Gibbs Sampling with VAR Applications

Frank Schorfheide
University of Pennsylvania

Econ 722 – Part 1

February 7, 2019

Extension 1: Alternative Priors

- Giannone, Lenza, Primiceri (2018): "Priors for the Long-Run," *Journal of American Statistical Association*, forthcoming.
- Estimation is typically based on conditional likelihood functions that ignore the likelihood of the initial observations.
- Example:

$$y_t = c + \phi y_{t-1} + u_t = \underbrace{\phi^{t-1} y_1 + c \sum_{s=0}^{t-2} \phi^s}_{DC_t} + \underbrace{\sum_{s=0}^{t-2} \phi^s u_{t-j}}_{SC_t}$$

- Write

$$DC_t = \begin{cases} y_1 + (t-1)c & \text{if } \phi = 1 \\ \frac{c}{1-\phi} + \phi^{t-1}\left(y_1 - \frac{c}{1-\phi}\right) & \text{if} |\phi| < 1 \end{cases}$$

- Deterministic component may absorb too much low frequency variation of the time series.

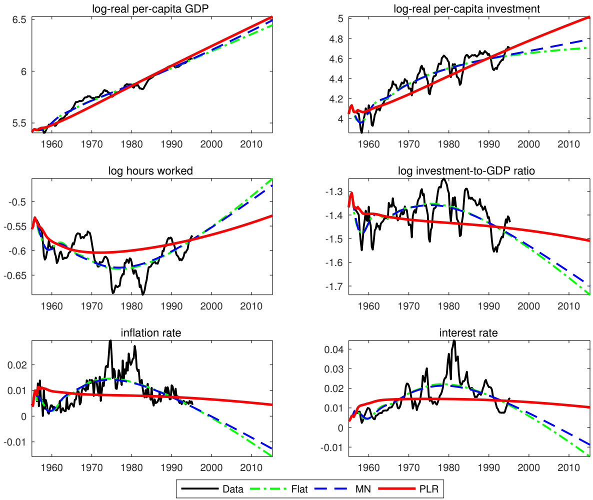


FIGURE 2.1. Deterministic component for selected variables implied by various 7-variable VARs. Flat: BVAR with a flat prior; MN: BVAR with the Minnesota prior; PLR: BVAR with the prior for the long run.

Extension 1: Alternative Priors – The Basic Idea

- Write VAR in VECM form:

$$\Delta y_t = \Pi_0 + \Pi_* y_{t-1} + \sum_{j=1}^{p-1} \Pi_i \Delta y_{t-j} + u_t$$

  where $\Pi_* = \alpha\beta'$.

- Reasonable prior for columns of α will depend on the rows of β':
  - if i'th row of β' corresponds to a linear combination that is stationary, then it makes sense to choose a prior for i'th column of α with mass away form zero.
  - if i'th row of β' corresponds to a linear combination that is non-stationary, then it makes sense to choose a prior for i'th column of α with mass away form zero.

- See paper for details on how to implement this.

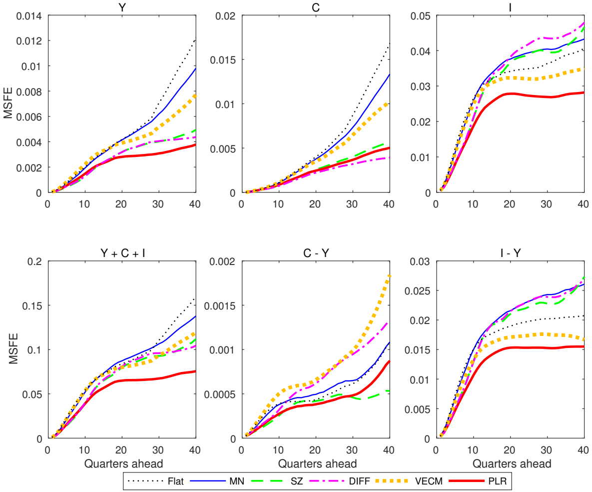


FIGURE 5.1. Mean squared forecast errors in models with three variables. Flat: BVAR with a flat prior; MN: BVAR with the Minnesota prior; SZ: BVAR with the Minnesota and sum-of-coefficient priors; DIFF: VAR with variables in first differences; VECM: vector error-correction model that imposes the existence of a common stochastic trend for Y, C and I, without any additional prior information; PLR: BVAR with the Minnesota prior and the prior for the long run.

