	0.726	12.419	7.537
14	16.236	17.233	12.995	9.634	6.059	1.146	12.611	8.481
16	19.122	21.575	13.238	10.116	5.823	2.389	13.428	9.512

NOTES: The rolling sample is 1975:III–1997:III (90 periods). At each date in the sample, 80 observations are used in order to estimate the VAR. The forecasts are computed based on the values $\hat{\lambda}$ and $\hat{\imath}$ that have the highest posterior probability based on the estimation sample.