Bayesian Methods for Empirical Macroeconomics

Hedibert Freitas Lopes

The University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, IL 60637
http://faculty.chicagobooth.edu/hedibert.lopes
hlopes@ChicagoBooth.edu

Contents

1 Overview of Bayesian Econometrics **7**
1.1 Example i. Sequential learning . . . 7
1.2 Example ii. Normal-normal . . . 9
1.3 Turning the Bayesian crank . . . 11
1.3.1 Prior predictive . . . 11
1.3.2 Posterior . . . 11
1.3.3 Posterior predictive . . . 12
1.3.4 Sequential Bayes . . . 12
1.3.5 Model probability . . . 13
1.3.6 Posterior odds . . . 14
1.3.7 Bayes factor . . . 14
1.3.8 Marginal likelihood . . . 15
1.4 Example iii. Multiple linear regression . . . 15
1.5 Real data exercise . . . 18
1.6 Example iv. SV model . . . 20

2 Bayesian Computation **23**
2.1 A bit of history . . . 23
2.2 MC integration . . . 25
2.3 MC via IS . . . 26
2.4 Rejection method . . . 28
2.5 SIR method . . . 31
2.6 MH algorithms . . . 39
2.7 Simulated annealing . . . 46
2.8 Gibbs sampler . . . 48
2.9 SUR model . . . 51
2.10 BVAR . . . 56
2.11 References . . . 60

3 Dynamic Models **63**
3.1 1st order DLM . . . 63
3.1.1 n-variate normal . . . 64
3.1.2 The Kalman filter . . . 64
3.1.3 The Kalman smoother . . . 65

3.1.4 Example . . . 66
3.1.5 Integrating out states x^n . . . 67
3.1.6 MCMC scheme . . . 68
3.1.7 Lessons . . . 69
3.2 Dynamic linear models (DLMs) . . . 69
3.2.1 Linear growth model . . . 70
3.2.2 Bivariate VAR(1) with TVP . . . 70
3.2.3 Sequential inference . . . 71
3.2.4 Smoothing . . . 72
3.2.5 The FFBS . . . 72
3.2.6 Individual sampling . . . 73
3.2.7 Joint sampling . . . 74
3.2.8 Example . . . 75
3.3 SV model . . . 76

4 Bayesian VAR **87**
4.1 VAR . . . 87
4.1.1 Model set up . . . 87
4.1.2 Forecasting . . . 89
4.1.3 Stationarity . . . 89
4.1.4 VMA . . . 89
4.1.5 VD . . . 90
4.1.6 IRF . . . 90
4.1.7 Examples . . . 91
4.1.8 Minnesota prior . . . 95
4.1.9 Other priors . . . 96
4.1.10 Example . . . 97
4.1.11 Structural VAR . . . 98
4.2 VAR-GARCH . . . 101
4.3 VAR-SV . . . 102
4.4 TVP-VAR-SV . . . 102
4.5 Dimensionality . . . 104

5 BVAR Extensions **107**
5.1 Variable selection . . . 107
5.1.1 ON/OFF switch . . . 108
5.1.2 SSVS . . . 110
5.2 Factor analysis . . . 111
5.2.1 Basic model . . . 111
5.2.2 PBFA . . . 118
5.2.3 FSV . . . 120
5.2.4 SDFM . . . 123
5.3 Forecasting with large q . . . 126
5.4 Forecasting with many predictors . . . 127
5.5 Large BVAR & BFAVAR . . . 128
5.6 Large TVP-BVAR & BFAVAR . . . 130

5.7 References 131

Chapter 1

Overview of Bayesian Econometrics

1.1 Example i. Sequential learning

Example i. Sequential learning

- John claims some discomfort and goes to the doctor.
- The doctor believes John may have the disease A.
- $\theta = 1$: John has disease A; $\theta = 0$: he does not.
- The doctor claims, based on his expertise (H), that

$$P(\theta = 1|H) = 0.70$$

- Examination X is related to θ as follows

$$\begin{cases} P(X = 1|\theta = 0) = 0.40, & \text{positive test given no disease} \\ P(X = 1|\theta = 1) = 0.95, & \text{positive test given disease} \end{cases}$$

Observe $X = 1$

Exam's result: $X = 1$

$$\begin{aligned} P(\theta = 1|X = 1) &\propto P(X = 1|\theta = 1)P(\theta = 1) \\ &\propto (0.95)(0.70) = 0.665 \\ P(\theta = 0|X = 1) &\propto P(X = 1|\theta = 0)P(\theta = 0) \\ &\propto (0.40)(0.30) = 0.120 \end{aligned}$$

Consequently

$$\begin{aligned} P(\theta = 0|X = 1) &= 0.120/0.785 = 0.1528662 \text{ and} \\ P(\theta = 1|X = 1) &= 0.665/0.785 = 0.8471338 \end{aligned}$$

The information $X = 1$ increases, for the doctor, the probability that John has the disease A from 70% to 84.71%.

Posterior predictive

John undertakes the test Y, which relates to θ as follows[1]

$$P(Y = 1|\theta = 1) = 0.99 \quad \text{and} \quad P(Y = 1|\theta = 0) = 0.04$$

Then, the predictive of $Y = 0$ given $X = 1$ is given by

$$\begin{aligned} P(Y = 0|X = 1) &= P(Y = 0|X = 1, \theta = 0)P(\theta = 0|X = 1) \\ &+ P(Y = 0|X = 1, \theta = 1)P(\theta = 1|X = 1) \\ &= P(Y = 0|\theta = 0)P(\theta = 0|X = 1) \\ &+ P(Y = 0|\theta = 1)P(\theta = 1|X = 1) \\ &= (0.96)(0.1528662) + (0.01)(0.8471338) \\ &= 15.52\% \end{aligned}$$

Key condition: X and Y are conditionally independent given θ.

Model criticism

Suppose the observed result was $Y = 0$. This is a reasonably unexpected result as the doctor only gave it roughly 15% chance.

He should at least consider rethinking the model based on this result. In particular, he might want to ask himself

1. Did 0.7 adequately reflect his $P(\theta = 1|H)$?
2. Is test X really so unreliable?
3. Is the sample distribution of X correct?
4. Is the test Y so powerful?
5. Have the tests been carried out properly?

[1] Recall that $P(X = 1|\theta = 1) = 0.95$ and $P(X = 1|\theta = 0) = 0.40$.

Observe $Y = 0$

Let $H_2 = \{X = 1, Y = 0\}$. Then, Bayes theorem leads to

$$\begin{aligned} P(\theta = 1|H_2) &\propto P(Y = 0|\theta = 1)P(\theta = 1|X = 1) \\ &\propto (0.01)(0.8471338) = 0.008471338 \\ P(\theta = 0|H_2) &\propto P(Y = 0|\theta = 0)P(\theta = 0|X = 1) \\ &\propto (0.96)(0.1528662) = 0.1467516 \end{aligned}$$

Therefore,

$$P(\theta = 1|X = 1, Y = 0) = \frac{P(Y = 0, \theta = 1|X = 1)}{P(Y = 0|X = 1)} = 0.0545753$$

$$P(\theta = 1|H_i) = \begin{cases} 0.7000 & , H_0\text{: before X and Y} \\ 0.8446 & , H_1\text{: after X=1 and before Y} \\ 0.0546 & , H_2\text{: after X=1 and Y=0} \end{cases}$$

1.2 Example ii. Normal-normal

Example ii. Normal-normal

Consider a simple measurement error model

$$X = \theta + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2)$$

where

$$\theta \sim N(\theta_0, \tau_0^2).$$

The quantities $(\sigma^2, \theta_0, \tau_0^2)$ are known.

The posterior distribution of θ (after $X = x$ is observed) is

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

More precisely,

$$\begin{aligned} p(\theta|x) &\propto \exp\{-0.5(\theta^2 - 2\theta x)/\sigma^2\}\exp\{-0.5(\theta^2 - 2\theta\theta_0)/\tau_0^2\} \\ &\times \exp\{-0.5(\theta^2(1/\sigma^2 + 1/\tau_0^2) + 2\theta(x/\sigma^2 + \theta_0/\tau_0^2)\} \\ &= \exp\{-0.5(\theta^2/\tau_1^2 + 2\theta\tau_1^2(x/\sigma^2 + \theta_0/\tau_0^2)/\tau_1^2\} \\ &= \exp\{-0.5(\theta^2 + 2\theta\theta_1)/\tau_1^2\}. \end{aligned}$$

Therefore, $\theta|x$ is normally distributed with

$$E(\theta|x) = \tau_1^2(x/\sigma^2 + \theta_0/\tau_0^2)$$

and

$$V(\theta|x) = \left(1/\sigma^2 + 1/\tau_0^2\right)^{-1}.$$

Notice that

$$E(\theta|x) = \omega\theta_0 + (1-\omega)x$$

where

$$\omega = \frac{\sigma^2}{\sigma^2 + \tau_0^2}$$

measures the relative information contained in the prior distribution with respect to the total information (prior plus likelihood).

Illustration

Prior A: Physicist A (large experience): $\theta \sim N(900, (20)^2)$

Prior B: Physicist B (not so experienced): $\theta \sim N(800, (80)^2)$.

Model: $(X|\theta) \sim N(\theta, (40)^2)$.

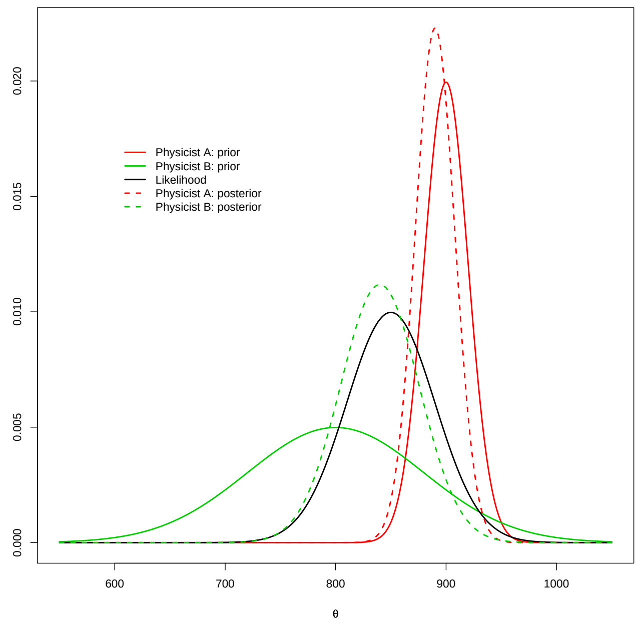
Observation: $X = 850$

$$\begin{aligned} (\theta|X = 850, H_A) &\sim N(890, (17.9)^2) \\ (\theta|X = 850, H_B) &\sim N(840, (35.7)^2) \end{aligned}$$

Information (precision)

Physicist A: from 0.002500 to 0.003120 (an increase of 25%)

Physicist B: from 0.000156 to 0.000781 (an increase of 400%)



1.3 Turning the Bayesian crank

1.3.1 Prior predictive

Turning the Bayesian crank

We usually decompose

$$p(\theta, x|H)$$

into

$$p(\theta|H) \quad \text{and} \quad p(x|\theta, H)$$

The prior predictive distribution

$$p(x|H) = \int_{\Theta} p(x|\theta, H)p(\theta|H)\, d\theta = E_\theta[p(x|\theta, H)]$$

if of key importance in Bayesian model assessment.

1.3.2 Posterior

Posterior distribution

The posterior distribution of θ is obtained, after x is observed, by Bayes' Theorem:

$$
\begin{aligned}
p(\theta|x,H) &= \frac{p(\theta,x|H)}{p(x|H)} \\
&= \frac{p(x|\theta,H)p(\theta|H)}{p(x|H)} \\
&\propto p(x|\theta,H)p(\theta|H).
\end{aligned}
$$

1.3.3 Posterior predictive

Posterior predictive distribution

Let y be a new set of observations conditionally independent of x given θ, ie.

$$p(x,y|\theta) = p(x|\theta,H)p(y|\theta,H).$$

Then,

$$
\begin{aligned}
p(y|x,H) &= \int_\Theta p(y,\theta|x,H)d\theta \\
&= \int_\Theta p(y|\theta,x,H)p(\theta|x,H)d\theta \\
&= \int_\Theta p(y|\theta,H)p(\theta|x,H)d\theta \\
&= E_{\theta|x}\left[p(y|\theta,H)\right]
\end{aligned}
$$

In general, but not always (time series, for example) x and y are independent given θ.

It might be more useful to concentrate on prediction rather than on estimation since the former is *verifiable*.

x and y can be (and usually are) observed; θ can not!

1.3.4 Sequential Bayes

Sequential Bayes theorem

Experimental result: $x_1 \sim p_1(x_1|\theta)$

$$p(\theta|x_1) \propto l_1(\theta;x_1)p(\theta)$$

Experimental result: $x_2 \sim p_2(x_2|\theta)$

$$\begin{aligned} p(\theta|x_2, x_1) &\propto l_2(\theta; x_2)p(\theta|x_1) \\ &\propto l_2(\theta; x_2)l_1(\theta; x_1)p(\theta) \end{aligned}$$

Experimental results: $x_i \sim p_i(x_i|\theta)$, for $i = 3, \ldots, n$

$$\begin{aligned} p(\theta|x_n, \ldots, x_1) &\propto l_n(\theta; x_n)p(\theta|x_{n-1}, \ldots, x_1) \\ &\propto \left[\prod_{i=1}^{n} l_i(\theta; x_i)\right] p(\theta) \end{aligned}$$

1.3.5 Model probability

Model probability

Suppose that the competing models can be enumerated and are represented by the set

$$M = \{M_1, M_2, \ldots\}$$

and that the *true model* is in M (Bernardo and Smith, 1994).

The posterior model probability of model M_j is given by

$$Pr(M_j|y) = \frac{f(y|M_j)Pr(M_j)}{f(y)}$$

Ingredients

Prior predictive density of model M_j

$$f(y|M_j) = \int f(y|\theta_j, M_j)p(\theta_j|M_j)d\theta_j$$

Prior model probability of model M_j

$$Pr(M_j)$$

Overall prior predictive

$$f(y) = \sum_{M_j \in M} f(y|M_j)Pr(M_j)$$

$\log_{10} B_{jk}$	B_{jk}	Evidence against k
0.0 to 0.5	1.0 to 3.2	Not worth more than a bare mention
0.5 to 1.0	3.2 to 10	Substantial
1.0 to 2.0	10 to 100	Strong
> 2	> 100	Decisive

$2\log_e B_{jk}$	B_{jk}	Evidence against k
0.0 to 2.0	1.0 to 3.0	Not worth more than a bare mention
2.0 to 6.0	3.0 to 20	Substantial
6.0 to 10.0	20 to 150	Strong
> 10	> 150	Decisive

1.3.6 Posterior odds

Posterior odds

The posterior odds of model M_j relative to M_k is given by

$$\underbrace{\frac{Pr(M_j|y)}{Pr(M_k|y)}}_{\text{posterior odds}} = \underbrace{\frac{Pr(M_j)}{Pr(M_k)}}_{\text{prior odds}} \times \underbrace{\frac{f(y|M_j)}{f(y|M_k)}}_{\text{Bayes factor}} .$$

The Bayes factor can be viewed as the weighted likelihood ratio of M_j to M_k.

The main difficulty is the computation of the marginal likelihood or normalizing constant $f(y|M_j)$.

Therefore, the posterior model probability for model j can be obtained from

$$\frac{1}{Pr(M_j|y)} = \sum_{M_k \in M} B_{kj} \frac{Pr(M_k)}{Pr(M_j)}.$$

1.3.7 Bayes factor

Bayes factor

Jeffreys (1961) recommends the use of the following rule of thumb to decide between models j and k:

Kass and Raftery (1995) argue that "it can be useful to consider twice the natural logarithm of the Bayes factor, which is on the same scale as the familiar deviance and likelihood ratio test statistics". Their slight modification is:

1.3.8 Marginal likelihood

Marginal likelihood

A basic ingredient for model assessment is given by the predictive density

$$f(y|M) = \int f(y|\theta, M)p(\theta|M)d\theta \ ,$$

which is the normalizing constant of the posterior distribution.

The predictive density can now be viewed as the likelihood of model M.

It is sometimes referred to as predictive likelihood, because it is obtained after marginalization of model parameters.

The predictive density can be written as the expectation of the likelihood with respect to the prior:

$$f(y) = E_p[f(y|\theta)].$$

1.4 Example iii. Multiple linear regression

Example iii. Multiple linear regression

The standard Bayesian approach to multiple linear regression is

$$y_i = x_i'\beta + \epsilon_i$$

for $i = 1, \dots, n$, x_i a q-dimensional vector of regressors and residuals ϵ_i iid $N(0, \sigma^2)$.

In matrix notation,

$$(y|X, \beta, \sigma^2) \sim N(X\beta, \sigma^2 I_n)$$

where $y = (y_1, \dots, y_n)$, $X = (x_1, \dots, x_n)'$ is the $(n \times q)$, design matrix and $q = p + 1$.

Example iii. Maximum likelihood estimation

It is well known that

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
\hat{\sigma}^2 &= \frac{S_e}{n-q} = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n-q}
\end{aligned}$$

are the OLS estimates of β and σ^2, respectively.

The conditional and unconditional sampling distributions of $\hat{\beta}$ are

$$\begin{aligned}(\hat{\beta}|\sigma^2, y, X) &\sim N(\beta, \sigma^2(X'X)^{-1})\\ (\hat{\beta}|y, X) &\sim t_{n-q}(\beta, S_e(X'X)^{-1})\end{aligned}$$

respectively, with

$$(\hat{\sigma}^2|\sigma^2) \sim IG\left((n-q)/2, ((n-q)\sigma^2/2\right).$$

Example iii. Conjugate prior

The prior distribution of (β, σ^2) is $NIG(b_0, B_0, n_0, S_0)$, i.e.

$$\begin{aligned}\beta|\sigma^2 &\sim N(b_0, \sigma^2 B_0)\\ \sigma^2 &\sim IG(n_0/2, n_0 S_0/2)\end{aligned}$$

for known hyperparameters b_0, B_0, n_0 and S_0.

