Variational Bayes Inference for Large Vector Autoregressions

Reza Hajargasht **Tomasz Wozniak**

University of Melbourne

Melbourne, July 2016

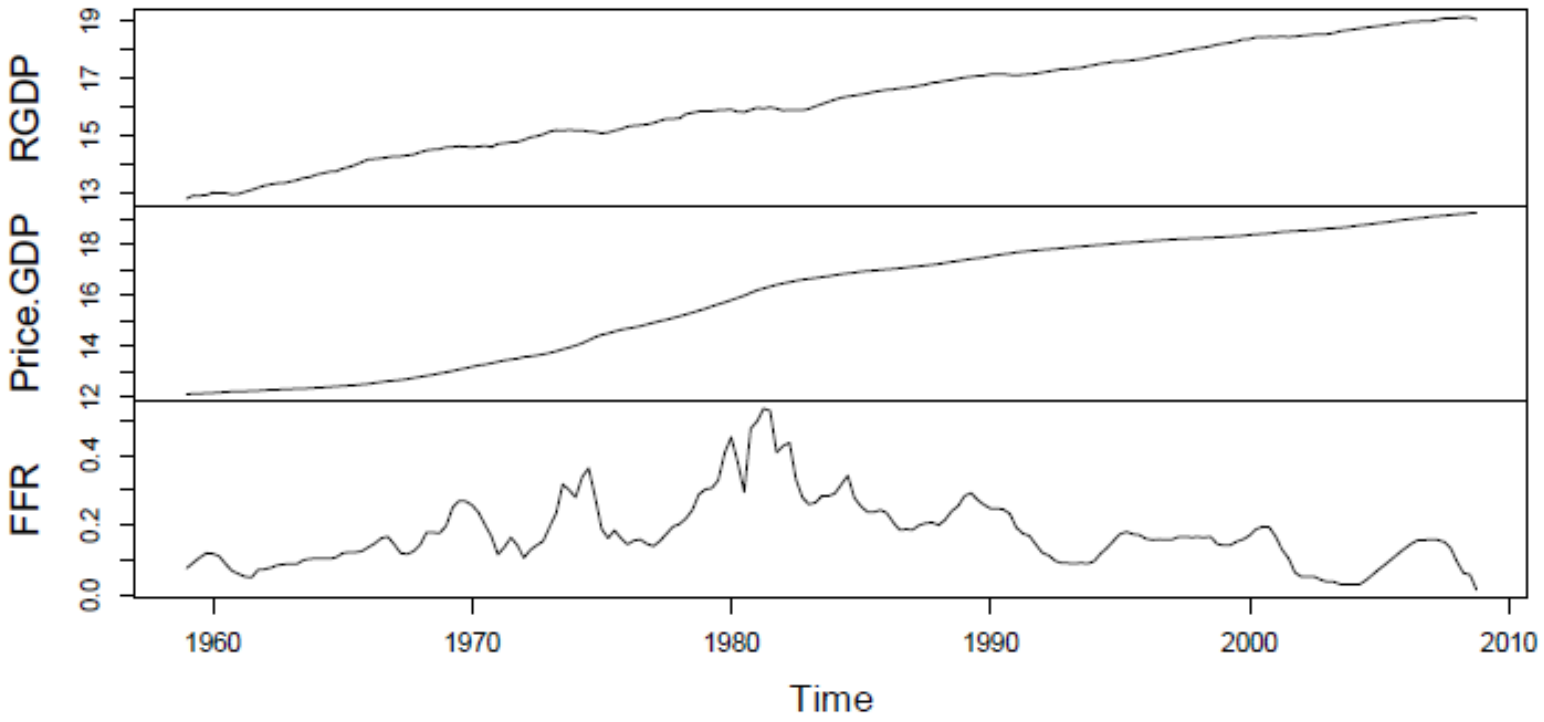
VAR Models

- *Understanding the dynamics of macro variables, their relationship and forecasting.*
- *VAR models popularized by Chris SIMS in 80s is one the most important tools*
- A typical small VAR model