Extension 2: Sparse versus Dense Models

- Giannone, Lenza, Primiceri (2018): "Economic Prediction With Big Data: The Illusion of Sparsity," *Manuscript*, FRB New York, ECB, and Northwestern University.

- Sparse models: only a few predictors are relevant.

- Dense models: many predictors are relevant but only have small individual effects.

- Model:

$$y_t = x_t'\phi + z_t'\beta + u_t.$$

Here x_t's are included in all specifications (low dimensional), z_t's are optional (high dimensional).

- Prior – part 1:

$$p(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \phi \propto c.$$

Extension 2: Sparse versus Dense Models

- Prior – part 2: "spike and slab"

$$\beta_i|(\sigma^2,\gamma^2,q) \sim \begin{cases} N(0,\sigma^2\gamma^2) & \text{with prob. } q \\ 0 & \text{with prob. } 1-q \end{cases}$$

- For $q = 1$ we obtain our "standard" prior ("Ridge Regression")
- Rewrite prior as

$$\beta_i|(\sigma^2,\gamma^2,\nu_i) \sim N(0,\sigma^2\gamma^2,\nu_i), \quad \nu_i \sim \text{Bernoulli}(q).$$

- By changing the mixing distribution, we can generate a wide variety of priors, including a Bayesian version of LASSO.

Extension 2: Sparse versus Dense Models

- In problems of this form it is often good to standardize and orthogonalize the regressors x_t prior to the estimation.

- To specify a prior on the hyperparameters (q, γ^2) they suggest to define

$$R^2(\gamma^2, q) = \frac{qk\gamma^2\bar{\sigma}_z^2}{qk\gamma^2\bar{\sigma}_z^2 + 1}$$

  where k is the number of regressors z and $\bar{\sigma}_z^2$ is the average sample variance of the z_j's.

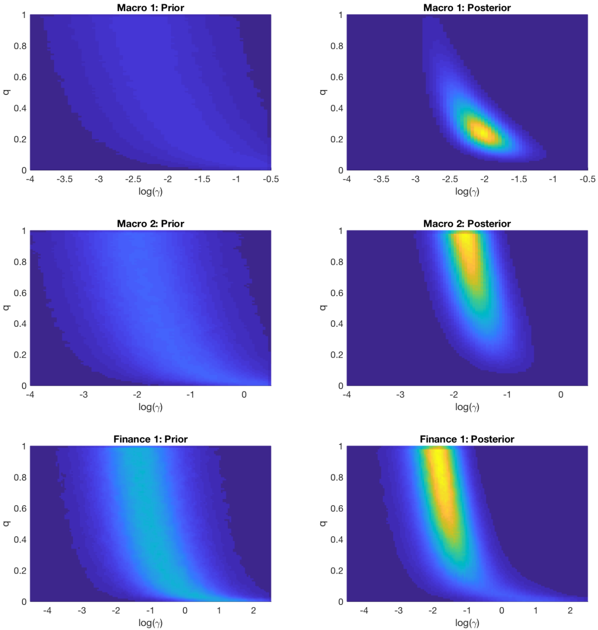
- The prior takes the form

$$q \sim \text{Beta}(a, b), \quad R^2 \sim \text{Beta}(A, B).$$

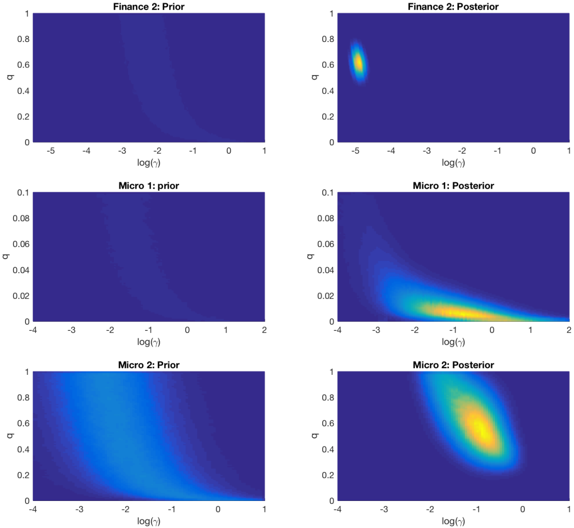
- The paper works out the posterior.

TABLE 1. Description of the datasets.