For clarification, when $\sigma^2 \sim IG(a, b)$, if follows that

$$p(\sigma^2) \propto (\sigma^2)^{-(a+1)} \exp\left\{-\frac{b}{\sigma^2}\right\}$$

with

$$E(\sigma^2) = \frac{b}{a-1} \quad \text{and} \quad V(\sigma^2) = \frac{b^2}{(a-1)^2(a-2)}$$

Example iii. Conditionals

It is easy to show that

$$(\beta|\sigma^2, y, X) \sim N(b_1, \sigma^2 B_1)$$

where

$$\begin{aligned}B_1^{-1} &= B_0^{-1} + X'X\\ B_1^{-1} b_1 &= B_0^{-1} b_0 + X'y.\end{aligned}$$

It is also easy to show that

$$(\sigma^2|\beta, y, X) \sim IG(n_1/2, n_1 S_{11}(\beta)/2)$$

where

$$\begin{aligned}n_1 &= n_0 + n\\ n_1 S_{11}(\beta) &= n_0 S_0 + (y - X\beta)'(y - X\beta).\end{aligned}$$

Example iii. Marginals

It can be shown that

$$(\sigma^2|y,X) \sim IG(n_1/2, n_1S_1/2)$$

where

$$n_1S_1 = n_0S_0 + (y - Xb_1)'y + (b_0 - b_1)'B_0^{-1}b_0.$$

Consequently,

$$(\beta|y,X) \sim t_{n_1}(b_1, S_1B_1).$$

Example iii. MLE versus Bayes

Distributions of the estimators $\hat{\beta}$ and $\hat{\sigma}^2$

$$\begin{aligned}(\hat{\sigma}^2|\sigma^2,y,X) &\sim IG\big((n-q)/2, ((n-q)\sigma^2/2\big)\\ (\hat{\beta}|\beta,y,X) &\sim t_{n-q}(\beta, S_e(X'X)^{-1}).\end{aligned}$$

Marginal posterior distributions of β and σ^2

$$\begin{aligned}(\sigma^2|y,X) &\sim IG(n_1/2, n_1S_1/2)\\ (\beta|y,X) &\sim t_{n_1}(b_1, S_1B_1).\end{aligned}$$

Vague prior: When $B_0^{-1} = 0$, $n_0 = -q$ and $S_0 = 0$

$$\begin{aligned}b_1 &= \hat{\beta}\\ B_1 &= (X'X)^{-1}\\ n_1 &= n - q\\ n_1S_1 &= (y - X\hat{\beta})'y = (y - X\hat{\beta})'(y - X\hat{\beta}) = (n-q)\hat{\sigma}^2\\ S_1B_1 &= \hat{\sigma}^2(X'X)^{-1}.\end{aligned}$$

Example iii. Predictive

The predictive density can be obtained by

$$p(y|X) = \int p(y|X,\beta,\sigma^2)p(\beta|\sigma^2)p(\sigma^2)d\beta d\sigma^2$$

or (via Bayes' theorem) by

$$p(y|X) = \frac{p(y|X,\beta,\sigma^2)p(\beta|\sigma^2)p(\sigma^2)}{p(\beta|\sigma^2,y,X)p(\sigma^2|y,X)}$$

which is valid for all (β, σ^2).

Closed form solution for the multiple normal linear regression:

$$(y|X) \sim t_{n_0}(Xb_0, S_0(I_n + XB_0X')).$$

1.5 Real data exercise

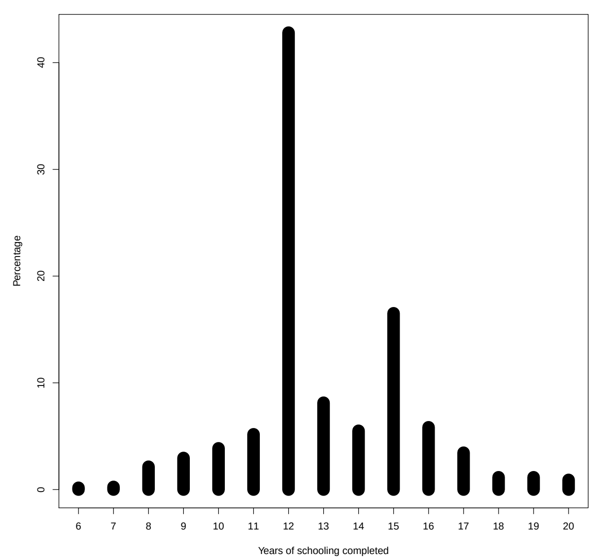
Real data exercise
To better understand the differential role of the prior in estimation and model comparison, consider the following simple linear regression application, illustrated using a sample of $n = 1,217$ observations from the National Longitudinal Survey of Youth (NLSY):

- $\mathcal{M}_0 : y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2).$
- $\mathcal{M}_1 : y_i = \beta_0 \qquad\qquad + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2).$

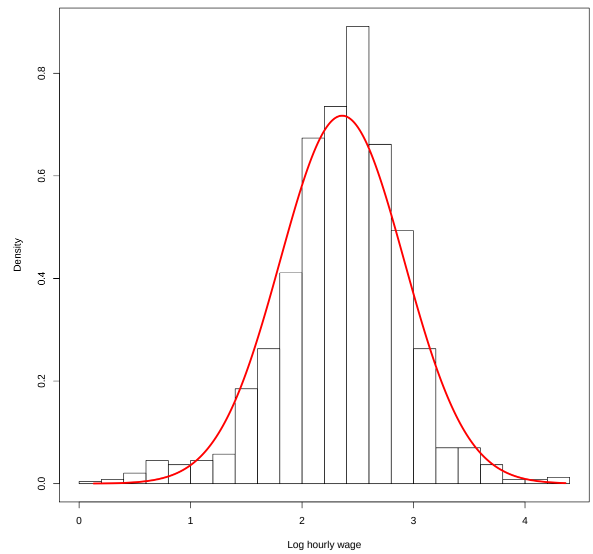
y_i: log hourly wage received by individual i.

x_i: education in years of schooling completed by individual i.

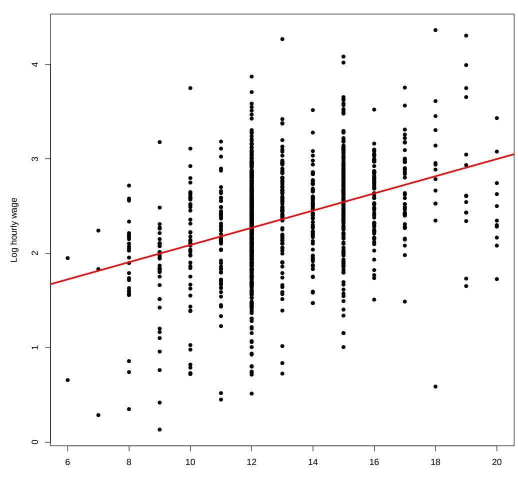
Years of schooling completed



Log hourly wage



MLE regression



$\hat{\beta} = (1.17766, 0.09101)'$ and $\hat{\sigma}^2 = 0.2668455$.

Recall the conjugate prior for (β, σ^2) is

$$\beta|\sigma^2 \sim N(b_0, \sigma^2 B_0) \quad \text{and} \quad \sigma^2 \sim IG(n_0/2, n_0 S_0/2).$$

Let us assume that $b_0 = 0$, $n_0 = 6$ and $S_0 = 0.1333$.

Let us consider two different prior variance for β:

- Prior I: $B_0 = 1.0 \times 10^1 I_2$,
- Prior II: $B_0 = 1.0 \times 10^{100} I_2$.

Posterior summary for the complete model $\mathcal{M}_0$				
Parameter	Prior I		Prior II	
	Post. Mean	Post Std.	Post Mean	Post Std.
β_0	1.17439	(0.08626)	1.17439	(0.08637)
β_1	0.09125	(0.00654)	0.09125	(0.00655)
σ^2	0.26587	(0.01073)	0.26587	(0.01073)
Posterior summary for the restricted model $\mathcal{M}_1$				
β_0	2.35963	(0.01592)	2.35963	(0.01591)
σ^2	0.30815	(0.01244)	0.30815	(0.01244)
Log Bayes factor of $\mathcal{M}_0$ versus $\mathcal{M}_1$				
$\log B_{01}$	84.7294		-29.8886	

$$\begin{aligned}\log B_{01}(\text{Prior I}) &= \log p(y|X, \mathcal{M}_0, \text{Prior I}) - \log p(y|X, \mathcal{M}_1, \text{Prior I}) \\ &= -2458.713 - (-2543.442) = 84.7294\end{aligned}$$

$$\begin{aligned}\log B_{01}(\text{Prior II}) &= \log p(y|X, \mathcal{M}_0, \text{Prior II}) - \log p(y|X, \mathcal{M}_1, \text{Prior II}) \\ &= -2686.406 - (-2656.517) = -29.8886\end{aligned}$$

1.6 Example iv. SV model

Example iv. Stochastic volatility

One of the most used models in financial econometrics is the diffusive stochastic volatility model, where log-returns are normally distributed

$$y_t|\theta_t, H \sim N(0; e^{\theta_t})$$

with heteroscedasticity modeled as

$$\theta_t|\theta_{t-1}, \gamma, H \sim N(\alpha + \beta\theta_{t-1}, \sigma^2)$$

for $t = 1, \ldots, T$ and $\gamma = (\alpha, \beta, \sigma^2)$, known for now.

The model is completed with

$$\theta_0|\gamma, H \sim N(m_0, C_0)$$

for known hyperparameters (m_0, C_0).

Example iv. Posterior distribution

For $\theta = (\theta_1, \ldots, \theta_T)'$, it follows that

$$\begin{aligned} p(\theta|y,H) &\propto \prod_{t=1}^{T} e^{-\theta_t/2} \exp\left\{-\frac{1}{2} y_t^2 e^{-\theta_t}\right\} \\ &\times \prod_{t=1}^{T} \exp\left\{-\frac{1}{2\sigma^2}(\theta_t - \alpha - \beta\theta_{t-1})^2\right\} \\ &\times \exp\left\{-\frac{1}{2C_0}(\theta_0 - m_0)^2\right\} \end{aligned}$$

Unfortunately, closed form solutions are rare!

- How to compute $E(\theta_{43}|y,H)$ or $V(\theta_{11}|y,H)$?
- How to obtain a 95% credible region for $(\theta_{35}, \theta_{36}|y,H)$?
- How to sample from $p(\theta|y,H)$?
- How to compute $p(y|H)$ or $p(y_{T+1}, \ldots, y_{T+k}|y,H)$?

Chapter 2

Bayesian Computation

2.1 A bit of history

MC in the 40s and 50s

Stan Ulam soon realized that computers could be used in this fashion to answer questions of neutron diffusion and mathematical physics;

He contacted John Von Neumann and they developed many Monte Carlo algorithms (importance sampling, rejection sampling, etc);

In the 1940s Nick Metropolis and Klari Von Neumann designed new controls for the state-of-the-art computer (ENIAC);

Metropolis and Ulam (1949) The Monte Carlo method. *Journal of the American Statistical Association.*
Metropolis *et al.* (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics.*

Monte Carlo methods

We introduce several Monte Carlo (MC) methods for integrating and/or sampling from nontrivial densities.

- MC integration
  - Simple MC integration
  - MC integration via importance sampling (IS)
- MC sampling
  - Rejection method
  - Sampling importance resampling (SIR)
- Iterative MC sampling
  - Metropolis-Hastings algorithms

- Simulated annealing
- Gibbs sampler

Based on the book by Gamerman and Lopes (1996).

A few references

- MC integration (Geweke, 1989)
- Rejection methods (Gilks and Wild, 1992)
- SIR (Smith and Gelfand, 1992)
- Metropolis-Hastings algorithm (Hastings, 1970)
- Simulated annealing (Metropolis *et al.*, 1953)
- Gibbs sampler (Gelfand and Smith, 1990)

Two main tasks

1. Compute high dimensional integrals:

$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta$$

2. Obtain

   a sample $\{\theta_1, \ldots, \theta_n\}$ from $\pi(\theta)$

   when only

   a sample $\{\tilde{\theta}_1, \ldots, \tilde{\theta}_m\}$ from $q(\theta)$

   is available.

$q(\theta)$ is known as the *proposal/auxiliary* density.

Bayes via MC

MC methods appear frequently, but not exclusively, in modern Bayesian statistics.

Posterior and predictive densities are hard to sample from:

$$\begin{aligned}
\text{Posterior} &: \quad \pi(\theta) = \frac{f(x|\theta)p(\theta)}{f(x)} \\
\text{Predictive} &: \quad f(x) = \int f(x|\theta)p(\theta)d\theta
\end{aligned}$$

Other important integrals and/or functionals of the posterior and predictive densities are:

- Posterior modes: $\max_\theta \pi(\theta)$;
- Posterior moments: $E_\pi[g(\theta)]$;
- Density estimation: $\hat{\pi}(g(\theta))$;
- Bayes factors: $f(x|M_0)/f(x|M_1)$;
- Decision: $\max_d \int U(d,\theta)\pi(\theta)d\theta$.

2.2 MC integration

MC integration

The objective here is to compute moments

$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta$$

If $\theta_1, \ldots, \theta_n$ is a random sample from $\pi(\cdot)$ then

$$\bar{h}_{mc} = \frac{1}{n}\sum_{i=1}^{n} h(\theta_i) \rightarrow E_\pi[h(\theta)] \qquad \text{as } n \rightarrow \infty.$$

If, additionally, $E_\pi[h^2(\theta)] < \infty$, then

$$V_\pi[\bar{h}_{mc}] = \frac{1}{n}\int \{h(\theta) - E_\pi[h(\theta)]\}^2 \pi(\theta)d\theta$$

and

$$v_{mc} = \frac{1}{n^2}\sum_{i=1}^{n}(h(\theta_i) - \bar{h}_{mc})^2 \rightarrow V_\pi[\bar{h}_{mc}] \qquad \text{as } n \rightarrow \infty.$$

Example i. MC integration

The objective here is to compute[1]

$$p = \int_0^1 [cos(50\theta) + sin(20\theta)]^2 d\theta$$

by noticing that the above integral can be rewritten as

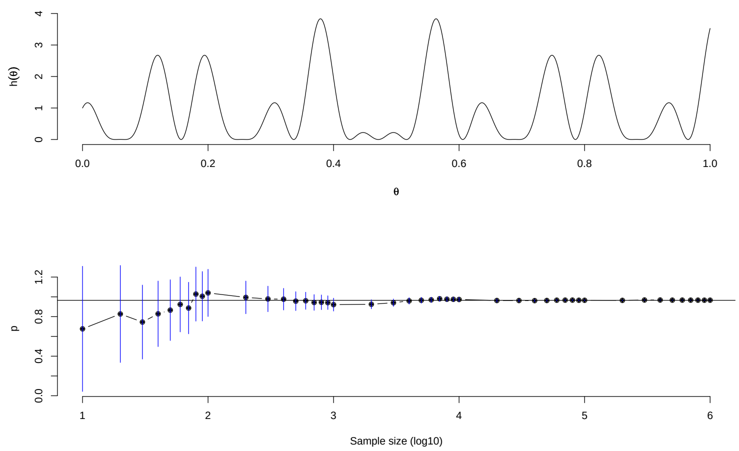
$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta$$

where $h(\theta) = [cos(50\theta) + sin(20\theta)]^2$ and $\pi(\theta) = 1$ is the density of a $U(0,1)$. Therefore

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} h(\theta_i)$$

where $\theta_1, \ldots, \theta_n$ are i.i.d. from $U(0,1)$.

[1] True value is 0.965.



2.3 MC via IS

MC via IS

The objective is still the same, ie to compute

$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta$$

by noticing that

$$E_\pi[h(\theta)] = \int \frac{h(\theta)\pi(\theta)}{q(\theta)}q(\theta)d\theta$$

where $q(\cdot)$ is an *importance function.*

If $\theta_1, \ldots, \theta_n$ is a random sample from $q(\cdot)$ then

$$\Rightarrow \bar{h}_{is} = \frac{1}{n}\sum_{i=1}^{n}\frac{h(\theta_i)\pi(\theta_i)}{q(\theta_i)} \longrightarrow E_\pi[h(\theta)]$$

as $n \longrightarrow \infty$.

Ideally, $q(\cdot)$ should be

- As *close* as possible to $h(\cdot)\pi(\cdot)$, and
- Easy to sample from.

Example ii. Cauchy tail

The objective here is to estimate

$$p = Pr(\theta > 2) = \int_2^\infty \frac{1}{\pi(1+\theta^2)}d\theta = 0.1475836$$

where θ is a standard Cauchy random variable.

A natural MC estimator of p is

$$\hat{p}_1 = \frac{1}{n}\sum_{i=1}^{n} I\{\theta_i \in (2,\infty)\}$$

where $\theta_1, \ldots, \theta_n \sim$ Cauchy(0,1).

A more elaborated estimator based on a change of variables from θ to $u = 1/\theta$ is

$$\hat{p}_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{u_i^{-2}}{2\pi[1+u_i^{-2}]}$$

where $u_1, \ldots, u_n \sim U(0, 1/2)$.

The true value is $p = 0.147584$.

n	$\hat{p}_1$	$\hat{p}_2$	$v_1^{1/2}$	$v_2^{1/2}$
100	0.100000	0.1467304	0.030000	0.001004
1000	0.137000	0.1475540	0.010873	0.000305
10000	0.148500	0.1477151	0.003556	0.000098
100000	0.149100	0.1475591	0.001126	0.000031
1000000	0.147711	0.1475870	0.000355	0.000010

With only $n = 1000$ draws, $\hat{p}_2$ has roughly the same precision that $\hat{p}_1$, which is based on $1000n$ draws, ie. three orders of magnitude.