$$\begin{cases} gdp_t = \alpha_1 + \alpha_{11} gdp_{t-1} + \ldots + \alpha_{14} gdp_{t-4} + \beta_{11} p_{t-1} + \ldots + \beta_{14} p_{t-4} + \gamma_{11} r_{t-1} + \ldots + \gamma_{14} r_{t-4} + \varepsilon_{1,t} \\ P_t = \alpha_2 + \alpha_{21} gdp_{t-1} + \ldots + \alpha_{24} gdp_{t-4} + \beta_{21} p_{t-1} + \ldots + \beta_{24} p_{t-4} + \gamma_{22} r_{t-1} + \ldots + \gamma_{24} r_{t-4} + \varepsilon_{2,t} \\ r_t = \alpha_3 + \alpha_{31} gdp_{t-1} + \ldots + \alpha_{34} gdp_{t-4} + \beta_{31} p_{t-1} + \ldots + \beta_{34} p_{t-4} + \gamma_{31} r_{t-1} + \ldots + \gamma_{34} r_{t-2} + \varepsilon_{3,t} \end{cases}$$

$$\operatorname{cov}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_{3\times 3}$$

UK Macroeconomics Series



We consider 3 cases with 3, 7 and 22 variables.
Source: Giannone, Lenza and Primiceri (2012) and Stock and Watson (2008)

Bayesian VAR

large number of parameters

leading to imprecise estimation and forecasting

- Bayesians address this by specifying Shrinkage Priors

Our Contribution

1. Derivation of algorithms:

   Variational Bayes estimation

   Forecasting

   for various VAR models [Conjugate and Independent Priors]

2. Optimal hyper-parameters

3. Providing an accurate method for computation of Marginal Likelihood using VB

Variational Bayes

- VB is an alternative to MCMC developed in Machine Learning finding its way into statistics and Econometrics

- The basic idea is to approximate the posterior $p(\boldsymbol{\theta} \mid \mathbf{y})$ with another tractable density function $q(\boldsymbol{\theta})$

Variational Bayes

- .

Problem setting.
Approximate the posterior distribution, $p(\theta|\mathbf{y})$, with another distribution, $q(\theta)$, that is tractable.

Characterisation.

	Variational Bayes	simulations
Calculations:	approximate	exact
Convergence:	deterministic	stochastic
Methods:	mean field theory calculus of variations	MCMC

Variational Bayes

- Minimize K-L Distance

$$Min_q KL(q \,\|\, p) = \int q(\boldsymbol{\theta}) \log \left\{ \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta} \,|\, \mathbf{y})} \right\} d\boldsymbol{\theta}$$

subject to $q(\boldsymbol{\theta}) = \prod_{k=1}^{M} q_k(\boldsymbol{\theta}_k)$

where $\{\boldsymbol{\theta}_1, \ldots., \boldsymbol{\theta}_K\}$ is a partition of $\boldsymbol{\theta}$

Variational Bayes

- This is a calculus of variation problem. It can be shown:
- Initialize with $q_2^*(\boldsymbol{\theta}_2),\ldots..q_M^*(\boldsymbol{\theta}_M)$ and cycle

$$\begin{cases} q_1^*(\boldsymbol{\theta}_1) \leftarrow \dfrac{\exp(E_{\boldsymbol{\theta}_{-1}} \ln p(\mathbf{y},\boldsymbol{\theta}))}{\int \exp(E_{\boldsymbol{\theta}_{-1}} \ln p(\mathbf{y},\boldsymbol{\theta}))d\boldsymbol{\theta}_1} \\ \vdots \\ q_M^*(\boldsymbol{\theta}_M) \leftarrow \dfrac{\exp(E_{\boldsymbol{\theta}_{-M}} \ln p(\mathbf{y},\boldsymbol{\theta}))}{\int \exp(E_{\boldsymbol{\theta}_{-M}} \ln p(\mathbf{y},\boldsymbol{\theta}))d\boldsymbol{\theta}_M} \end{cases}$$