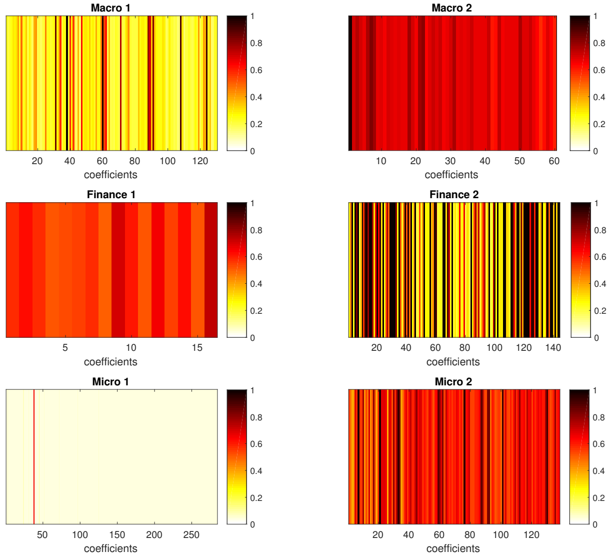
	Dependent variable	Possible predictors	Sample
Macro 1	Monthly growth rate of US industrial production	130 lagged macroeconomic indicators	659 monthly time-series observations, from February 1960 to December 2014
Macro 2	Average growth rate of GDP over the sample 1960-1985	60 socio-economic, institutional and geographical characteristics, measured at pre-60s value	90 cross-sectional country observations
Finance 1	US equity premium (S&P 500)	16 lagged financial and macroeconomic indicators	58 annual time-series observations, from 1948 to 2015
Finance 2	Stock returns of US firms	144 dummies classifying stock as very low, low, high or very high in terms of 36 lagged characteristics	1400k panel observations for an average of 2250 stocks over a span of 624 months, from July 1963 to June 2015
Micro 1	Per-capita crime (murder) rates	Effective abortion rate and 284 controls including possible covariate of crime and their transformations	576 panel observations for 48 US states over a span of 144 months, from January 1986 to December 1997
Micro 2	Number of pro-plaintiff eminent domain decisions in a specific circuit and in a specific year	Characteristics of judicial panels capturing aspects related to gender, race, religion, political affiliation, education and professional history of the judges, together with some interactions among the latter, for a total of 138 regressors	312 panel circuit/year observations, from 1975 to 2008



Macro 1: Prior
Macro 1: Posterior
Macro 2: Prior
Macro 2: Posterior
Finance 1: Prior
Finance 1: Posterior
q
log(γ)



Finance 2: Prior
Finance 2: Posterior
Micro 1: prior
Micro 1: Posterior
Micro 2: Prior
Micro 2: Posterior
q
log(γ)



Macro 1
coefficients
Macro 2
coefficients
Finance 1
coefficients
Finance 2
coefficients
Micro 1
coefficients
Micro 2
coefficients

Extension 3: Time-Varying Coefficients

- Most Common Versions of TVP Models
  - Parameters follow AR law of motion.
  - Parameters follow regime switching process
- Even though this is a lecture about VARs, we focus mostly on univariate models to discuss the key ideas.

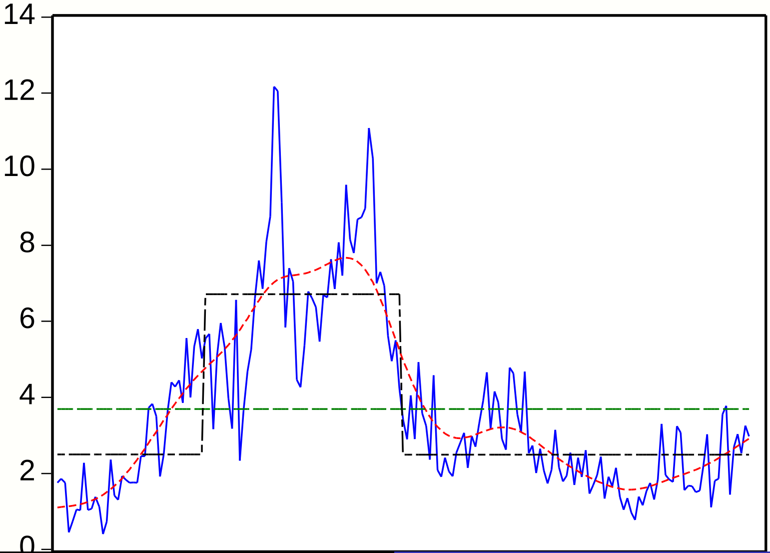
From Constant Coefficients to Time-Varying Coefficients

- We previously considered constant-coefficient autoregressive models.
- For instance, a simple model for inflation could be

$$\pi_t = \pi^* + \tilde{\pi}_t, \quad \tilde{\pi}_t = \rho\tilde{\pi}_{t-1} + \sigma_\epsilon \epsilon_t$$

- where
  - π^* is steady state or target inflation;
  - $\tilde{\pi}_t$ captures fluctuations around the target

Is a Constant π^* Plausible?