2.4 Rejection method

Rejection method

The objective is to draw from a target density

$$\pi(\theta) = c_\pi \tilde{\pi}(\theta)$$

when only draws from an auxiliary density

$$q(\theta) = c_q \tilde{q}(\theta)$$

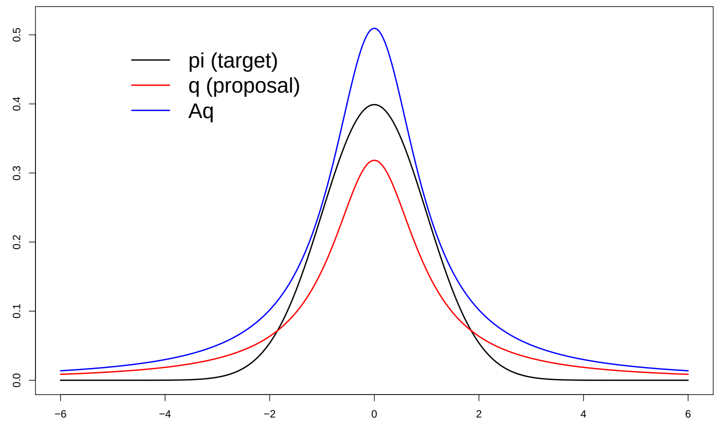
is available, for normalizing constants c_π and c_q.

If there exist a constant $A < \infty$ such that

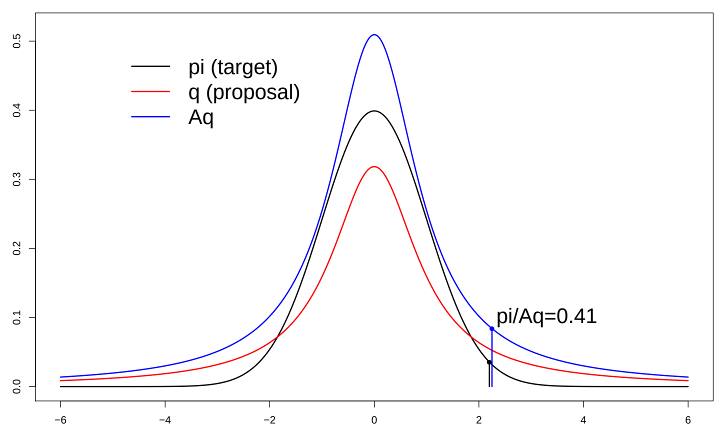
$$0 \leq \frac{\tilde{\pi}(\theta)}{A\tilde{q}(\theta)} \leq 1 \quad \text{for all } \theta$$

then $q(\theta)$ becomes a *blanketing density* or an *envelope* and A the *envelope constant.*

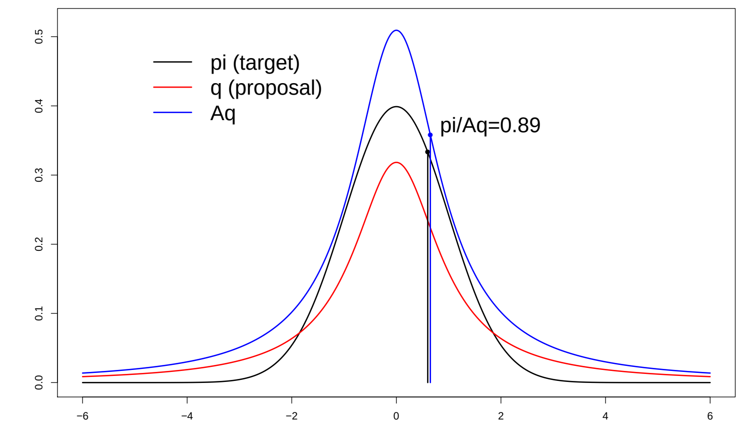
Blanket distribution



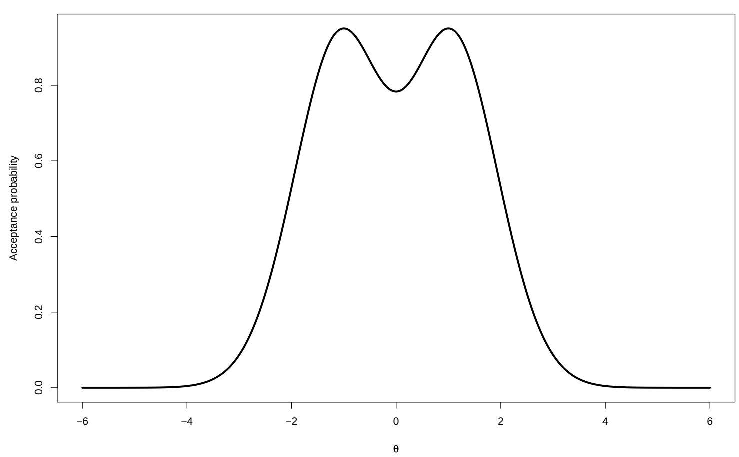
Bad draw



Good draw



Acceptance probability



Algorithm

Drawing from $\pi(\theta)$.

1. Draw θ^* from $q(\cdot)$;
2. Draw u from $U(0,1)$;
3. Accept θ^* if $u \leq \frac{\tilde{\pi}(\theta^*)}{A\tilde{q}(\theta^*)}$;
4. Repeat 1, 2 and 3 until n draws are accepted.

Normalizing constants c_π and c_q are not needed.

The theoretical acceptance rate is $\frac{c_q}{Ac_\pi}$.

The smaller the A, the larger the acceptance rate.

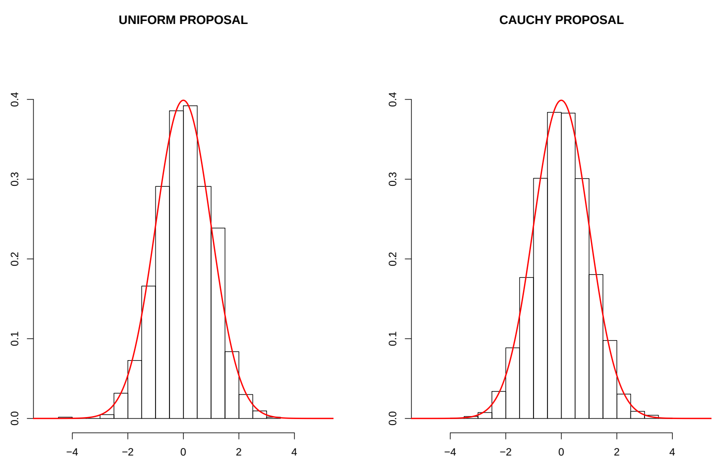
Example iii. Sampling $N(0,1)$
Enveloping the standard normal density

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}} \exp\{-0.5\theta^2\}$$

by a Cauchy density $q_C(\theta) = 1/(\pi(1+\theta^2))$, or a uniform density $q_U(\theta) = 0.05$ for $\theta \in (-10, 10)$.

Bad proposal:The maximum of $\pi(\theta)/q_U(\theta)$ is roughly $A_U = 7.98$ for $\theta \in (-10, 10)$. The theoretical acceptance rate is 12.53%.

Good proposal: The max of $\pi(\theta)/q_C(\theta)$ is equal to $A_C = \sqrt{2\pi/e} \approx 1.53$. The theoretical acceptance rate is 65.35%.



Empirical rates: 0.1265 (Uniform) and 0.6483 (Cauchy)
Theoretical rates: 0.1253 (Uniform) and 0.6535 (Cauchy)

2.5 SIR method

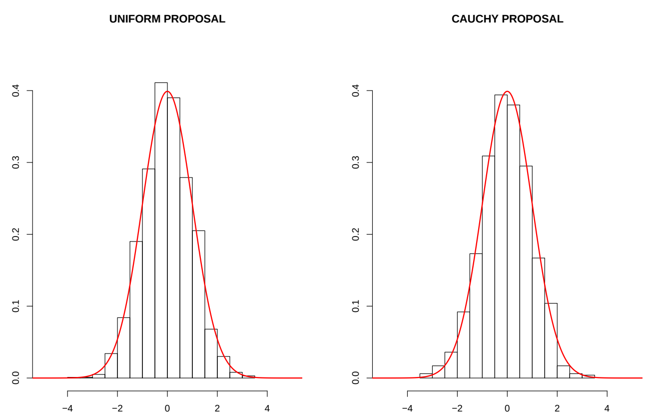
SIR method
No need to rely on the existance of A!

Algorithm

1. Draw $\theta_1^*, \ldots, \theta_n^*$ from $q(\cdot)$
2. Compute (unnormalized) weights
$$\omega_i = \pi(\theta_i^*)/q(\theta_i^*) \qquad i = 1, \ldots, n$$
3. Sample θ from $\{\theta_1^*, \ldots, \theta_n^*\}$ such that
$$Pr(\theta = \theta_i^*) \propto \omega_i \qquad i = 1, \ldots, n.$$
4. Repeat m times step 3.

Rule of thumb: $n/m = 20$.
Ideally, $\omega_i = 1/n$ and $Var(\omega) = 0$.

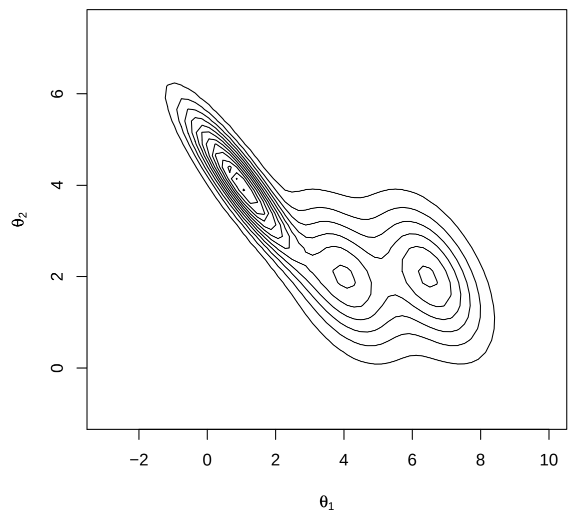
Example iii. revisited



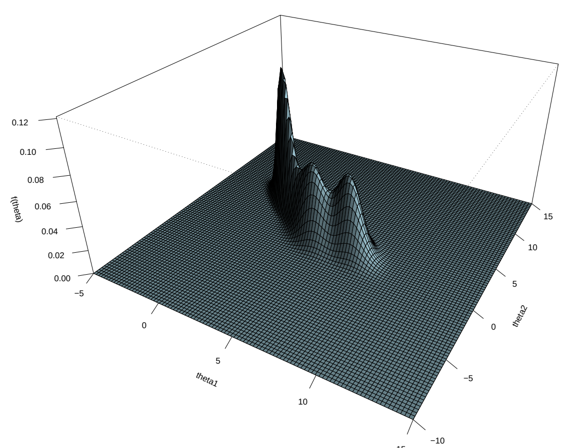
Fraction of redraws: 0.391 (Uniform) and 0.1335 (Cauchy)
Variance of weights: 4.675 (Uniform) and 0.332 (Cauchy)

Example iv. 3-component mixture
Assume that we are interested in sampling from
$$\pi(\theta) = \alpha_1 p_N(\theta; \mu_1, \Sigma_1) + \alpha_2 p_N(\theta; \mu_2, \Sigma_2) + \alpha_3 p_N(\theta; \mu_3, \Sigma_3)$$
where $p_N(\cdot; \mu, \Sigma)$ is the density of a bivariate normal distribution with mean vector μ and covariance matrix Σ. The mean vectors are
$$\mu_1 = (1,4)' \quad \mu_2 = (4,2)' \quad \mu_3 = (6.5,2),$$
the covariance matrices are
$$\Sigma_1 = \begin{pmatrix} 1.0 & -0.9 \\ -0.9 & 1.0 \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \Sigma_3 = \begin{pmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{pmatrix},$$
and weights $\alpha_1 = \alpha_2 = \alpha_3 = 1/3$.

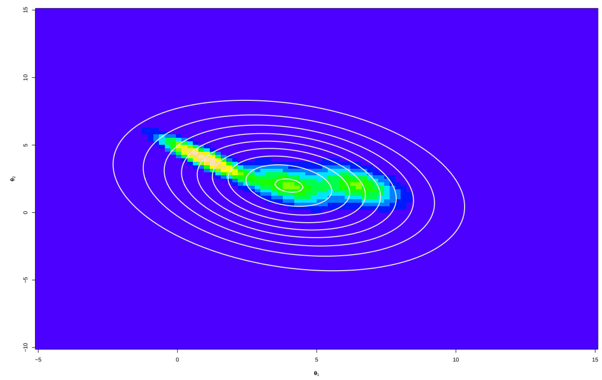
Target $\pi(\theta)$



Target $\pi(\theta)$



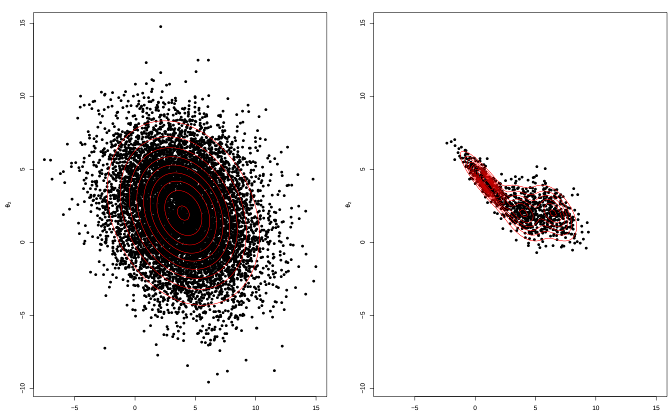
Proposal $q(\theta)$



$q(\theta) \sim N(\mu, \Sigma)$ where

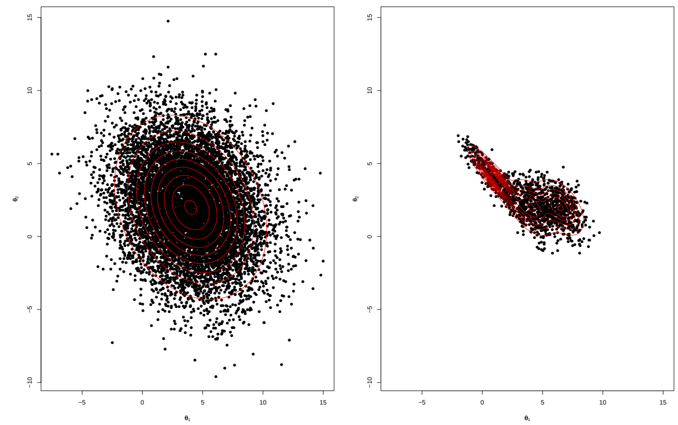
$$\mu_2 = (4,2)' \quad \text{and} \quad \Sigma = 9 \begin{pmatrix} 1.0 & -0.25 \\ -0.25 & 1.0 \end{pmatrix}$$

Rejection method



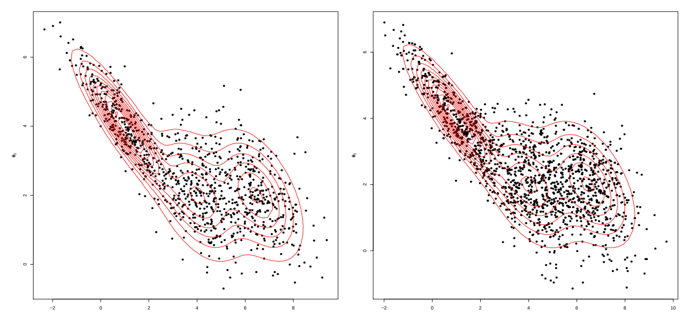
Acceptance rate: 9.91% of $n = 10,000$ draws.

SIR method



Fraction of redraws: 29.45% of $(n = 10,000, m = 2,000)$.

Rejection & SIR



Example v. 2-component mixture

Let us now assume that

$$\pi(\theta) = \alpha_1 p_N(\theta; \mu_1, \Sigma_1) + \alpha_3 p_N(\theta; \mu_3, \Sigma_3)$$

where mean vectors are

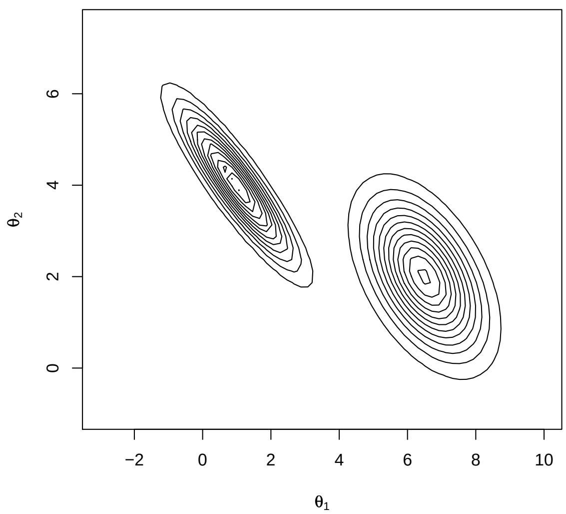
$$\mu_1 = (1, 4)' \quad \mu_3 = (6.5, 2),$$

the covariance matrices are

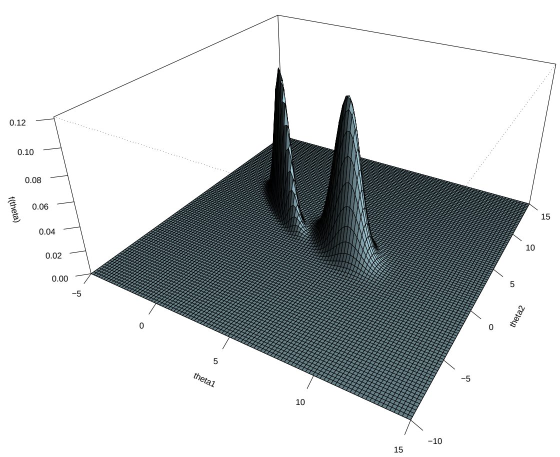
$$\Sigma_1 = \begin{pmatrix} 1.0 & -0.9 \\ -0.9 & 1.0 \end{pmatrix} \text{ and } \Sigma_3 = \begin{pmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{pmatrix},$$

and weights $\alpha_1 = 1/3$ and $\alpha_3 = 2/3$.