Lower Bound for Marginal Likelihood

- It has been shown that ML or MDD

$$\ln ML = E_q \ln p(\mathbf{y}, \boldsymbol{\theta}) - E_q \ln q(\boldsymbol{\theta}) + KL(q, p)$$

- Define

$$\ln \underline{ML} = E_q \ln p(\mathbf{y}, \boldsymbol{\theta}) - E_q \ln q(\boldsymbol{\theta})$$

- Since KL is always positive, $\underline{ML}$ provides a lower bound to marginal likelihood

VAR Model

M = number of endogenous variables

T = number of time periods

d= number of lags

$$\mathbf{y}_t = \boldsymbol{\gamma}_0 + \sum_{j=1}^{d} \boldsymbol{\Gamma}_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t$$

$$\operatorname{cov}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_{M \times M}$$

Matrix Form

$$\mathbf{Y}_{T \times M} = \mathbf{X}_{T \times K} \boldsymbol{\Gamma}_{K \times M} + \mathbf{E}_{T \times M}$$

Conjugate Priors

$$P(\mathbf{\Gamma}, \mathbf{\Sigma}^{-1}) = P(\mathbf{\Gamma} \mid \mathbf{\Sigma}^{-1}) P(\mathbf{\Sigma}^{-1})$$

where

$$P(\mathbf{\Gamma} \mid \mathbf{\Sigma}^{-1}) = MN(\underline{\mathbf{\Gamma}}, \mathbf{\Sigma}, \underline{\mathbf{V}})$$

$$P(\mathbf{\Sigma}^{-1}) = W(\underline{\mathbf{S}}^{-1}, \underline{\upsilon})$$

1) Normal-Wishart
2) Minnesota (proposed by Litterman (1980) and been used in many studies such as BGR 2010)
3) Conjugate SSVS (Koop 2012)
4) Hierarchical (Giannone, Lenza and Primiceri, 2012)

Minnesota

- For $\underline{\gamma} = vec(\underline{\Gamma})$

$$\underline{\gamma}_i = 0 \; except \; for \; first \; own \; lag \; which \; is \; 0.9 \; or \; 1$$

- $\underline{V}$ is a diagonal matrix with (e.g. size 13)

$$\begin{cases} \text{Off-diagonal elements of } \underline{V} \text{ set to zeroes} \\ \underline{V}_1 = \lambda_0 \text{ for intercept} \\ \underline{V}_j = \dfrac{\lambda_1^2}{r^2 \sigma_{jj}^2} \text{ corresponding to lag r} \\ \sigma_{ii} = s_i^2 \end{cases}$$

Posterior without Factorization

- Exact Posterior

$$\begin{cases} \boldsymbol{\Gamma} \mid \boldsymbol{\Sigma}^{-1}, \mathbf{Y} \sim MN\{\bar{\boldsymbol{\Gamma}}, \boldsymbol{\Sigma}\ , \bar{\mathbf{V}}\} \\ \boldsymbol{\Sigma}^{-1} \mid \mathbf{Y} \sim W(\bar{\mathbf{S}}^{-1}, \bar{\upsilon}) \end{cases}$$

where

$$\bar{\mathbf{V}} = [\underline{\mathbf{V}}^{-1} + \mathbf{X}'\mathbf{X}]^{-1} \quad \bar{\boldsymbol{\Gamma}} = \bar{\mathbf{V}}[\underline{\mathbf{V}}^{-1}\underline{\boldsymbol{\Gamma}} + \mathbf{X}'\mathbf{Y}]$$

$$\bar{\mathbf{S}} = (\mathbf{Y} - \mathbf{X}\bar{\boldsymbol{\Gamma}})'(\mathbf{Y} - \mathbf{X}\bar{\boldsymbol{\Gamma}}) + \underline{\mathbf{S}} + (\bar{\boldsymbol{\Gamma}}' - \underline{\boldsymbol{\Gamma}}')\underline{\mathbf{V}}^{-1}(\bar{\boldsymbol{\Gamma}}' - \underline{\boldsymbol{\Gamma}}) \quad \bar{\upsilon} = T + \underline{\upsilon}$$