Let's Introduce Time-Variation into Inflation Target

- Inflation evolves according to:

$$\pi_t = \pi_t^* + \tilde{\pi}_t$$

where

$$\tilde{\pi}_t = \rho\tilde{\pi}_{t-1} + \sigma_\epsilon \epsilon_t, \quad \pi_t^* = \pi_{t-1}^* + \sigma_\eta \eta_t.$$

- This looks like a state-space model:

$$\begin{aligned} y_t &= \begin{bmatrix} 1 & 1 \end{bmatrix} s_t \\ s_t &= \begin{bmatrix} \pi_t^* \\ \tilde{\pi}_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix} s_{t-1} + \begin{bmatrix} \sigma_\eta & 0 \\ 0 & \sigma_\epsilon \end{bmatrix} \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix}. \end{aligned}$$

Parameter Estimation

- Suppose that all the non-redundant parameters of the state space model are collected in the vector θ.

- Construct a Gibbs sampler that iterates over parameters and states $S_{1:T}$:

$$\begin{aligned} p(S_{1:T}|Y_{1:T},\theta) &\propto p(S_{1:T}|\theta)p(Y_{1:T}|S_{1:T},\theta) \\ p(\theta|Y_{1:T},S_{1:T}) &\propto p(\theta)p(S_{1:T}|\theta)p(Y_{1:T}|S_{1:T},\theta) \end{aligned}$$

Estimation: Gibbs-Sampling Algorithm

- Generate draws from $p(\theta, S_{1:T}|Y_{1:T})$ using Carter and Kohn (1994)'s approach.
- Gibbs-sampling algorithm iterates over the conditional posteriors of θ and $S_{1:T}$.
- Recall the linear Gaussian state space representation

$$
\begin{aligned}
y_t &= A + Bs_t + u_t, \quad u_t \sim N(0, H) \\
s_t &= \Phi s_{t-1} + e_t, \quad e_t \sim N(0, Q)
\end{aligned}
$$

  with $\theta = (A, B, H, \Phi, Q)$
- For $i = 1, ..., n_{sim}$
  (a) Draw $\theta^{(i)}$ from $p\left(\theta \mid Y_{1:T}, S_{1:T}^{(i-1)}\right)$
    - Conditional on $S_{1:T}^{(i-1)}$, drawing θ is a standard linear regression
    - (Measurement) $y_t = A + Bs_t + u_t$
    - (Transition) $s_t = \Phi s_{t-1} + e_t$

  (b) Draw $S_{1:T}^{(i)}$ from $p\left(S_{1:T} \mid Y_{1:T}, \theta^{(i)}\right)$
    - Kalman / simulation smoother

Drawing the States: Carter and Kohn (1994)

- How can we draw $S_{1:T}$ from the conditional posterior $p(S_{1:T}|Y_{1:T},\theta)$?
- It turns out that we can draw the states sequentially, starting from $s_T|Y_{1:T}$, which is obtained in the T'th iteration of the filter.
- We then continue with

$$p(s_t|S_{t+1:T},Y_{1:T}) \propto p(s_t,S_{t+1:T},Y_{1:T})$$

Drawing the States: Carter and Kohn (1994)

Consider the following factorization

$$
\begin{aligned}
& p(s_t, S_{t+1:T}, Y_{1:T}) \\
&= \int p(S_{1:T}, Y_{1:T}) dS_{1:t-1} \\
&= \int p(S_{1:t}, Y_{1:t}) \cdot \big[p(s_{t+1}|s_t) p(y_{t+1}|s_{t+1}) \big] \cdot \big[p(s_{t+2}|s_{t+1}) p(y_{t+2}|s_{t+2}) \big] \\
&\quad \ldots \big[p(s_T|s_{T-1}) p(y_T|s_T) \big] dS_{1:t-1} \\
&= p(s_t, Y_{1:t}) \cdot \big[p(s_{t+1}|s_t) p(y_{t+1}|s_{t+1}) \big] \cdot \text{terms without } s_t.
\end{aligned}
$$