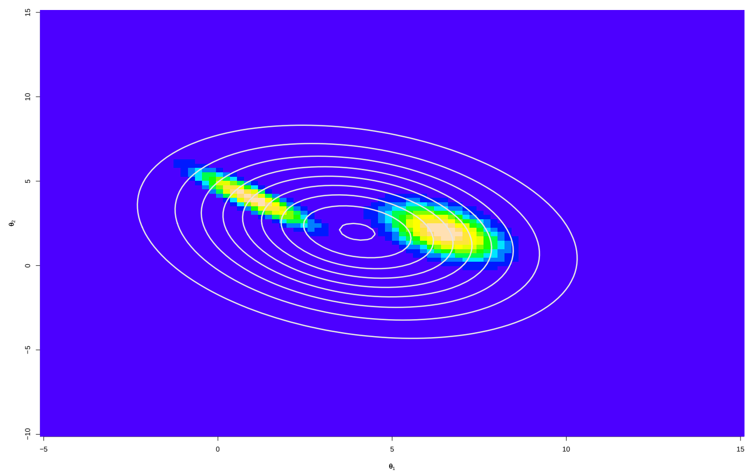
Target $\pi(\theta)$



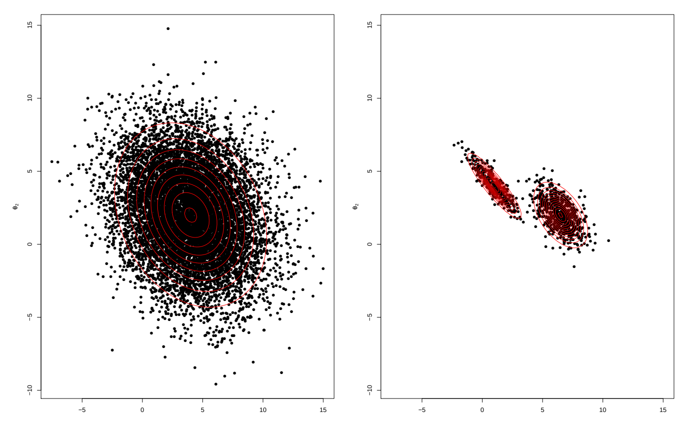
Target $\pi(\theta)$



Proposal $q(\theta)$

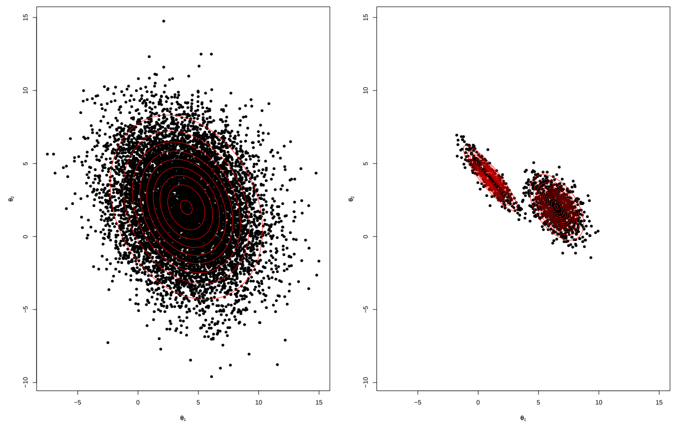


Rejection method



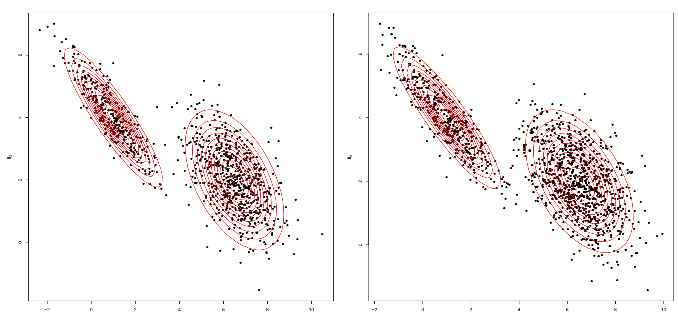
Acceptance rate: 10.1% of $n = 10,000$ draws.

SIR method



Fraction of redraws: 37.15% of $(n = 10,000, m = 2,000)$.

Rejection & SIR



MCMC history

Dongarra and Sullivan (2000) list the top algorithms with the greatest influence on the development and practice of science and engineering in the 20th century (in chronological order):

- Metropolis Algorithm for Monte Carlo
- Simplex Method for Linear Programming
- Krylov Subspace Iteration Methods
- The Decompositional Approach to Matrix Computations
- The Fortran Optimizing Compiler
- QR Algorithm for Computing Eigenvalues

- Quicksort Algorithm for Sorting
- Fast Fourier Transform

70s and 80s
Metropolis-Hastings:

Hastings (1970) and his student Peskun (1973) showed that Metropolis and the more general Metropolis-Hastings algorithm are particular instances of a larger family of algorithms.

Gibbs sampler:

Besag (1974) Spatial Interaction and the Statistical Analysis of Lattice Systems.
Geman and Geman (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images.
Pearl (1987) Evidential reasoning using stochastic simulation.
Tanner and Wong (1987). The calculation of posterior distributions by data augmentation.
Gelfand and Smith (1990) Sampling-based approaches to calculating marginal densities.

2.6 MH algorithms

MH algorithms

A sequence $\{\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots\}$ is drawn from a Markov chain whose *limiting equilibrium distribution* is the posterior distribution, $\pi(\theta)$.

Algorithm

1. Initial value: $\theta^{(0)}$
2. Proposed move: $\theta^* \sim q(\theta^*|\theta^{(i-1)})$
3. Acceptance scheme:

$$\theta^{(i)} = \begin{cases} \theta^* & \text{com prob. } \alpha \\ \theta^{(i-1)} & \text{com prob. } 1-\alpha \end{cases}$$

where

$$\alpha = \min\left\{1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})} \frac{q(\theta^{(i-1)}|\theta^*)}{q(\theta^*|\theta^{(i-1)})}\right\}$$

Special cases

1. Symmetric chains: $q(\theta|\theta^*) = q(\theta^*|\theta)$

$$\alpha = \min\left\{1, \frac{\pi(\theta^*)}{\pi(\theta)}\right\}$$

2. Independence chains: $q(\theta|\theta^*) = q(\theta)$

$$\alpha = \min\left\{1, \frac{\omega(\theta^*)}{\omega(\theta)}\right\}$$

where $\omega(\theta^*) = \pi(\theta^*)/q(\theta^*)$.

Random walk Metropolis

The most famous symmetric chain is the random walk Metropolis:

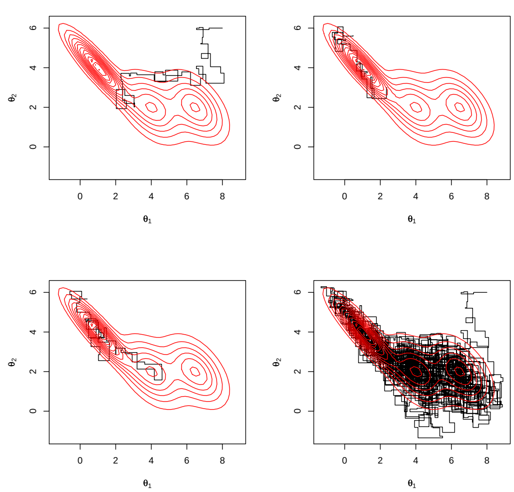
$$q(\theta|\theta^*) = q(|\theta - \theta^*|)$$

Hill climbing: when

$$\alpha = \min\left\{1, \frac{\pi(\theta^*)}{\pi(\theta)}\right\}$$

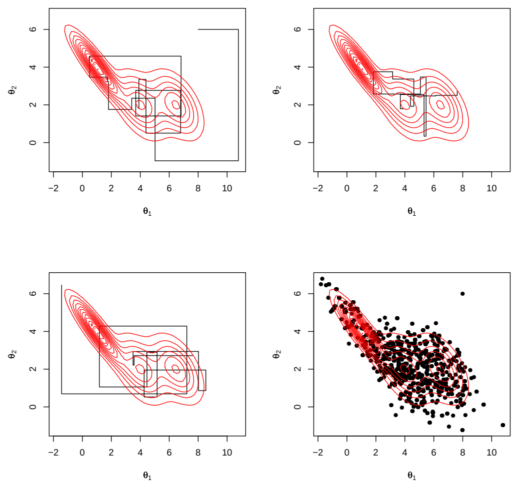
a value θ^* with higher density $\pi(\theta^*)$ greater than $\pi(\theta)$ is automatically accepted.

Example iv. RW Metropolis



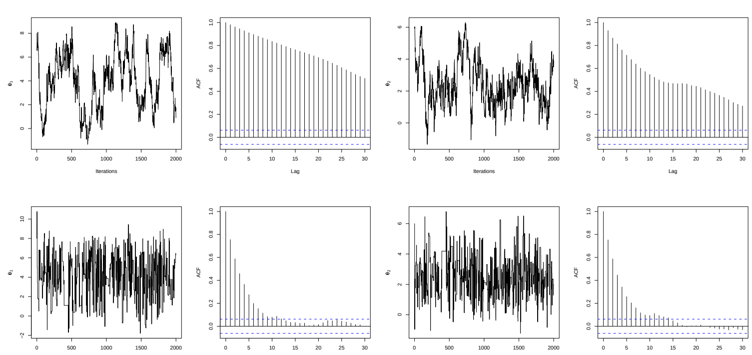
$q(\theta|\theta_i) \sim N(\theta_i, 0.25\Sigma_2)$.

Example iv. Ind. Metropolis

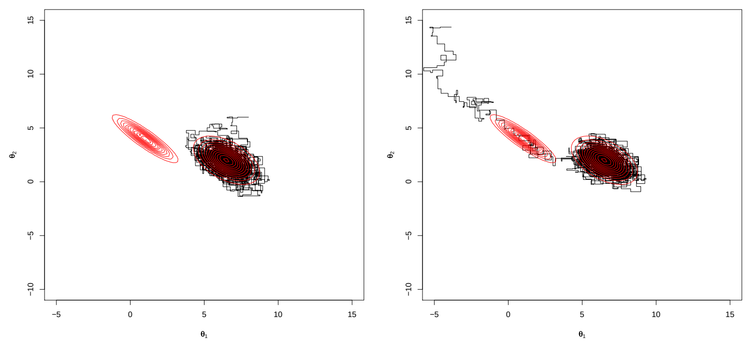


$q(\theta) \equiv q_{SIR}(\theta) \sim N(\mu, \Sigma).$

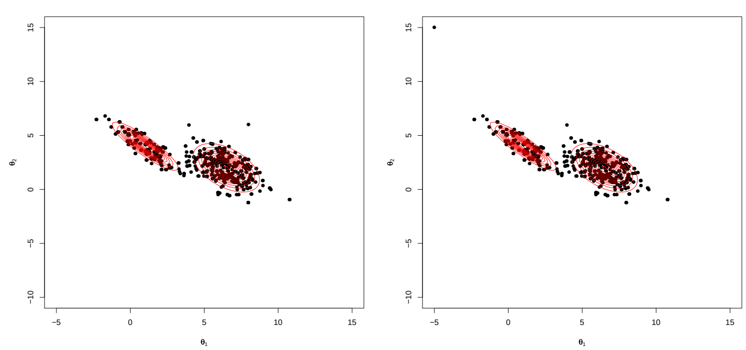
Example iv. Autocorrelations



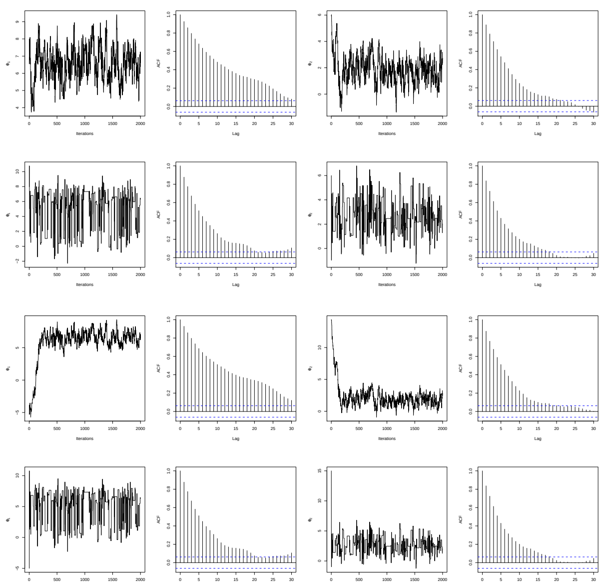
Example v. RW Metropolis



Example v. Ind. Metropolis



Example v. Autocorrelations



Example vi. tuning selection

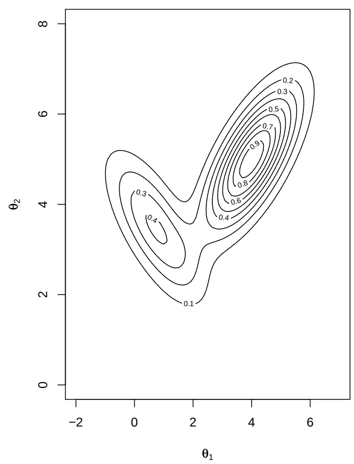
Tthe target distribution is a two-component mixture of bivariate normal densities, ie:

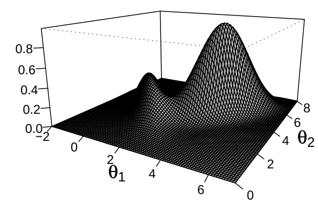
$$\pi(\theta) = 0.7 f_N(\theta; \mu_1, \Sigma_1) + 0.3 f_N(\theta; \mu_2, \Sigma_2).$$

where

$$\begin{aligned}
\mu_1' &= (4.0, 5.0) \\
\mu_2' &= (0.7, 3.5) \\
\Sigma_1 &= \begin{pmatrix} 1.0 & 0.7 \\ 0.7 & 1.0 \end{pmatrix} \\
\Sigma_2 &= \begin{pmatrix} 1.0 & -0.7 \\ -0.7 & 1.0 \end{pmatrix} .
\end{aligned}$$

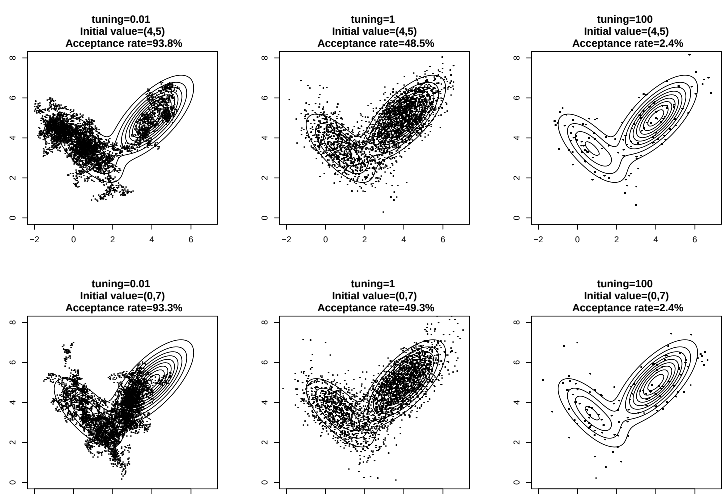
Target distribution



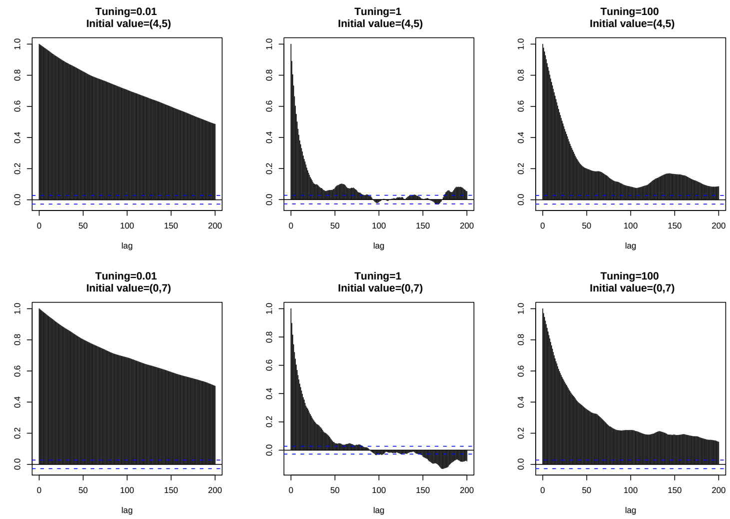


RW Metropolis

$q(\theta,\phi) = f_N(\phi;\theta,\nu I_2)$ and $\nu =$tuning.