- Posterior Moments

$$E(\boldsymbol{\Gamma}) = \bar{\boldsymbol{\Gamma}}$$

$$Var(\boldsymbol{\gamma}) = \frac{1}{T + \underline{\upsilon} - p - 1}\bar{\mathbf{S}} \otimes \bar{\mathbf{V}}$$

$$E(\boldsymbol{\Sigma}^{-1}) = (T + \underline{\upsilon})\bar{\mathbf{S}}^{-1}$$

$$Var(\sigma_{ij}^{+}) = \bar{\upsilon}\left(s_{ij}^{+2} + s_{ii}^{+}s_{jj}^{+}\right)$$

Posterior with Factorization

- Factorization $q(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}^{-1}) = q(\boldsymbol{\Gamma}) q(\boldsymbol{\Sigma}^{-1})$
- VB Posterior

$$\begin{cases} \boldsymbol{\Gamma} \sim MN\{\bar{\boldsymbol{\Gamma}}, \overline{\boldsymbol{\Sigma}^{-1}}^{-1}, \bar{\mathbf{V}}\} \\ \boldsymbol{\Sigma}^{-1} \sim W(\bar{\mathbf{S}}_{VB}^{-1}, \bar{\upsilon}_{VB}) \end{cases}$$

where

$$\bar{\mathbf{V}} = [\underline{\mathbf{V}}^{-1} + \mathbf{X}'\mathbf{X}]^{-1} \quad \bar{\boldsymbol{\Gamma}} = \bar{\mathbf{V}}[\underline{\mathbf{V}}^{-1}\underline{\boldsymbol{\Gamma}} + \mathbf{X}'\mathbf{Y}]$$

$$\bar{\mathbf{S}}_{VB} = \frac{\bar{\upsilon}_{VB}}{\bar{\upsilon}} \bar{\mathbf{S}} \quad \bar{\upsilon}_{VB} = T + p + \underline{\upsilon}$$

- Posterior Means

$$E(\boldsymbol{\Gamma}_{VB}) = \bar{\boldsymbol{\Gamma}}$$

$$Var(\boldsymbol{\gamma}_{VB}) = \frac{1}{T + \underline{\upsilon}} \bar{\mathbf{S}} \otimes \bar{\mathbf{V}}$$

$$E(\boldsymbol{\Sigma}^{-1}) = (T + \underline{\upsilon})\bar{\mathbf{S}}^{-1}$$

$$Var(\sigma_{ij}^{+}) = \frac{\bar{\upsilon}_G^2}{\bar{\upsilon}_{VB}} \left(s_{ij}^{+2} + s_{ii}^{+} s_{jj}^{+} \right)$$

VB vs Full Bayes

- Posterior Means for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}^{-1}$ are the same
- Posterior SE for factorized model are smaller

$$\frac{Var(\boldsymbol{\gamma}_{VB})}{Var(\boldsymbol{\gamma}_{G})} = \frac{T + \underline{\upsilon} - p - 1}{T + \underline{\upsilon}} \qquad \frac{Var(\sigma^{+}_{ijVB})}{Var(\sigma^{+}_{ijG})} = \frac{T + \underline{\upsilon}}{T + \underline{\upsilon} + p}$$

- KL Distance

$$KL = -\frac{Mp}{2}\ln 2e + \frac{M}{2}\ln\left(\frac{\bar{\upsilon}_{VB}^{\bar{\upsilon}_{VB}}}{\bar{\upsilon}^{\bar{\upsilon}}}\right) - \ln\left(\frac{\Gamma_M(\bar{\upsilon}_{VB}/2)}{\Gamma_M(\bar{\upsilon}/2)}\right)$$

- KL increases with increase in *M* and *d* and decreases with increase in *T* and $\underline{\upsilon}$. It does not depend on data and other hyper-parameters!