We deduce

$$
\begin{aligned}
p(s_t|S_{t+1:T}, Y_{1:T}) &\propto p(s_t, Y_{1:t}) p(s_{t+1}|s_t) \\
&\propto p(s_t, s_{t+1}, Y_{1:t}) \\
&= p(s_t|s_{t+1}, Y_{1:t})
\end{aligned}
$$

Drawing the States: Carter and Kohn (1994)

- We now can write

$$\begin{aligned} p(S_{1:T}|Y_{1:T}) &= p(s_T|Y_{1:T}) \prod_{t=1}^{T-1} p(s_t|s_{t+1}, {\color{red}Y_{1:T}}) \\ &= p(s_T|Y_{1:T}) \prod_{t=1}^{T-1} p(s_t|s_{t+1}, {\color{blue}Y_{1:t}}) \end{aligned}$$

- where $p(s_t|s_{t+1}, Y_{1:t}) \propto p(s_t|Y_{1:t})p(s_{t+1}|s_t)$.
- Draw $s_t^{(i)} \sim p(s_t \mid s_{t+1}^{(i)}, Y_{1:T})$ for $t = T, ..., 1$.
  (a) Run the Kalman filter to get $\{\hat{s}_{t|t}, P_{t|t}\}_{t=1}^{T}$ where $s_t|Y_{1:t} \sim N(\hat{s}_{t|t}, P_{t|t})$;
  (b) Draw $s_T^{(i)} \sim N(\hat{s}_{T|T}, P_{T|T})$;
  (c) For $t = T-1, ..., 1$,

$$s_t^{(i)} \sim N\left(\hat{s}_{t|t+1},\ P_{t|t+1}\right)$$

  where

$$\begin{aligned} \hat{s}_{t|t+1} &= \hat{s}_{t|t} + P_{t|t}\Phi' P_{t+1|t}^{-1}(s_{t+1}^{(i)} - \Phi\hat{s}_{t|t}) \\ P_{t|t+1} &= P_{t|t} - P_{t|t}\Phi' P_{t+1|t}^{-1}\Phi P_{t|t} \end{aligned}$$

Generalizations

- Multivariate instead of univariate framework;
- Time-varying slope coefficients;
- Time-varying shock variances.
- Some references:
    - Cogley and Sargent (2002, NBER Macro Annual)
    - Cogley and Sargent (2005, RED)
    - Primiceri (2005, REStud)

A TVP VAR with Stochastic Volatility

- Reduced-form VAR:

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_c + u_t.$$

- We defined $x_t = [y'_{t-1}, \ldots, y'_{t-p}, 1]'$ and $\Phi = [\Phi_1, \ldots, \Phi_p, \Phi_c]'$.
- Let $X_t = I_n \otimes x_t$ and $\phi = vec(\Phi)$. Write the VAR as

$$y_t = X'_t \phi_t + u_t. \tag{1}$$

- Parameters evolve according to the random walk process:

$$\phi_t = \phi_{t-1} + \nu_t, \quad \nu_t \sim iidN(0, Q). \tag{2}$$

- Q is diagonal; ν_t and u_t are uncorrelated.
- VAR innovations:

$$u_t \sim N(0, \Sigma_t), \quad \Sigma_t = B^{-1} H_t (B^{-1})'. \tag{3}$$

B is a lower-triangular matrix with ones on the diagonal; H_t is a diagonal matrix with elements $h^2_{i,t}$:

$$\ln h_{i,t} = \ln h_{i,t-1} + \eta_{i,t}, \quad \eta_{i,t} \sim iidN(0, \sigma^2_i). \tag{4}$$

A Gibbs Sampler

- $\phi_{1,T}^{(s)}$ conditional on $(B^{(s-1)}, H_{1,T}^{(s-1)}, Q^{(s-1)}, \sigma_1^{(s-1)} \dots \sigma_n^{(s-1)}, Y)$.
  (1) and (2) provide a state-space representation for y_t. Thus, $\phi_{1,T}$ can be sampled using the algorithm developed by Carter and Kohn.