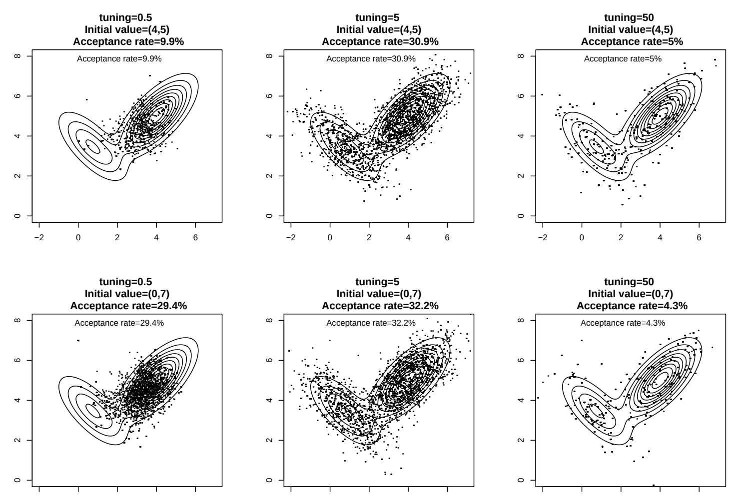


Autocorrelations

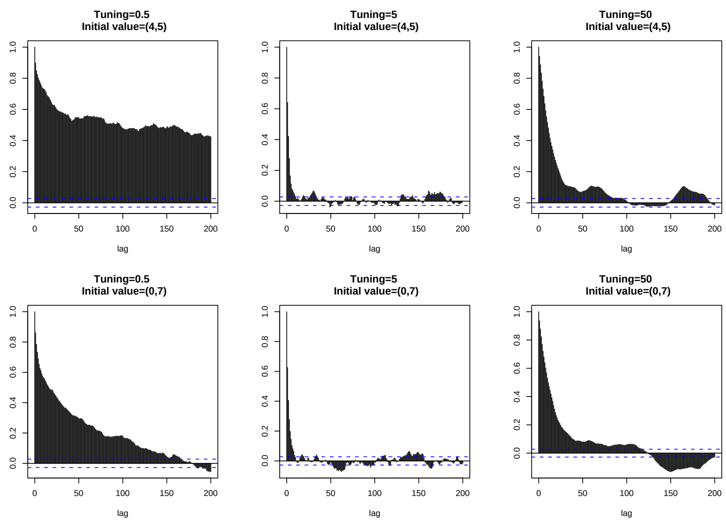


Independent Metropolis

$q(\theta, \phi) = f_N(\phi; \mu_3, \nu I_2)$ and $\mu_3 = (3.01, 4.55)'$.



Autocorrelations



2.7 Simulated annealing

Simulated annealing

Simulated annealing[2] is an optimization technique designed to find maxima of functions.

It can be seen as a M-H algorithm that *tempers* with the target distribution:

$$q(\theta) \propto \pi(\theta)^{1/T}$$

where the constant $T > 1$ receives the physical interpretation of system temperature, hence the nomenclature used (Jennison, 1993).

The *heated* distribution q is flattened with respect to π and its density gets closer to the uniform distribution, which is particularly relevant for the case of a distribution with distant modes.

By flattening the modes, the moves required to cover adequately the parameter space become more likely.

Example vii: Nonlinear surface

Assume that the goal is to find the mode/maximum of

$$\pi(\beta_1, \beta_2) \propto \prod_{i=1}^{4} \frac{e^{(\beta_1+\beta_2 x_i)y_i}}{(1+e^{\beta_1+\beta_2 x_i})^5},$$

with $x = (-0.863, -0.296, -0.053, 0.727)$ and $y = (0, 1, 3, 5)$.

[2]Kirkpatrick, Gelatt and Vecchi (1983)

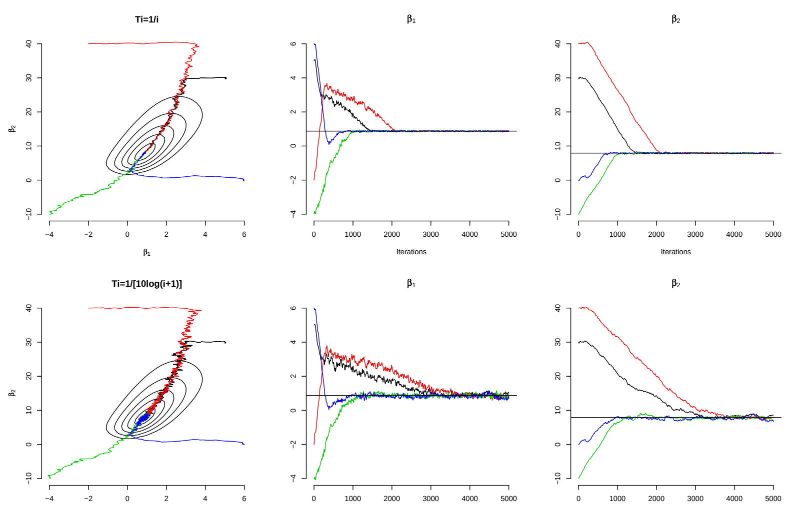
The simulated annealing algorithm is implemented for four initial values:

$$(5,30) \quad (-2,40) \quad (-4,-10) \quad (6,0)$$

and two cooling schedules:

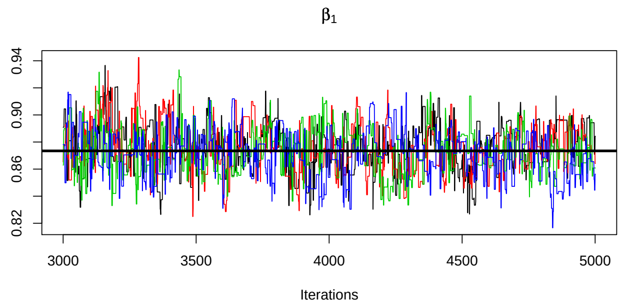
$$T_i = 1/i \quad \text{and} \quad T_i = 1/[10\log(1+i)].$$

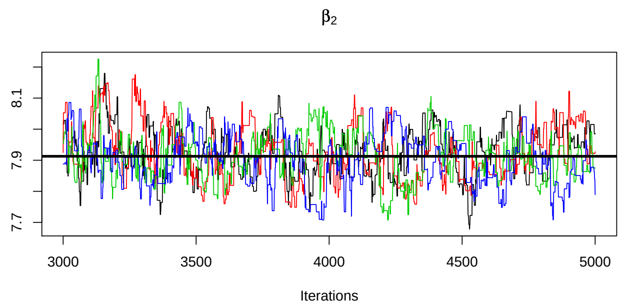
The proposal distribution is $q(\beta|\beta^{(i)}) = f_N(\beta;\beta^{(i)}, 0.05^2 I_2)$.



Newton-Raphson mode: $(0.87, 7.91)$.

$T_i = 1/i$: mode is $(0.88, 7.99)$ when $(\beta_1^{(0)}, \beta_2^{(0)}) = (5,30)$.





2.8 Gibbs sampler

Gibbs sampler

Technically, the Gibbs sampler is an MCMC scheme whose transition kernel is the product of the full conditional distributions.

Algorithm

1. Start at $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots)$
2. Sample the components of $\theta^{(j)}$ iteratively:

$$\begin{array}{lcl} \theta_1^{(j)} & \sim & \pi(\theta_1 | \theta_2^{(j-1)}, \theta_3^{(j-1)}, \ldots) \\ \theta_2^{(j)} & \sim & \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \ldots) \\ \theta_3^{(j)} & \sim & \pi(\theta_3 | \theta_1^{(j)}, \theta_2^{(j)}, \ldots) \\ & \vdots & \end{array}$$

The Gibbs sampler opened up a new way of approaching statistical modeling by combining simpler structures (the full conditional models) to address the more general structure (the full model).

Example viii: Bivariate normal

Assume that the target distribution is the bivariate normal with mean vector and covariance matrix given by

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix},$$

respectively.

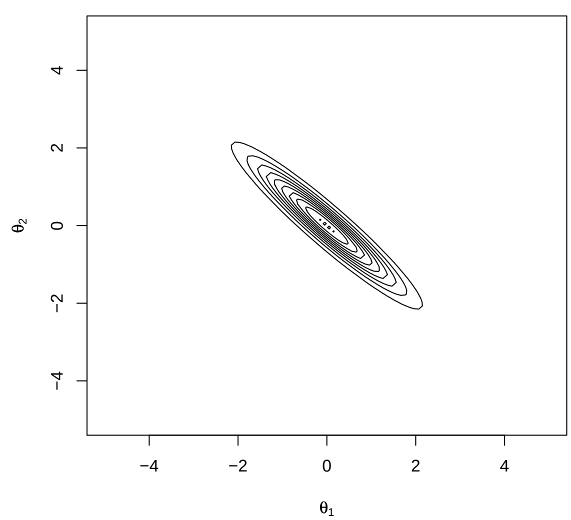
In this case, the two full conditionals are given by

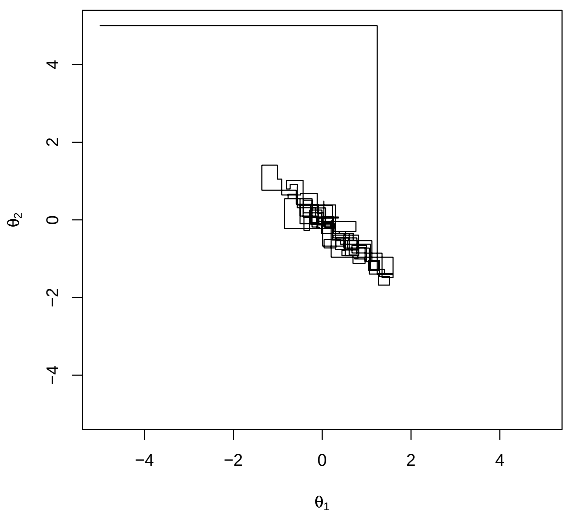
$$\theta_1|\theta_2 \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(\theta_2 - \mu_2), \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\right)$$

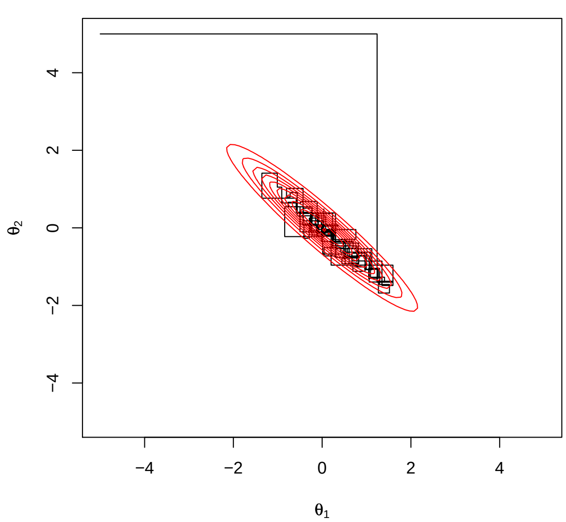
and

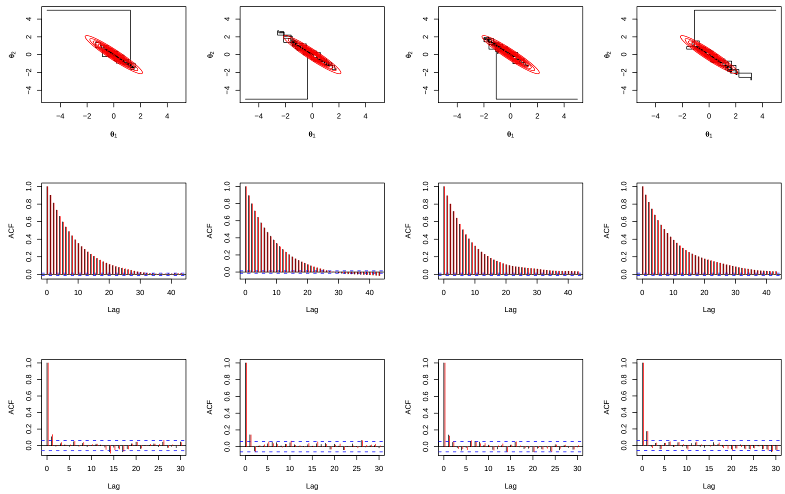
$$\theta_2|\theta_1 \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(\theta_1 - \mu_1), \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}\right)$$

$\mu_1 = \mu_2 = 0$
$\sigma_1^2 = \sigma_2^2 = 1$
$\sigma_{12} = -0.95$



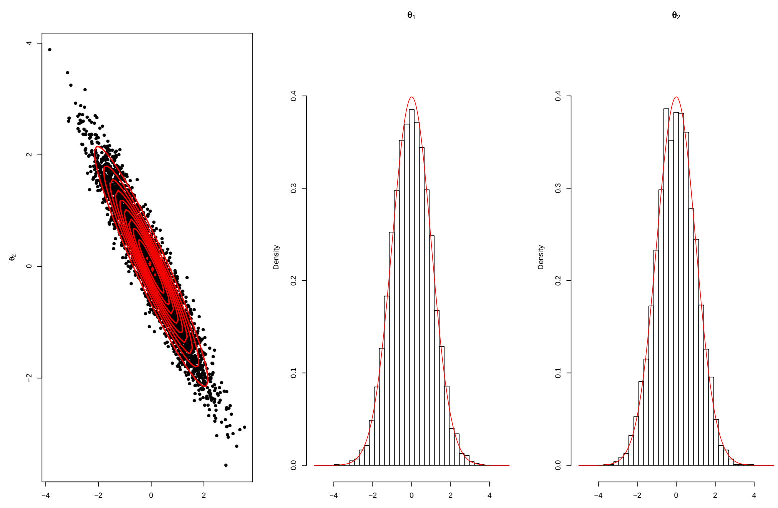

θ₁
θ₂
−4
−2
0
2
4


θ₁
θ₂
−4
−2
0
2
4



Middle frame: Based on $M = 21,000$ consecutive draws.

Bottom frame: Based on $M = 1000$ draws, after initial $M_0 = 1000$ draws and saving every 20th draws.



2.9 SUR model

Example: Seemingly Unrelated Regressions

Investments Grunfeld (1958), Boot and White (1960) and Zellner (1962,1963) study

$$I_{mt} = \beta_{m1} + \beta_{m2}F_{mt} + \beta_{m3}C_{mt} + \varepsilon_{mt}$$

- Should the parameters be the same across firms?
- Do the ε_{mt} share unobserved common factors?
- Staking the observation for firm m:

$$y_m = X_m \beta_m + \varepsilon_m$$

Capital asset pricing For a given security, the CAPM specifies that

$$r_{mt} - r_{ft} = \alpha_m + \beta_m (r_{mt} - r_{ft}) + \varepsilon_{mt}$$

for return on security m, r_{mt}, return on a risk-free security, r_{ft}, and market return, r_{mt}.

- Are the disturbances correlated across securities?
- Are the α_ms and/or β_ms related in any way?
- Staking the observation for security m:

$$y_m = X_m \beta_m + \varepsilon_m$$

Gross State Product Greene (2008) examines (his examples 9.9 and 9.12) Munnell's (1990) model for output by the 48 continental US states:

$$\begin{aligned} \log GSP_{mt} &= \beta_{m1} + \beta_{m2} \log pcap_{mt} + \beta_{m3} \log hwy_{mt} \\ &+ \beta_{m4} \log water_{mt} + \beta_{m5} \log util_{mt} \\ &+ \beta_{m6} \log emp_{mt} + \beta_{m7} unemp_{mt} + \varepsilon_{mt} \end{aligned}$$

- Should the coefficient vector be the same across states?
- Should the disturbances correlated across states?
- Should the disturbances correlated across time?
- Staking the observation for state m:

$$y_m = X_m \beta_m + \varepsilon_m$$

SUR

For $m = 1, \ldots, M$ and $t = 1, \ldots, T$

$$y_{mt} = x'_{mt} \beta_m + \epsilon_{mt},$$

with x_{mt} a k_m-dimensional vector of regressors.

Let us stack all equations:

$$\begin{array}{rcll} y_t & = & (y_{1t}, \ldots, y_{Mt})' & (M \times 1) \\ \varepsilon_t & = & (\varepsilon_{1t}, \ldots, \varepsilon_{Mt})' & (M \times 1) \\ \beta & = & (\beta_1', \ldots, \beta_M')' & (k \times 1) \\ X_t & = & \mathrm{diag}(x_{1t}', \ldots, x_{Mt}') & (M \times k) \end{array}$$

where $k = \sum_{m=1}^{M} k_m$. Therefore,

$$y_t = X_t\beta + \varepsilon_t$$

We can now stack all observations $t = 1, \ldots, T$ together:

$$\begin{array}{rcl} y & = & (y_1, \ldots, y_T)' \\ \varepsilon & = & (\varepsilon_t, \ldots, \varepsilon_T)' \\ X & = & (X_1', \ldots, X_T')', \end{array}$$

such that

$$y = X\beta + \varepsilon.$$

NLRM: ε_{mt} are i.i.d. $N(0, \sigma^2)$ for all m and t.

SUR: ε_t are i.i.d. $N(0, \Sigma)$ for all t.

This leads to $\varepsilon \sim N(0, \Omega)$, where

$$\Omega = \mathrm{diag}(\Sigma, \ldots, \Sigma) = I_T \otimes \Sigma$$

is an $MT \times MT$ block-diagonal covariance matrix.

Prior distribution

Conditionally conjugate prior for β and $\Phi = \Sigma^{-1}$:

$$p(\beta, \Phi) = p(\beta)p(\Phi),$$

where

$$\beta \sim N(\beta_0, V_0)$$

and

$$\Phi \sim Wishart(\nu_0, \Phi_0).$$

See Dreze and Richard (1983) and Richard and Steel (1988) for further discussion regarding alternative prior specifications.