VB vs Full Bayes

$$\lim_{T\to\infty} KL \to \frac{1}{2}\sum_{j=1}^{m} \ln(1) = 0$$

$$\lim_{\underline{\upsilon}\to\infty} KL \to \frac{1}{2}\sum_{j=1}^{m} \ln(1) = 0$$

$$\lim_{p\to\infty} KL \to \frac{1}{2}\sum_{j=1}^{m} \ln(0) = -\infty$$

$$\lim_{p\to 0} KL \to \frac{1}{2}\sum_{j=1}^{m} \ln(1) = 0$$

Another way of writing VAR

Another way of writing VAR

$$\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_T \end{pmatrix} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_T \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_M \end{pmatrix} + \boldsymbol{\varepsilon} \qquad \text{where} \quad Z_t = \begin{pmatrix} z_{1t}' & 0 & \cdots & 0 \\ 0 & z_{2t}' & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & z_{Mt}' \end{pmatrix}$$

$$z_t = (y_{1,t}, \ldots, y_{m,t}, y_{1,t-1}, \ldots, y_{m,t-1}, \ldots\ldots, y_{1,t-d}, \ldots, y_{m,t-d})$$

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}_T \otimes \boldsymbol{\Sigma}_M)$$

Independent Priors

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\beta})\, p(\boldsymbol{\Sigma}^{-1})$$

where

$$P(\boldsymbol{\beta}) = N(\underline{\boldsymbol{\beta}}, \underline{\mathbf{V}})$$

$$P(\boldsymbol{\Sigma}^{-1}) = W(\underline{\mathbf{S}}^{-1}, \underline{\upsilon})$$

1- Normal-Wishart

2- Minnesota [Kadiyala and Karlsson, 1997]

3- Lasso Type Priors [Korobilis 2012, Gefang 2012]

4- SSVS [George, Sun and Ni 2008, Korobilis 2013]

Minnesota

- For $\underline{\boldsymbol{\beta}}$

$$\underline{\beta}_i = 0 \; except \; for \; first \; own \; lag \; which \; is \; 0.9 \; or \; 1$$

- $\underline{\mathbf{V}}$ is a diagonal matrix of size p e.g. 39

Off-diagonal elements of $\underline{\mathbf{V}}$ set to zeroes

$$\underline{V}_1 = \lambda_0^2 \sigma_{ii}^2 \quad \text{for intercepts}$$

$$\underline{V}_{i,jj} = \frac{\lambda_1^2 \sigma_{ii}^2}{r^2} \quad \text{for own lag r}$$

$$\underline{V}_{i,jj} = \frac{\lambda_2^2 \sigma_{ii}^2}{r^2 \sigma_{jj}^2} \quad \text{for other variables lag}$$

Exact Posterior

Gibbs Sampler

$$\begin{cases} \boldsymbol{\beta} \mid \boldsymbol{\Sigma}^{-1}, \mathbf{Y} \sim N(\bar{\boldsymbol{\beta}}, \bar{\mathbf{V}}) \\ \boldsymbol{\Sigma}^{-1} \mid \boldsymbol{\beta}, \mathbf{Y} \sim W(\bar{\mathbf{S}}^{-1}, \bar{\upsilon}) \end{cases}$$

where

$$\bar{\mathbf{V}} = [\underline{\mathbf{V}}^{-1} + \sum_{t=1}^{T} \mathbf{Z}_t' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_t]^{-1} \quad \bar{\boldsymbol{\beta}} = \bar{\mathbf{V}}[\underline{\mathbf{V}}^{-1} \underline{\boldsymbol{\beta}} + \sum_{t=1}^{T} \mathbf{Z}_t' \boldsymbol{\Sigma}^{-1} \mathbf{y}_t]$$

$$\bar{\mathbf{S}} = \underline{\mathbf{S}} + \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{Z}_t \boldsymbol{\beta})(\mathbf{y}_t - \mathbf{Z}_t \boldsymbol{\beta})' \quad \bar{\upsilon} = T + \underline{\upsilon}$$