- $B^{(s)}$ conditional on $(\phi_{1,T}^{(s)}, H_{1,T}^{(s-1)}, Q^{(s-1)}, \sigma_1^{(s-1)} \dots \sigma_n^{(s-1)}, Y)$.
  Conditional on the VAR parameters ϕ_t, the innovations to equation (1) are known. According to (3) Bu_t is normally distributed with variance H_t, or:

$$Bu_t = H_t^{\frac{1}{2}} \epsilon_t, \tag{5}$$

  where ϵ_t is a vector of standard normals.

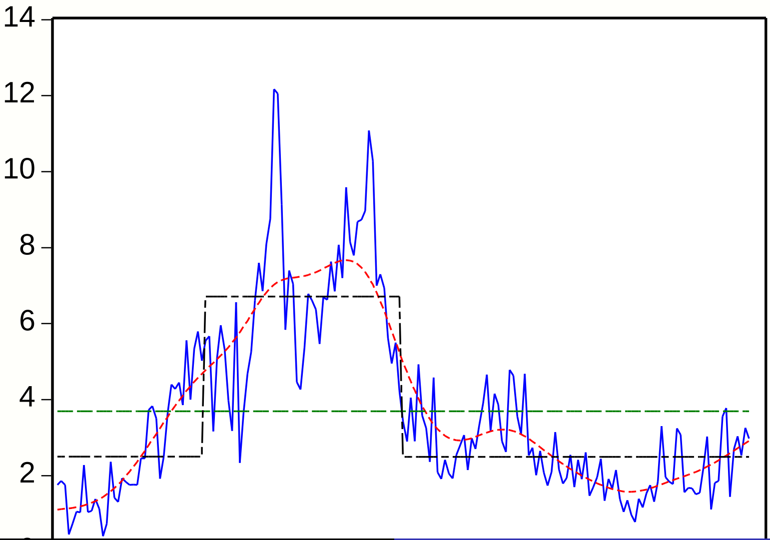
A Gibbs Sampler

- $H_{1,T}^{(s)}$ conditional on $(\phi_{1,T}^{(s)}, B^{(s)}, Q^{(s-1)}, \sigma_1^{(s-1)} \dots \sigma_n^{(s-1)}, Y)$.
  Conditional on ϕ_t and B we can write the i'th equation of (5) as
  $z_{i,t} = B_{(i.)} u_t \sim N(0, h_{i,t}^2)$. (see literature on stochastic volatility models...)

- $Q^{(s)}$ conditional on $(\phi_{1,T}^{(s)}, B^{(s)}, H_{1,T}^{(s)}, \sigma_1^{(s-1)} \dots \sigma_n^{(s-1)}, Y)$.
  Use appropriate Inverted Wishart distribution derived from (2).

- $\sigma_1^{(s)} \dots \sigma_n^{s}$ conditional on $(\phi_{1,T}^{(s)}, B^{(s)}, H_{1,T}^{(s)}, Q^{(s)}, Y)$.
  Use appropriate Inverted Gamma distributions derived from (4).

Back To Our Inflation Series



A Simple Markov Switching Model

- Suppose

$$y_t = \alpha_0 + \alpha_1 s_t + \epsilon_t, \quad \epsilon_t \sim iidN(0,1),$$

- where $s_t \in \{0,1\}$ evolves according to a Markov-switching process with transition probabilities q_{00} (from 0 to 0) and q_{11} (from 1 to 1).
- Priors:

$$\begin{aligned} \alpha &\sim N(\underline{\alpha}, \underline{V}_\alpha) \\ p(q_{00}, q_{11}) &\propto q_{00}^{\alpha_0 - 1}(1 - q_{00})^{\beta_0 - 1} q_{11}^{\alpha_1 - 1}(1 - q_{11})^{\beta_1 - 1} \end{aligned}$$

- Joint:

$$p(Y_{1:T}, S_{1:T}, \alpha, q) = p(Y_{1:T}|S_{1:T}, \alpha)p(S_{1:T}|q)p(q)p(\alpha)$$

Gibbs Sampler

- $S_{1:T}|(q, \alpha, Y_{1:T})$: use a smoother tailored toward MS models.
- $\alpha|(q, S_{1:T}, Y_{1:T})$: regression on split sample

$$\begin{aligned} y_t &= \alpha_0 + \epsilon_t \quad \text{if } s_t = 0 \\ y_t &= \alpha_0 + \alpha_1 + \epsilon_t \quad \text{if } s_t = 1 \end{aligned}$$