Full conditionals

The full conditional distributions are

$$\begin{aligned}\beta|y,X,\Sigma &\sim N(\beta_1,V_1)\\ \Phi|y,X,\beta &\sim Wishart(\nu_1,\Phi_1)\end{aligned}$$

where $\nu_1=\nu_0+T$,

$$\begin{aligned}V_1^{-1} &= V_0^{-1}+\sum_{t=1}^{T}X_t'\Phi X_t\\ V_1^{-1}\beta_1 &= V_0^{-1}\beta_0+\sum_{t=1}^{n}X_t'\Phi y_t\\ \Phi_1^{-1} &= \Phi_0^{-1}+\sum_{t=1}^{T}(y_t-X_t\beta)(y_t-X_t\beta)'\end{aligned}$$

Grunfeld's (1958) data

M=10 U.S. firms over T=20 years, 1935-1954.[3]

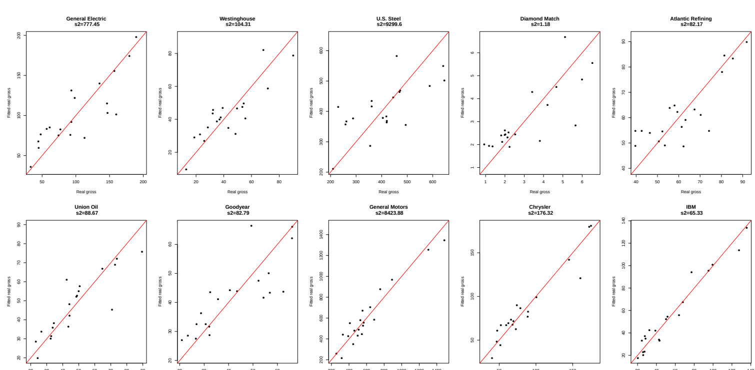
Variables:

FN = Firm Number; YR = Year; I = Annual real gross investment; F = Real value of the firm (shares outstanding); and C = Real value of the capital stock.

Firms:

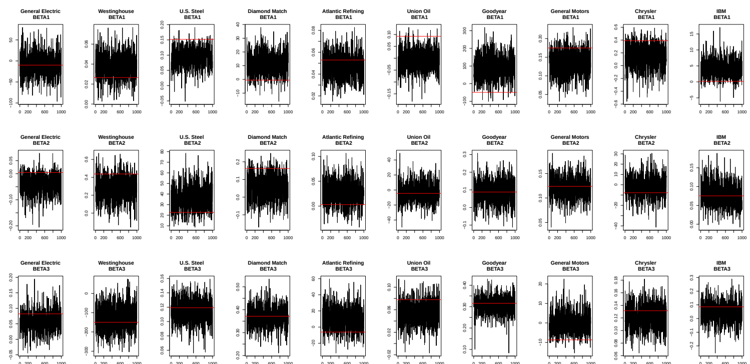
General Electric, Westinghouse, U.S. Steel, Diamond Match, Atlantic Refining, Union Oil, Goodyear, General Motors, Chrysler and IBM

Individual regressions

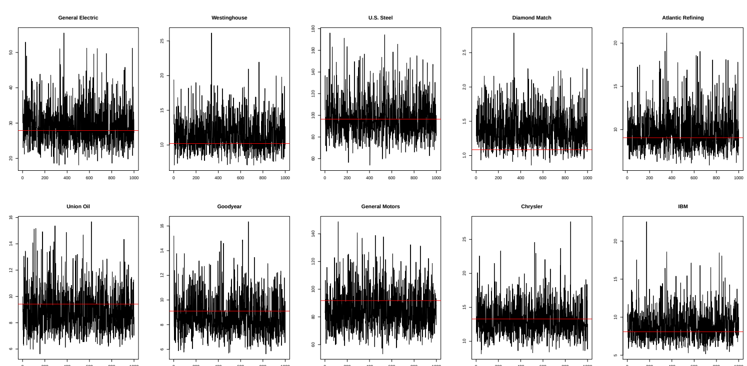


[3]Zellner (1971), pages 240-246.

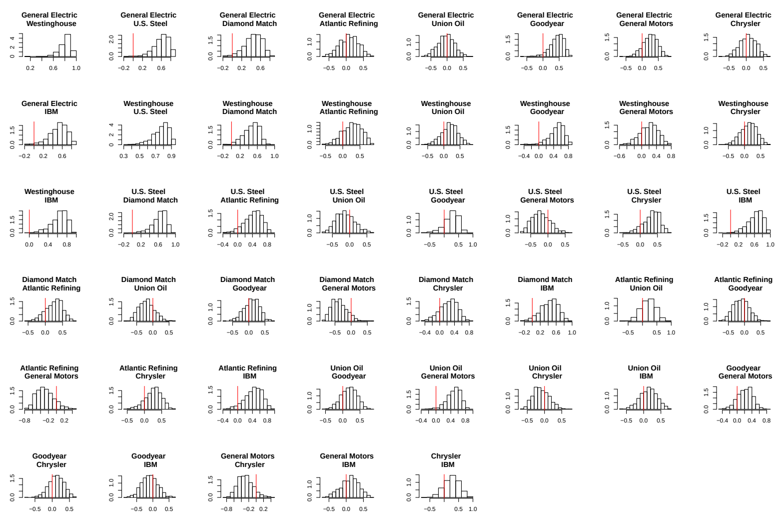
Regression coefficients



Standard deviations



Correlations



SUR references

Boot and De Wit (1960) Investment Demand: An Empirical Contribution to the Aggregation. Problem, *International Economic Review*, 1, 3-30.

Zellner (1971) *An Introduction to Bayesian Inference and Econometrics*, Wiley.

Koop (2003) *Bayesian Econometrics*, Wiley.

Kleiber and Zeileis (2010) The Grunfeld Data at 50, *German Economic Review*, 11(4), 404-417.

2.10 BVAR

Example: Bivariate BVAR(1)

Let $y_t = (y_{t1}, y_{t2})'$ contain 2 time series observed at time t.

The (basic) VAR(1) can be written as

$$(y_t | y_{t-1}, B, \Sigma) \sim N(By_{t-1}, \Sigma)$$

where

$$B = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

VAR(1) as a SUR model

The above VAR(1) model can be rewritten as a SUR model as

$$(y_t|z_t, \beta, \Sigma) \sim N(z_t\beta, \Sigma)$$

where

$$z_t = \begin{pmatrix} y_{t-1,1} & y_{t-1,2} & 0 & 0 \\ 0 & 0 & y_{t-1,1} & y_{t-1,2} \end{pmatrix}$$

and

$$\beta = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{pmatrix}.$$

Therefore,

$$(y|\beta, \Sigma) \sim N(Z\beta, \Sigma)$$

where $y = (y_1', \ldots, y_T')'$ and $Z = (z_1', \ldots, z_T')'$.

Prior of (β, Σ)

We assume that β and Σ are independent *a priori*.

Prior of β:

$$\beta \sim N(b_0, B_0).$$

Prior of Σ:

$$\Sigma \sim IW(v_0, V_0).$$

This conditionally conjugate prior DOES NOT lead to closed form posterior inference, but the implementation of the Gibbs sampler is straightforward.

Full conditional of β

It is easy to see that

$$\begin{aligned} p(\beta|\Sigma, y) &\propto \exp\left\{-0.5\left[\beta' B_0^{-1}\beta - 2\beta' B_0^{-1}\beta_0\right]\right\} \\ &\times \exp\left\{-0.5\left[\beta' Z'\Sigma^{-1}Z\beta - 2\beta' Z'\Sigma^{-1}y\right]\right\}. \end{aligned}$$

Therefore,

$$\beta|\Sigma, \gamma, y \sim N(\beta_1, V_1)$$

where

$$\beta_1 = B_1(B_0^{-1}\beta_0 + Z'\Sigma^{-1}y) \quad \text{and} \quad B_1^{-1} = B_0^{-1} + Z'\Sigma^{-1}Z.$$

Full conditional of Σ

It is easy to see that

$$
\begin{aligned}
p(\Sigma|\beta, y) &\propto |\Sigma|^{\frac{q+v_0+1}{2}} \exp\left\{-0.5\mathrm{tr}(\Sigma^{-1}V_0)\right\} \\
&\times |\Sigma|^{\frac{q+T+1}{2}} \exp\left\{-0.5\sum_{t=1}^{T}(y_t - z_t\beta)'\Sigma^{-1}(y_t - z_t\beta)\right\}.
\end{aligned}
$$

Therefore,

$$\Sigma|\beta, y \sim IW(v_1, V_1)$$

where $v_1 = v_0 + T$ and

$$S = V_0 + \sum_{t=1}^{T}(y_t - z_t\beta)(y_t - z_t\beta)'.$$

Simulated data

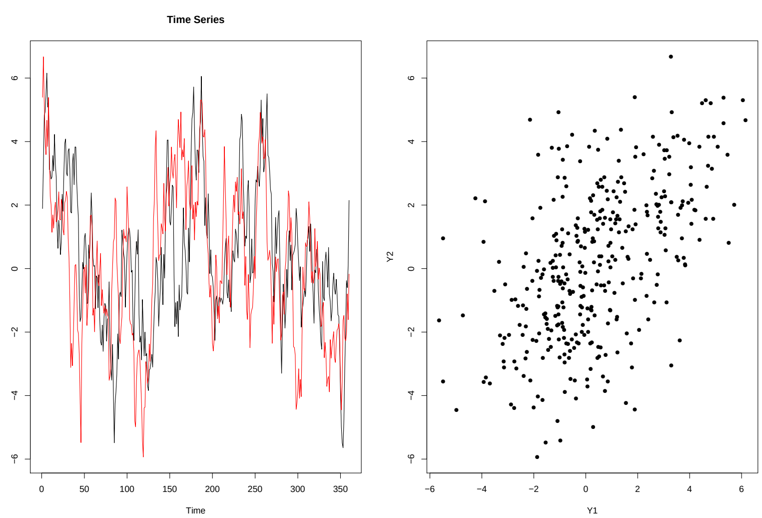
We simulated $n = 360$ observations (30 years of monthly data) from the above bivariate VAR(1) with

$$B = \begin{pmatrix} 0.85 & 0.10 \\ 0.00 & 0.95 \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{pmatrix}$$

Data



Posterior inference

The prior hyperparameters are

$$b_0 = 0_4 \quad \text{and} \quad B_0 = 1000 I_4$$

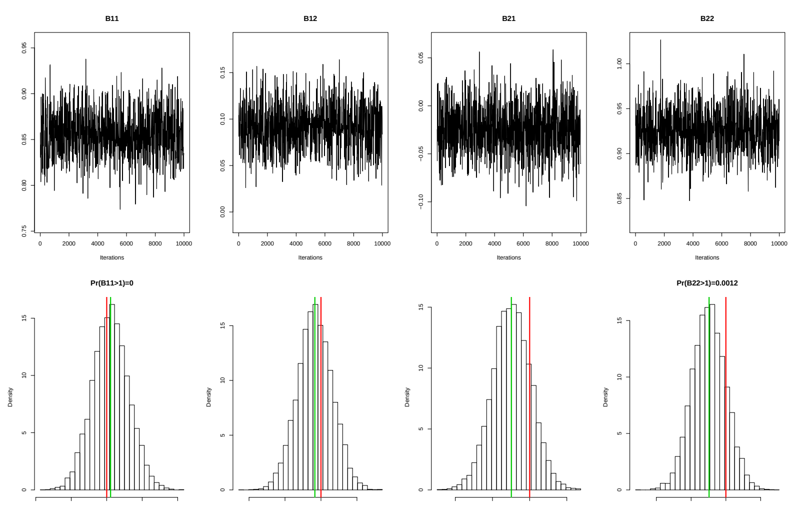
and

$$v_0 = 5 \quad \text{and} \quad V_0 = 0.001$$

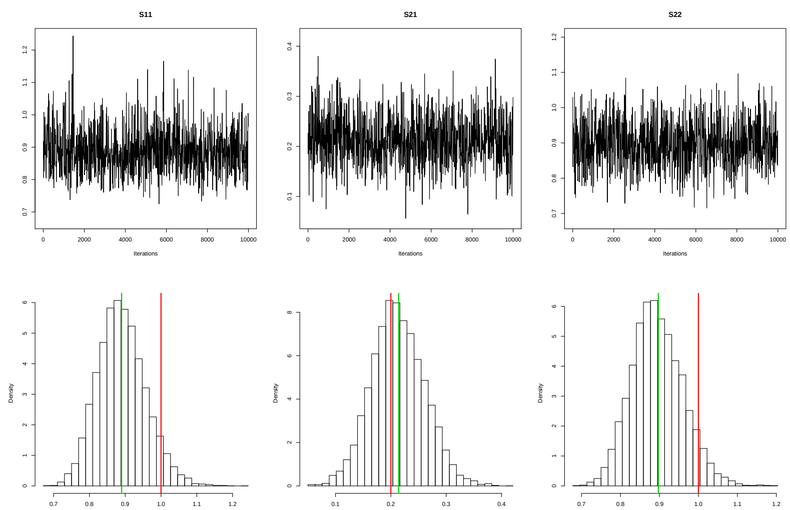
We started the Gibbs sampler with $B^{(0)} = B$ (true value).

We runt he Gibbs sampler for $M = 10,000$ iterations.

$p(B|\textbf{data})$



$p(\Sigma|\textbf{data})$



2.11 References

Book references

1. Ripley (1987) Stochastic Simulation. Wiley. (Chapters 5 and 7).

2. Thisted (1988) Elements of Statistical Computing. Chapman & Hall/CRC. (Chapter 5).

3. Gentle (1998) Random Number Generation and Monte Carlo Methods. Springer. (Chapter 5).

4. Liu (2001) Monte Carlo Strategies in Scientific Computing. Springer. (Chapters 2, 5 and 6).

5. Robert and Casella (2004) Monte Carlo Statistical Methods. Springer. (Chapters 3, 4 and 7).

6. Givens and Hoeting (2005) Computational Statistics. Wiley. (Chapters 6, 7 and 8).

7. **Gamerman and Lopes (2006) MCMC: Stochastic Simulation for Bayesian Inference. Chapman & Hall/CRC. (Chapters 3, 5 and 6).**

8. Rizzo (2008) Statistical Computing with R. Chapman & Hall/CRC. (Chapters 5 and 9).

9. Rubinstein and Kroese (2008) Simulation and the Monte Carlo Method. Wiley. (Chapters 5 and 6).

10. Robert and Casella (2010) Introducing Monte Carlo Methods with R. Springer. (Chapters 3, 6 and 7).

Seminal papers

1. Metropolis and Ulam (1949) The Monte Carlo method. JASA, 44, 335-341.

2. Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) Equation of state calculations by fast computing machines. Journal of Chemical Physics, Number 21, 1087-1092.

3. Hastings (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57, 97-109.

4. Peskun (1973) Optimum Monte-Carlo sampling using Markov chains. Biometrika, 60, 607-612.

5. Besag (1974) Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, Series B, 36, 192-236.

6. Geman and Geman (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.

7. Jennison (1993) Discussion of the meeting on Gibbs sampling and other Markov chain Monte Carlo methods. JRSS-B, 55, 54-6.

8. Kirkpatrick, Gelatt and Vecchi (1983) Optimization by simulated annealing. Science, 220, 671-80.

9. Pearl (1987) Evidential reasoning using stochastic simulation. Articial Intelligence, 32, 245-257.

10. Tanner and Wong (1987) The calculation of posterior distributions by data augmentation. JASA, 82, 528-550.

11. Geweke (1989) Bayesian inference in econometric models using Monte Carlo integration. Econometrica, 57, 1317-1339.

12. Gelfand and Smith (1990) Sampling-based approaches to calculating marginal densities, JASA, 85, 398-409.

13. Casella and George (1992) Explaining the Gibbs sampler. The American Statistician,46,167-174.

14. Smith and Gelfand (1992) Bayesian statistics without tears: a sampling-resampling perspective. American Statistician, 46, 84-88.

15. Gilks and Wild (1992) Adaptive rejection sampling for Gibbs sampling. Applied Statistics, 41, 337-48.

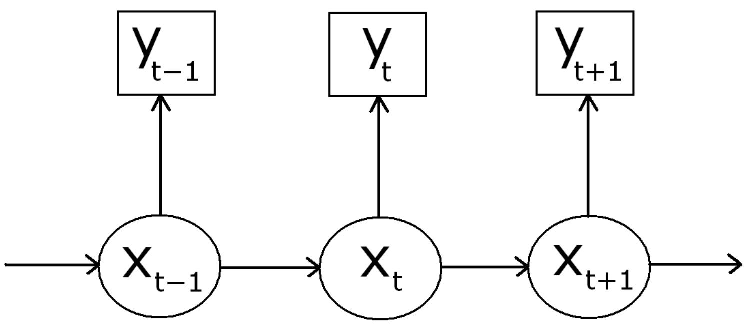
16. Chib and Greenberg (1995) Understanding the Metropolis-Hastings algorithm. The American Statistician, 49, 327-335.

17. Dongarra and Sullivan (2000) Guest editors' introduction: The top 10 algorithms. Computing in Science and Engineering, 2, 22-23.

Chapter 3

Dynamic Models

Dynamic models (DMs)



3.1 1st order DLM

1st order DLM

The local level model (West and Harrison, 1997) has

Observation equation:

$$y_{t+1}|x_{t+1}, \theta \sim N(x_{t+1}, \sigma^2)$$

System equation:

$$x_{t+1}|x_t, \theta \sim N(x_t, \tau^2)$$

where

$$x_0 \sim N(m_0, C_0)$$

and

$$\theta = (\sigma^2, \tau^2)$$

fixed (for now).