VB Posterior

- VB Posterior $\begin{cases} \boldsymbol{\beta} \sim N(\bar{\boldsymbol{\beta}}, \bar{\mathbf{V}}) \\ \boldsymbol{\Sigma}^{-1} \sim W(\bar{\mathbf{S}}^{-1}, \bar{\upsilon}) \end{cases}$
- Initialize with $\overline{\boldsymbol{\Sigma}^{-1}}$ and cycle

$$\bar{\mathbf{V}} \leftarrow [\underline{\mathbf{V}}^{-1} + \sum_{t=1}^{T} \mathbf{Z}_t' \overline{\boldsymbol{\Sigma}^{-1}} \mathbf{Z}_t]^{-1}$$

$$\bar{\boldsymbol{\beta}} \leftarrow \bar{\mathbf{V}}[\underline{\mathbf{V}}^{-1}\underline{\boldsymbol{\beta}} + \sum_{t=1}^{T} \mathbf{Z}_t' \overline{\boldsymbol{\Sigma}^{-1}} \mathbf{y}_t]$$

$$\bar{\mathbf{S}} \leftarrow \underline{\mathbf{S}} + \sum_{t=1}^{T} \{(\mathbf{y}_t - \mathbf{Z}_t\bar{\boldsymbol{\beta}})(\mathbf{y}_t - \mathbf{Z}_t\bar{\boldsymbol{\beta}})' + \mathbf{Z}_t\bar{\mathbf{V}}\mathbf{Z}_t'\}$$

$$\bar{\upsilon} \leftarrow T + \underline{\upsilon}$$

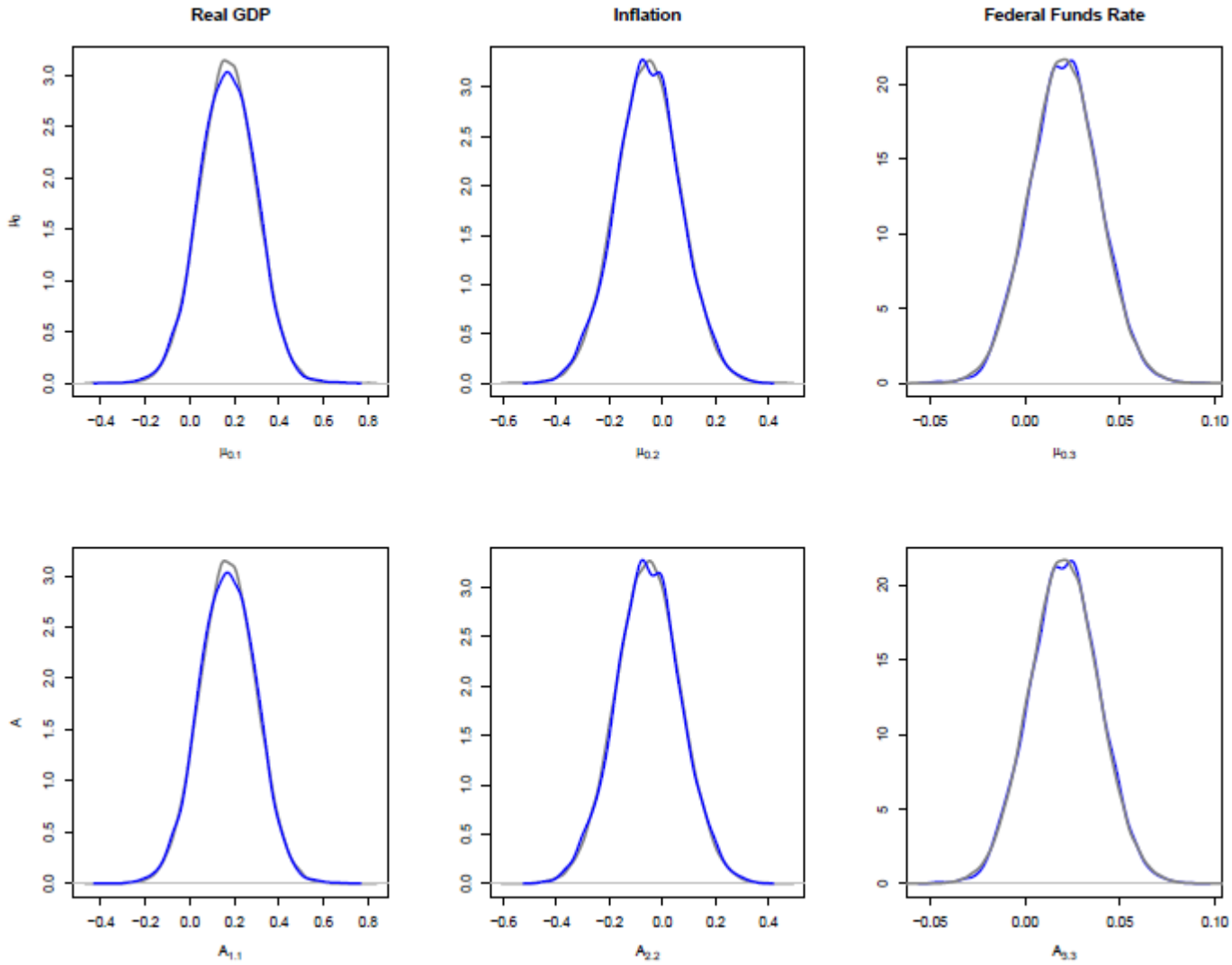
$$\overline{\boldsymbol{\Sigma}^{-1}} \leftarrow \bar{\upsilon}\bar{\mathbf{S}}^{-1}$$