- $q|(\alpha, S_{1:T}, Y_{1:T})$: count state transitions n_{00}, n_{01}, ...; (ignoring initialization of Markov process) posterior for q has Beta distribution

$$p(q|\cdot) \propto q_{00}^{n_{00}+\alpha_0-1}(1-q_{00})^{n_{01}+\beta_0-1} q_{11}^{n_{11}+\alpha_1-1}(1-q_{11})^{n_{10}+\beta_1-1}.$$

Generalization: Markov-Switching VARs

- We add regime-switching to the coefficients of the reduced form VAR:

$$y_t' = x_t'\Phi(s_t) + u_t', \quad u_t \sim iidN(0, \Sigma(s_t)) \tag{6}$$

- Here s_t is a discrete M-state Markov process with time-invariant transition probabilities

$$q_{lm} = \mathbb{P}[s_t = m \mid s_{t-1} = l], \quad l, m \in \{1, \ldots, M\}.$$

- Suppose that $M = 2$ and all elements of $\Phi(s_t)$ and $\Sigma(s_t)$ switch simultaneously, without any restrictions. Denote the values of the VAR parameter matrices in state $s_t = l$ by $\Phi(l)$ and $\Sigma(l)$, $l = 1, 2$, respectively. Specify MNIW priors for $(\Phi(l), \Sigma(l))$ and Beta priors for q_{11} and q_{22}

Gibbs Sampler

For $i = 1, \dots, n_{sim}$:

1. Draw $(\Phi^{(i)}(l), \Sigma^{(i)}(l))$ conditional on $(s_{1,T}^{(i-1)}, q_{11}^{(i-1)}, q_{22}^{(i-1)}, Y)$. Let $\mathcal{T}_l$ be a set that contains the time periods when $s_t = l$, $l = 1, 2$. Under a conjugate prior, the posterior of $\Phi(l)$ and $\Sigma(l)$ is MNIW, obtained from the regression $y_t' = x_t'\Phi(l) + u_t$, $u_t \sim N(0, \Sigma(l))$, $t \in \mathcal{T}_l$.
2. Draw $s_{1,T}^{(i)}$ conditional on $(\Phi^{(i)}(l), \Sigma^{(i)}(l), q_{11}^{(i-1)}, q_{22}^{(i-1)}, Y)$ using a smoother.
3. Draw $q_{11}^{(i)}$ and $q_{22}^{(i)}$ conditional on $(\Phi^{(i)}(s), \Sigma^{(i)}(s), s_{1,T}^{(i)}, Y)$ from the appropriate Beta distribution. $\square$

Extension 4

In macroeconomic applications, vector autoregressions (VARs) are typically estimated either exclusively based on

- quarterly observations
  $\Longrightarrow$ large set of macroeconomic series is available
- monthly information
  $\Longrightarrow$ VAR is able to track the economy more closely in real time

State-Space Representation of MF-VAR

- State-Transition Equation
  - Economy evolves at monthly frequency according to the following VAR(p) dynamics:

$$x_t = \Phi_1 x_{t-1} + \ldots + \Phi_p x_{t-p} + \Phi_c + u_t, \quad u_t \sim iidN(0, \Sigma) \tag{7}$$

  - Partition: $x_t' = [x_{m,t}', \; x_{q,t}']$
- Measurement Equation
  - Actual observations are denoted by y_t and subscript indicates the observation frequency

$$\begin{aligned} y_{m,t} &= x_{m,t} \\ y_{q,t} &= \frac{1}{3}(x_{q,t} + x_{q,t-1} + x_{q,t-2}) \quad \text{if observed in period } t \end{aligned} \tag{8}$$

Bayesian Inference

- Minnesota Prior (Sims and Zha, 1998 IER) for (Φ, Σ) indexed by hyperparameter vector λ
- Gibbs Sampler along the lines of Carter and Kohn (1994)
  - Draw latent states (Kalman Smoother)

$$X_{1:T}^{(j)} \mid (\Phi^{(j-1)}, \Sigma^{(j-1)}, Y_{1:T})$$

  - Draw VAR parameters (Direct Sampling from MNIW)

$$(\Phi^{(j)}, \Sigma^{(j)}) \mid (X_{1:T}^{(j)}, Y_{1:T})$$

  for $j = 1, ..., n_{sim}$

- Implementation issues:
  - Reduce dimension of state-space by removing observed states $x_{m,t}$
  - Adjustments for *ragged edge* at the end of estimation sample