3.1.1 n-variate normal

n-variate normal
It is worth noticing that the model can be rewritten as

$$\begin{aligned} y|x,\theta &\sim N(x,\sigma_2 I_n) \\ x|x_0,\theta &\sim N(x_0 1_n,\tau^2\Omega) \\ x_0 &\sim N(m_0,C_0) \end{aligned}$$

where

$$\Omega = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & \dots & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & \dots & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & \dots & 4 & 4 & 4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & 2 & 3 & 4 & \dots & n-2 & n-2 & n-2 \\ 1 & 2 & 3 & 4 & \dots & n-1 & n-1 & n-1 \\ 1 & 2 & 3 & 4 & \dots & n-2 & n-1 & n \end{pmatrix}$$

Therefore, the prior of x given θ is

$$x|\theta \sim N(m_0 1_n; C_0 1_n 1_n' + \tau^2\Omega),$$

while its full conditional posterior distribution is

$$x|y,\theta \sim N(m_1,C_1)$$

where

$$C_1^{-1} = (C_0 1_n 1_n' + \tau^2\Omega)^{-1} + \sigma^{-2} I_n$$

and

$$C_1^{-1} m_1 = (C_0 1_n 1_n' + \tau^2\Omega)^{-1} m_0 1_n + \sigma^{-2} y$$

3.1.2 The Kalman filter

The Kalman filter
Let $y^t = (y_1,\dots,y_t)$. The previous joint posterior posterior for x given y (omitting θ for now) can be constructed as

$$p(x|y^n) = p(x_1|y^n,x_2)\prod_{t=1}^{n} p(x_t|y^n,x_{t+1}),$$

which is obtained from

$$p(x^n|y^n)$$

and noticing that given y^t and x_{t+1},

- x_t and x_{t+h} are independent, and
- x_t and y_t are independent,

for all integer $h > 1$.

Therefore, we first need to derive the above joint and this is done forward via the well-known Kalman filter recursions.

$$p(x_t|y^t) \Longrightarrow p(x_{t+1}|y^t) \Longrightarrow p(y_{t+1}|x_t) \Longrightarrow p(x_{t+1}|y^{t+1})$$

- Posterior at t: $(x_t|y^t) \sim N(m_t, C_t)$
- Prior at $t+1$: $(x_{t+1}|y^t) \sim N(m_t, R_{t+1})$
$$R_{t+1} = C_t + \tau^2$$
- Marginal likelihood: $(y_{t+1}|y^t) \sim N(m_t, Q_{t+1})$
$$Q_{t+1} = R_{t+1} + \sigma^2$$
- Posterior at $t+1$: $(x_{t+1}|y^{t+1}) \sim N(m_{t+1}, C_{t+1})$
$$\begin{aligned} m_{t+1} &= (1 - A_{t+1})m_t + A_{t+1}y_{t+1} \\ C_{t+1} &= A_{t+1}\sigma^2 \end{aligned}$$
where $A_{t+1} = R_{t+1}/Q_{t+1}$.

3.1.3 The Kalman smoother

The Kalman smoother

For $t = n$, $x_n|y^n \sim N(m_n^n, C_n^n)$, where $m_n^n = m_n$ and $C_n^n = C_n$.

For $t < n$,

$$\begin{aligned} x_t|y^n &\sim N(m_t^n, C_t^n) \\ x_t|x_{t+1}, y^n &\sim N(a_t^n, R_t^n) \end{aligned}$$

where

$$\begin{aligned} m_t^n &= (1 - B_t)m_t + B_t m_{t+1}^n \\ C_t^n &= (1 - B_t)C_t + B_t^2 C_{t+1}^n \\ a_t^n &= (1 - B_t)m_t + B_t x_{t+1} \\ R_t^n &= B_t \tau^2 \end{aligned}$$

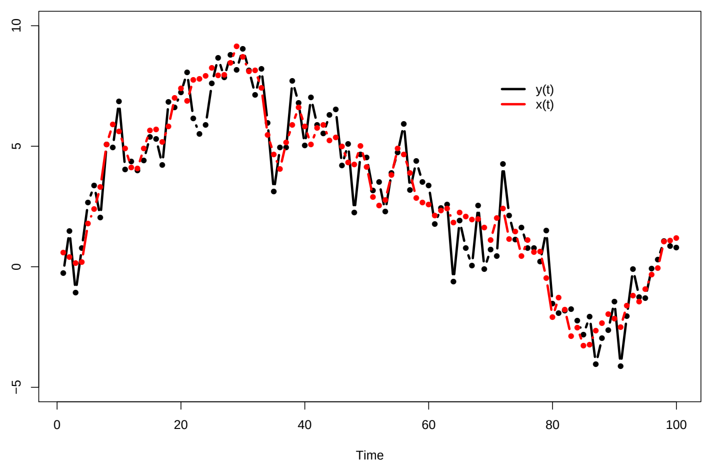
and

$$B_t = C_t/(C_t + \tau^2).$$

3.1.4 Example

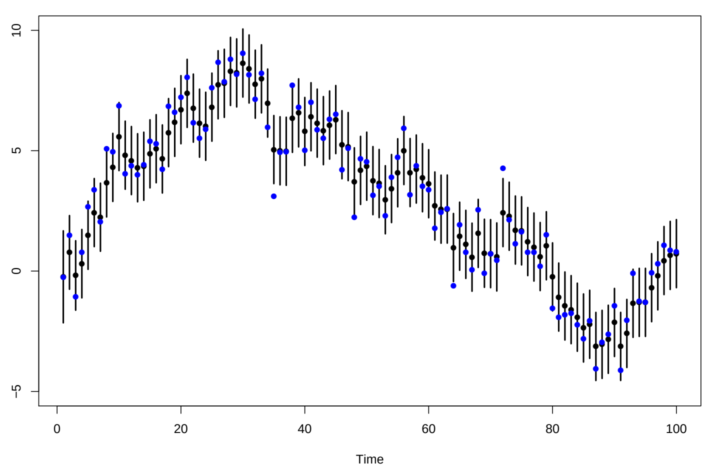
Example

$n = 100$, $\sigma^2 = 1.0$
$\tau^2 = 0.5$ and $x_0 = 0$.

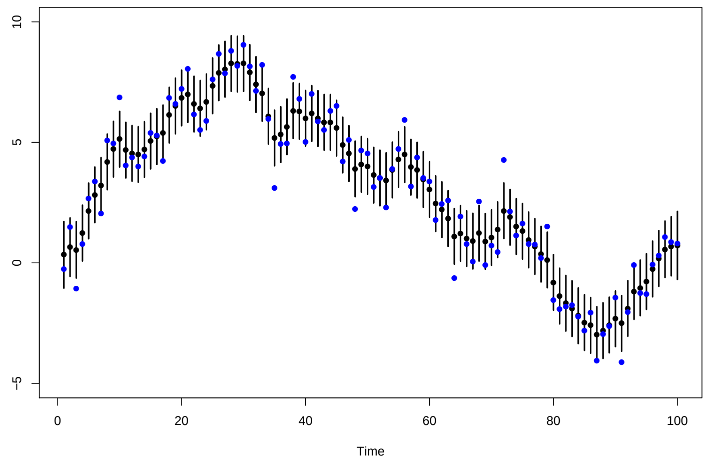


$p(x_t|y^t)$ via Kalman filter

$m_0 = 0.0$ and $C_0 = 10.0$
given τ^2 and σ^2



$p(x_t|y^n)$ via Kalman smoother
$m_0 = 0.0$ and $C_0 = 10.0$
given τ^2 and σ^2



3.1.5 Integrating out states x^n

Integrating out states x^n
We showed earlier that

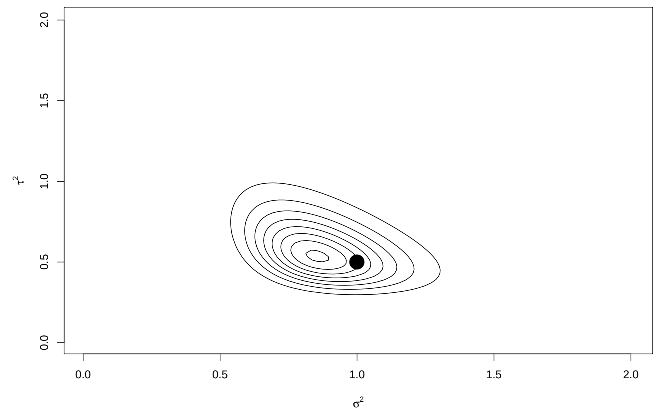
$$(y_t|y^{t-1}) \sim N(m_{t-1}, Q_t)$$

where both m_{t-1} and Q_t were presented before and are functions of $\theta = (\sigma^2, \tau^2)$, y^{t-1}, m_0 and C_0.

Therefore, by Bayes' rule,

$$\begin{aligned} p(\theta|y^n) &\propto p(\theta)p(y^n|\theta) \\ &= p(\theta)\prod_{t=1}^{n} f_N(y_t; m_{t-1}, Q_t). \end{aligned}$$

Example: $p(y|\sigma^2, \tau^2)p(\sigma^2)p(\tau^2)$
$\sigma^2 \sim IG(\nu_0/2, \nu_0\sigma_0^2/2)$, where $\nu_0 = 5$ and $\sigma_0^2 = 1$.
$\tau^2 \sim IG(n_0/2, n_0\tau_0^2/2)$, where $n_0 = 5$ and $\tau_0^2 = 0.5$



3.1.6 MCMC scheme

MCMC scheme

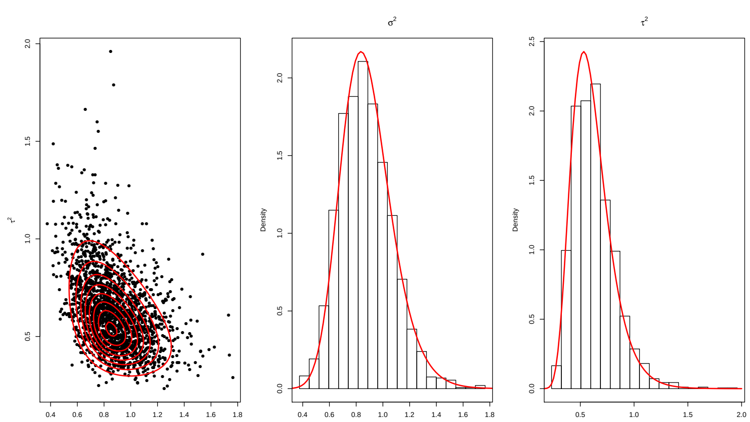
- Sample θ from $p(\theta|y^n, x^n)$

$$p(\theta|y^n, x^n) \propto p(\theta) \prod_{t=1}^{n} p(y_t|x_t, \theta) p(x_t|x_{t-1}, \theta).$$

- Sample x^n from $p(x^n|y^n, \theta)$

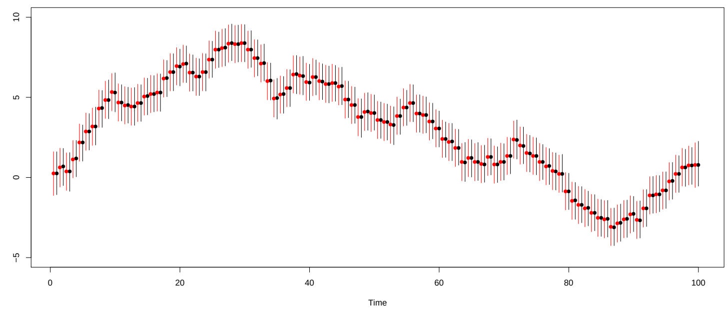
$$p(x^n|y^n, \theta) = \prod_{t=1}^{n} f_N(x_t|a_t^n, R_t^n)$$

Example: $p(x_t|y^n)$



Example: Comparison

$p(x_t|y^n)$ versus $p(x_t|y^n, \tilde{\sigma}^2 = 0.87, \tilde{\tau}^2 = 0.63)$.



3.1.7 Lessons

Lessons from the 1st order DLM

Sequential learning in non-normal and nonlinear dynamic models $p(y_{t+1}|x_{t+1})$ and $p(x_{t+1}|x_t)$ in general rather difficult since

$$\begin{aligned} p(x_{t+1}|y^t) &= \int p(x_{t+1}|x_t)p(x_t|y^t)dx_t \\ p(x_{t+1}|y^{t+1}) &\propto p(y_{t+1}|x_{t+1})p(x_{t+1}|y^t) \end{aligned}$$

are usually unavailable in closed form.

Over the last 20 years:

- FFBS for conditionally Gaussian DLMs;
- Gamerman (1998) for generalized DLMs;
- Carlin, Polson and Stoffer (2002) for more general DMs.

3.2 Dynamic linear models (DLMs)

Dynamic linear models (DLMs)

Large class of models with time-varying parameters.

Dynamic linear models are defined by a pair of equations, the *observation equation* and the *evolution/system equation*:

$$\begin{aligned} y_t &= F_t'\beta_t + \epsilon_t, \qquad \epsilon_t \sim N(0, V) \\ \beta_t &= G_t\beta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W) \end{aligned}$$

- y_t: sequence of observations;
- F_t: vector of explanatory variables;

- β_t: d-dimensional state vector;
- G_t: $d \times d$ evolution matrix;
- $\beta_1 \sim N(a, R)$.

3.2.1 Linear growth model

Linear growth model

The linear growth model is slightly more elaborate by incorporation of an extra time-varying parameter β_2 representing the growth of the level of the series:

$$\begin{aligned} y_t &= \beta_{1,t} + \epsilon_t \quad \epsilon_t \sim N(0, V) \\ \beta_{1,t} &= \beta_{1,t-1} + \beta_{2,t} + \omega_{1,t} \\ \beta_{2,t} &= \beta_{2,t-1} + \omega_{2,t} \end{aligned}$$

where $\omega_t = (\omega_{1,t}, \omega_{2,t})' \sim N(0, W)$ and

$$\begin{aligned} F_t &= (1, 0)' \\ G_t &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

3.2.2 Bivariate VAR(1) with TVP

Bivariate VAR(1) with TVP

Observation equation

$$y_t = F_t'\beta_t + \epsilon_t \qquad \epsilon_t \sim N(0, V)$$

where

$$y_t = (y_{1t}, y_{2t})',$$

$$F_t' = \begin{pmatrix} y_{1,t-1} & y_{2,t-1} & 0 & 0 \\ 0 & 0 & y_{1,t-1} & y_{2,t-1} \end{pmatrix}$$

and

$$\beta_t = (\beta_{11t}, \beta_{12t}, \beta_{21t}, \beta_{22t})'$$

Evolution equation

$$\beta_t = G_t\beta_{t-1} + \omega_t \qquad \omega_t \sim N(0, W)$$

Prior, updated and smoothed distributions

Prior distributions

$$p(\beta_t|y^{t-k}) \qquad k > 0$$

Updated/online distributions

$$p(\beta_t|y^t)$$

Smoothed distributions

$$p(\beta_t|y^{t+k}) \qquad k > 0$$

3.2.3 Sequential inference

Sequential inference

Let $y^t = \{y_1, \ldots, y_t\}$.

Posterior at time $t-1$:

$$\beta_{t-1}|y^{t-1} \sim N(m_{t-1}, C_{t-1})$$

Prior at time t:

$$\beta_t|y^{t-1} \sim N(a_t, R_t)$$

with $a_t = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t' + W$.

predictive at time t:

$$y_t|y^{t-1} \sim N(f_t, Q_t)$$

with $f_t = F_t' a_t$ and $Q_t = F_t' R_t F_t + V$.

Posterior at time t

$$p(\beta_t|y^t) = p(\beta_t|y_t, y^{t-1}) \propto p(y_t|\beta_t)\, p(\beta_t|y^{t-1})$$

The resulting posterior distribution is

$$\beta_t|y^t \sim N(m_t, C_t)$$

with

$$\begin{aligned} m_t &= a_t + A_t e_t \\ C_t &= R_t - A_t A_t' Q_t \\ A_t &= R_t F_t / Q_t \\ e_t &= y_t - f_t \end{aligned}$$

By induction, these distributions are valid for all times.

3.2.4 Smoothing

Smoothing

In dynamic models, the smoothed distribution $\pi(\beta|y^n)$ is more commonly used:

$$\begin{aligned}\pi(\beta|y^n) &= p(\beta_n|y^n)\prod_{t=1}^{n-1}p(\beta_t|\beta_{t+1},\ldots,\beta_n,y^n)\\ &= p(\beta_n|y^n)\prod_{t=1}^{n-1}p(\beta_t|\beta_{t+1},y^t)\end{aligned}$$

Integrating with respect to $(\beta_1,\ldots,\beta_{t-1})$:

$$\begin{aligned}\pi(\beta_t,\ldots,\beta_n|y^n) &= p(\beta_n|y^n)\prod_{k=t}^{n-1}p(\beta_k|\beta_{k+1},y^t)\\ \pi(\beta_t,\beta_{t+1}|y^n) &= p(\beta_{t+1}|y^n)p(\beta_t|\beta_{t+1},y^t)\end{aligned}$$

for $t=1,\ldots,n-1$.

Smoothing: $p(\beta_t|y^n)$

It can be shown that

$$\beta_t|V,W,y^n \sim N(m_t^n,C_t^n)$$

where

$$\begin{aligned}m_t^n &= m_t + C_tG'_{t+1}R_{t+1}^{-1}(m_{t+1}^n - a_{t+1})\\ C_t^n &= C_t - C_tG'_{t+1}R_{t+1}^{-1}(R_{t+1}-C_{t+1}^n)R_{t+1}^{-1}G_{t+1}C_t\end{aligned}$$

Smoothing: $p(\beta|y^n)$

It can be shown that

$$(\beta_t|\beta_{t+1},V,W,y^n)$$

is normally distributed with mean

$$(G_t'W^{-1}G_t + C_t^{-1})^{-1}(G_t'W^{-1}\beta_{t+1}+C_t^{-1}m_t)$$

and variance $(G_t'W^{-1}G_t + C_t^{-1})^{-1}$.

3.2.5 The FFBS

Forward filtering, backward sampling (FFBS)

Sampling from $\pi(\beta|y^n)$ can be performed by

- Sampling β_n from $N(m_n,C_n)$ and then
- Sampling β_t from $(\beta_t|\beta_{t+1},V,W,y^t)$, for $t=n-1,\ldots,1$.