- Lower bound for Log of Marginal Likelihood

$$\ln ML = \frac{p}{2} - \frac{MT}{2}\ln\pi + \ln\Gamma_M(\bar{\upsilon}/2) - \ln\Gamma_M(\underline{\upsilon}/2) + \frac{1}{2}\left(\ln\left|\bar{\mathbf{V}}\right| - \ln|\underline{\mathbf{V}}|\right)$$

$$+ \frac{1}{2}\left(\bar{\upsilon}\ln\left|\bar{\mathbf{S}}^{-1}\right| - \underline{\upsilon}\ln\left|\underline{\mathbf{S}}^{-1}\right|\right) - \frac{1}{2}tr\left\{\underline{\mathbf{V}}^{-1}[(\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})(\bar{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}})' + \bar{\mathbf{V}}]\right\}$$

Comparison


Real GDP
Inflation
Federal Funds Rate
μ
A
$\mu_{0.1}$
$\mu_{0.2}$
$\mu_{0.3}$
$A_{1.1}$
$A_{2.2}$
$A_{3.3}$

Optimal Hyper-parameter Selection

- Let $\boldsymbol{\lambda}$ be the vector of hyper-parameters and define

$$\ln \underline{ML}(\boldsymbol{\lambda}) = \ln \underline{ML}(\phi \mid \boldsymbol{\lambda})$$

- Then optimal hyper-parameters are obtained

$$\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda}} \ln \underline{ML}(\boldsymbol{\lambda})$$

Other available techniques.

- Maximisation of ln *MDD* - available only for natural-conjugate prior distribution
- Point forecasting performance measures (Bańbura, Giannone, Reichlin, 2010, JAE)
- Hierarchical prior structures (see e.g. Giannone, Lenza, Primiceri, 2013, NBER)

Point Forecasting

- We are interested in

$$E(y_{i,\tau+h} \mid Data_{\tau})$$

- Even with VB posterior this does not seem to have an analytic form [with more than one step ahead forecast].
- We can use simulation though: simulate

$$\boldsymbol{\Sigma}^{-1(j)} \sim W(\overline{S}^{-1}, \overline{\upsilon}) \quad \boldsymbol{\beta}^{(j)} \sim N(\overline{\boldsymbol{\beta}}, V_{\beta}) \quad \mathbf{u}_{\tau+h}^{(j)} \sim N(0, \boldsymbol{\Sigma})$$

and calculate recursively for $h=1,\ldots,H$

$$\tilde{\mathbf{y}}_{\tau+h}^{(j)} = \mathbf{a}_0^{(j)} + \sum_{i=1}^{h-1} \mathbf{A}_i^{(j)} \tilde{\mathbf{y}}_{\tau+h-i}^{(j)} + \sum_{i=h}^{p_1} \mathbf{A}_i^{(j)} \mathbf{y}_{\tau+h-i}^{(j)} + \mathbf{u}_{\tau+h}^{(j)}$$

Point Forecasting

- But note this does not increase running time very significantly because:

  1- The draws are independent, therefore a small number of draws e.g. 1000 should be enough

  2- This doesn't have to be done in a loop

- Forecasting Metrics

$$MSFE = \frac{\sum_{\tau=\tau_0}^{T-h} [y_{i,\tau+h}^{0} - E(y_{i,\tau+h} \mid Data_{\tau})]^2}{T-h-\tau_0+1}$$

Overall Forecasting

- Predictive likelihood

$$p[y^0_{i,\tau+h} \mid Data_\tau] = \int p[y^0_{i,\tau+h} \mid Data_\tau, \boldsymbol{\theta}]\, \pi(\boldsymbol{\theta} \mid Data_\tau) d\boldsymbol{\theta}$$

where $\pi(\boldsymbol{\theta} \mid Data_\tau)$ is the posterior.

- Even with VB posterior this doesn't seem analytically tractable but we can draw from the VB posterior and calculate the integral
- A quadratic approximation

$$\ln p[y^0_{i,\tau+h} \mid Data_\tau] \sim -\frac{1}{2}\left\{\ln 2\pi + \ln V_{i,\tau+h|\tau} + (y^0_{i,\tau+h} - \bar{y}_{i,\tau+h|\tau})^2 \Big/ V_{i,\tau+h|\tau}\right\}$$

- Forecasting Metrics

$$\sum_{t=\tau_0}^{T-h} \ln p[y^0_{i,\tau+h} \mid Data_\tau]$$

Forecasting Performance

- Models estimated with VB methods forecast as accurately as those estimated with the exact methods
  - for the natural-conjugate prior
    benchmark: analytical solutions
  - for the independent prior
    benchmark: Gibbs sampling output
- Hyper-parameters optimisation improves the forecasting performance
  - mixed evidence for point forecasts at some horizons
- Hyper-parameters optimization using InMDD_{VBLB} gives exactly the same results as using InMDD
- Nearly no improvements for independent priors

Accurate Computation of ML

- VB can be used for accurate computation of Marginal Likelihood under a reciprocal importance sampling framework e.g.

$$ML = M\left[\sum_{m=1}^{M} \frac{\mathbf{q}(\boldsymbol{\theta}^m \mid \mathbf{y})}{p(\mathbf{y} \mid \boldsymbol{\theta}^m)\, p(\boldsymbol{\theta}^m)}\right]^{-1}$$

- Where $\boldsymbol{\theta}^m$s are draws from MCMC and $q(\boldsymbol{\theta} \mid \mathbf{y})$ is the posterior from VB.
- This method has good properties if the candidate density has narrower tail which is usually the case with VB.

Marginal Likelihood

•

		VB_LB	GD_Gew	GD_VB	Bridge_VB	Chib	IS_VB	VB_UB
VAR_S 3 variable	Mean	635.25	635.27	635.42	635.42	635.42	635.42	635.62
	SD	NA	0.0462	0.0077	0.0056	0.0079	0.0107	0.0081
VAR_M 7 variable	Mean	961.63	958.23	962.77	962.77	962.71	962.76	964.09
	SD	NA	0.2940	0.0438	0.0164	0.0162	0.0528	0.0292

Conclusion

1. Derivation of fast algorithms:

   VB Estimation

   VB Forecasting

   for Bayesian VARs with conjugate, independent and hierarchical priors

2. Selection of optimal values for hyper-parameters of the model.

3- Accurate Computation of Marginal Likelihood using VB Posterior in par with Chib method

4- Can be extended to more elaborate models