The above scheme is known as the forward filtering, backward sampling (FFBS) algorithm (Carter and Kohn, 1994 and Frühwirth-Schnatter, 1994).

3.2.6 Individual sampling

Individual sampling from $\pi(\beta_t|\beta_{-t}, y^n)$

Let $\beta_{-t} = (\beta_1, \ldots, \beta_{t-1}, \beta_{t+1}, \ldots, \beta_n)$.

For $t = 2, \ldots, n-1$

$$\begin{aligned}
\pi(\beta_t|\beta_{-t}, y^n) &\propto p(y_t|\beta_t)\, p(\beta_{t+1}|\beta_t)\, p(\beta_t|\beta_{t-1}) \\
&\propto f_N(y_t; F_t'\beta_t, V) f_N(\beta_{t+1}; G_{t+1}\beta_t, W) \\
&\times f_N(\beta_t; G_t\beta_{t-1}, W) \\
&= f_N(\beta_t; b_t, B_t)
\end{aligned}$$

where

$$\begin{aligned}
b_t &= B_t(\sigma^{-2}F_t y_t + G_{t+1}'W^{-1}\beta_{t+1} + W^{-1}G_t\beta_{t-1}) \\
B_t &= (\sigma^{-2}F_tF_t' + G_{t+1}'W^{-1}G_{t+1} + W^{-1})^{-1}
\end{aligned}$$

for $t = 2, \ldots, n-1$.

For $t = 1$ and $t = n$,

$$\pi(\beta_1|\beta_{-1}, y^n) = f_N(\beta_1; b_1, B_1)$$

and

$$\pi(\beta_n|\beta_{-n}, y^n) = f_N(\beta_t; b_n, B_n)$$

where

$$\begin{aligned}
b_1 &= B_1(\sigma_1^{-2}F_1y_1 + G_2'W^{-1}\beta_2 + R^{-1}a) \\
B_1 &= (\sigma_1^{-2}F_1F_1' + G_2'W^{-1}G_2 + R^{-1})^{-1} \\
b_n &= B_n(\sigma_n^{-2}F_ny_n + W^{-1}G_n\beta_{n-1}) \\
B_n &= (\sigma_n^{-2}F_nF_n' + W^{-1})^{-1}
\end{aligned}$$

Sampling from $\pi(V, W|y^n, \beta)$

Assume that

$$\begin{aligned}
\phi = V^{-1} &\sim Gamma(n_\sigma/2, n_\sigma S_\sigma/2) \\
\Phi = W^{-1} &\sim Wishart(n_W/2, n_W S_W/2)
\end{aligned}$$

Full conditionals

$$\begin{aligned}
\pi(\phi|\beta, \Phi) &\propto \prod_{t=1}^{n} f_N(y_t; F_t'\beta_t, \phi^{-1})\ f_G(\phi; n_\sigma/2, n_\sigma S_\sigma/2) \\
&\propto f_G(\phi; n_\sigma^*/2, n_\sigma^* S_\sigma^*/2) \\
\pi(\Phi|\beta, \phi) &\propto \prod_{t=2}^{n} f_N(\beta_t; G_t\beta_{t-1}, \Phi^{-1})\ f_W(\Phi; n_W/2, n_W S_W/2) \\
&\propto f_W(\Phi; n_W^*/2, n_W^* S_W^*/2)
\end{aligned}$$

where $n_\sigma^* = n_\sigma + n$, $n_W^* = n_W + n - 1$,

$$\begin{aligned} n_\sigma^* S_\sigma^* &= n_\sigma S_\sigma + \sigma(y_t - F_t'\beta_t)^2 \\ n_W^* S_W^* &= n_W S_W + \Sigma_{t=2}^n (\beta_t - G_t\beta_{t-1})(\beta_t - G_t\beta_{t-1})' \end{aligned}$$

3.2.7 Joint sampling

MCMC scheme to sample from $p(\beta, V, W|y^n)$

- Sample V^{-1} from its full conditional

$$f_G(\phi; n_\sigma^*/2, n_\sigma^* S_\sigma^*/2)$$

- Sample W^{-1} from its full conditional

$$f_W(\Phi; n_W^*/2, n_W^* S_W^*/2)$$

- Sample β from its full conditional

$$\pi(\beta|y^n, V, W)$$

  by the FFBS algorithm.

Likelihood for (V, W)
It is easy to see that

$$p(y^n|V, W) = \prod_{t=1}^n f_N(y_t|f_t, Q_t)$$

which is the integrated likelihood of (V, W).

Jointly sampling (β, V, W)
(β, V, W) can be sampled jointly by

- Sampling (V, W) from its marginal posterior

$$\pi(V, W|y^n) \propto l(V, W|y^n)\pi(V, W)$$

  by a rejection or Metropolis-Hastings step;

- Sampling β from its full conditional

$$\pi(\beta|y^n, V, W)$$

  by the FFBS algorithm.

Jointly sampling (β, V, W) avoids MCMC convergence problems associated with the posterior correlation between model parameters (Gamerman and Moreira, 2002).

3.2.8 Example

Example: Comparing schemes[1]

First order DLM with $V = 1$

$$
\begin{aligned}
y_t &= \beta_t + \epsilon_t\,, & \epsilon_t &\sim N(0,1) \\
\beta_t &= \beta_{t-1} + \omega_t\,, & \omega_t &\sim N(0,W),
\end{aligned}
$$

with $(n, W) \in \{(100, .01), (100, .5), (1000, .01), (1000, .5)\}$.

400 runs: 100 replications per combination.

Priors: $\beta_1 \sim N(0, 10)$ and V and W have inverse Gammas with means set at true values and coefficients of variation set at 10.

Posterior inference: based on 20,000 MCMC draws.

Effective sample size

For a given θ, let $t^{(n)} = t(\theta^{(n)})$, $\gamma_k = Cov_\pi(t^{(n)}, t^{(n+k)})$, the variance of $t^{(n)}$ as $\sigma^2 = \gamma_0$, the autocorrelation of lag k as $\rho_k = \gamma_k/\sigma^2$ and $\tau_n^2/n = Var_\pi(\bar{t}_n)$. It can be shown that, as $n \to \infty$,

$$
\tau_n^2 = \sigma^2 \left(1 + 2 \sum_{k=1}^{n-1} \frac{n-k}{n} \rho_k \right) \to \sigma^2 \underbrace{\left(1 + 2 \sum_{k=1}^{\infty} \rho_k \right)}_{\text{inefficiency factor}} .
$$

The *inefficiency factor* measures how far $t^{(n)}$s are from being a random sample and how much $Var_\pi(\bar{t}_n)$ increases because of that.

The *effective sample size* is defined as

$$
n_{\text{eff}} = \frac{n}{1 + 2\sum_{k=1}^{\infty} \rho_k}
$$

or the size of a random sample with the same variance.

Schemes

Scheme I: Sampling $\beta_1, \ldots, \beta_n$, V, W from their conditionals.
Scheme II: Sampling β, V and W from their conditionals.
Scheme III: Jointly sampling (β, V, W).

Scheme	n=100	n=1000
II	1.7	1.9
III	1.9	7.2

[1] Gamerman, Reis and Salazar (2006) Comparison of sampling schemes for dynamic linear models. *International Statistical Review*, 74, 203-214.

Computing times relative to scheme I.

		Scheme		
W	n	I	II	III
0.01	1000	242	8938	2983
0.01	100	3283	13685	12263
0.50	1000	409	3043	963
0.50	100	1694	3404	923

Sample averages (based on the 100 replications) of n_{eff} based on V.

3.3 SV model

Stochastic volatility model

The canonical stochastic volatility model (SVM), is

$$\begin{aligned} y_t &= e^{h_t/2}\varepsilon_t \\ h_t &= \mu + \phi h_{t-1} + \tau\eta_t \end{aligned}$$

where ε_t and η_t are $N(0,1)$ shocks with $E(\varepsilon_t\eta_{t+h}) = 0$ for all h and $E(\varepsilon_t\varepsilon_{t+l}) = E(\eta_t\eta_{t+l}) = 0$ for all $l \neq 0$.

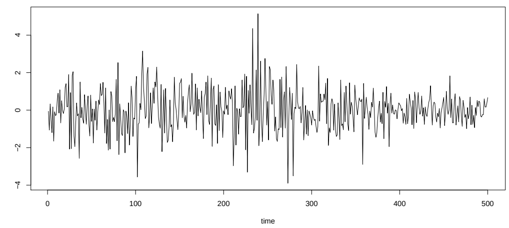
τ^2: volatility of the log-volatility.

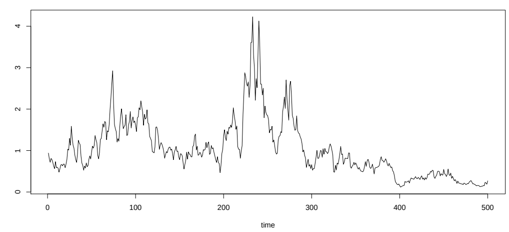
$|\phi| < 1$ then h_t is a stationary process.

Let $y^n = (y_1, \ldots, y_n)'$, $h^n = (h_1, \ldots, h_n)'$ and $h_{a:b} = (h_a, \ldots, h_b)'$.

Simulated data

$n = 500$, $h_0 = 0.0$ and $(\mu, \phi, \tau^2) = (-0.00645, 0.99, 0.0225)$





Prior information

Uncertainty about the initial log volatility is $h_0 \sim N(m_0, C_0)$.

Let $\theta = (\mu, \phi)'$, then the prior distribution of (θ, τ^2) is normal-inverse gamma, i.e. $(\theta, \tau^2) \sim NIG(\theta_0, V_0, \nu_0, s_0^2)$:

$$\begin{aligned}\theta|\tau^2 &\sim N(\theta_0, \tau^2 V_0)\\ \tau^2 &\sim IG(\nu_0/2, \nu_0 s_0^2/2)\end{aligned}$$

For example, if $\nu_0 = 10$ and $s_0^2 = 0.018$ then

$$\begin{aligned}E(\tau^2) &= \frac{\nu_0 s_0^2/2}{\nu_0/2 - 1} = 0.0225\\ Var(\tau^2) &= \frac{(\nu_0 s_0^2/2)^2}{(\nu_0/2-1)^2(\nu_0/2-2)} = (0.013)^2\end{aligned}$$

Hyperparameters: m_0, C_0, θ_0, V_0, ν_0 and s_0^2.

Simulated data

Simulation setup

- $h_0 = 0.0$
- $\mu = -0.00645$
- $\phi = 0.99$
- $\tau^2 = 0.0225$

Prior distribution

- $h_0 \sim N(0, 100)$
- $\mu \sim N(0, 100)$
- $\phi \sim N(0, 100)$
- $\tau^2 \sim IG(5, 0.14)$ (Mode=0.0234; 95% c.i.=(0.014; 0.086))

Posterior inference

The SVM is a dynamic model and posterior inference via MCMC for the the latent log-volatility states h_t can be performed in at least two ways.

Let $h_{-t} = (h_{0:(t-1)}, h_{(t+1):n})$, for $t = 1, \ldots, n-1$ and $h_{-n} = h_{1:(n-1)}$.

- Individual moves for h_t
  - $(\theta, \tau^2 | h^n, y^n)$
  - $(h_t | h_{-t}, \theta, \tau^2, y^n)$, for $t = 1, \ldots, n$

- Block move for h^n
  - $(\theta, \tau^2 | h^n, y^n)$
  - $(h^n | \theta, \tau^2, y^n)$

Sampling $(\theta, \tau^2 | h^n, y^n)$

Conditional on $h_{0:n}$, the posterior distribution of (θ, τ^2) is also normal-inverse gamma:

$$(\theta, \tau^2 | y^n, h_{0:n}) \sim NIG(\theta_1, V_1, \nu_1, s_1^2)$$

where $X = (1_n, h_{0:(n-1)})$, $\nu_1 = \nu_0 + n$

$$\begin{aligned} V_1^{-1} &= V_0^{-1} + X'X \\ V_1^{-1}\theta_1 &= V_0^{-1}\theta_0 + X'h_{1:n} \\ \nu_1 s_1^2 &= \nu_0 s_0^2 + (y - X\theta_1)'(y - X\theta_1) \\ &+ (\theta_1 - \theta_0)'V_0^{-1}(\theta_1 - \theta_0) \end{aligned}$$

Sampling $(h_0|\theta, \tau^2, h_1)$

Combining

$$h_0 \sim N(m_0, C_0)$$

and

$$h_1|h_0 \sim N(\mu + \phi h_0, \tau^2)$$

leads to (by Bayes' theorem)

$$h_0|h_1 \sim N(m_1, C_1)$$

where

$$\begin{aligned} C_1^{-1} m_1 &= C_0^{-1} m_0 + \phi \tau^{-2}(h_1 - \mu) \\ C_1^{-1} &= C_0^{-1} + \phi^2 \tau^{-2} \end{aligned}$$

Conditional prior distribution of h_t

Given h_{t-1}, θ and τ^2, it can be shown that

$$\begin{pmatrix} h_t \\ h_{t+1} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu + \phi h_{t-1} \\ (1+\phi)\mu + \phi^2 h_{t-1} \end{pmatrix}, \tau^2 \begin{pmatrix} 1 & \phi \\ \phi & (1+\phi^2) \end{pmatrix} \right\}.$$

$E(h_t|h_{t-1}, h_{t+1}, \theta, \tau^2)$ and $V(h_t|h_{t-1}, h_{t+1}, \theta, \tau^2)$ are

$$\begin{aligned} \mu_t &= \left(\frac{1-\phi}{1+\phi^2}\right)\mu + \left(\frac{\phi}{1+\phi^2}\right)(h_{t-1} + h_{t+1}) \\ \nu^2 &= \tau^2(1+\phi^2)^{-1}. \end{aligned}$$

Therefore,

$$\begin{aligned} (h_t|h_{t-1}, h_{t+1}, \theta, \tau^2) &\sim N(\mu_t, \nu^2) \qquad t = 1, \ldots, n-1 \\ (h_n|h_{n-1}, \theta, \tau^2) &\sim N(\mu_n, \tau^2) \end{aligned}$$

where $\mu_n = \mu + \phi h_{n-1}$.

Sampling h_t **via RWM**

Let $\nu_t^2 = \nu^2$ for $t = 1, \ldots, n-1$ and $\nu_n^2 = \tau^2$, then

$$p(h_t|h_{-t}, y^n, \theta, \tau^2) = f_N(h_t; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t})$$

for $t = 1, \ldots, n$.

RWM with tuning v_h^2 $(t = 1, \ldots, n)$:

1. Current state: $h_t^{(j)}$
2. Sample h_t^* from $N(h_t^{(j)}, v_h^2)$

3. Compute the acceptance probability

$$\alpha = \min\left\{1, \frac{f_N(h_t^*; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^*})}{f_N(h_t^{(j)}; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^{(j)}})}\right\}$$

4. New state:

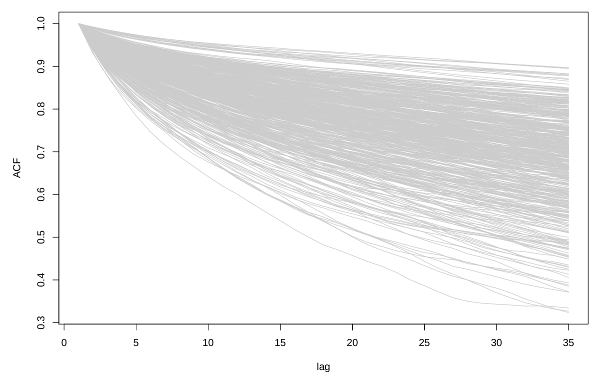
$$h_t^{(j+1)} = \begin{cases} h_t^* & \text{w. p. } \alpha \\ h_t^{(j)} & \text{w. p. } 1-\alpha \end{cases}$$

Simulated data

MCMC setup

- $M_0 = 1,000$
- $M = 1,000$

Autocorrelation of h_ts



Sampling h_t via IMH

The full conditional distribution of h_t is given by

$$\begin{aligned} p(h_t|h_{-t}, y^n, \theta, \tau^2) &= p(h_t|h_{t-1}, h_{t+1}, \theta, \tau^2) p(y_t|h_t) \\ &= f_N(h_t; \mu_t, \nu^2) f_N(y_t; 0, e^{h_t}). \end{aligned}$$

Kim, Shephard and Chib (1998) explored the fact that

$$\log p(y_t|h_t) = \text{const} - \frac{1}{2}h_t - \frac{y_t^2}{2}\exp(-h_t)$$

and that a Taylor expansion of $\exp(-h_t)$ around μ_t leads to

$$\begin{aligned} \log p(y_t|h_t) &\approx \text{const} - \frac{1}{2}h_t - \frac{y_t^2}{2}\left(e^{-\mu_t} - (h_t - \mu_t)e^{-\mu_t}\right) \\ g(h_t) &= \exp\left\{-\frac{1}{2}h_t(1 - y_t^2 e^{-\mu_t})\right\} \end{aligned}$$

Proposal distribution

Let $\nu_t^2 = \nu^2$ for $t = 1, \ldots, n-1$ and $\nu_n^2 = \tau^2$.

Then, by combining $f_N(h_t; \mu_t, \nu_t^2)$ and $g(h_t)$, for $t = 1, \ldots, n$, leads to the following proposal distribution:

$$q(h_t|h_{-t}, y^n, \theta, \tau^2) \equiv N\left(h_t; \tilde{\mu}_t, \nu_t^2\right)$$

where $\tilde{\mu}_t = \mu_t + 0.5\nu_t^2(y_t^2 e^{-\mu_t} - 1)$.

IMH algorithm

For $t = 1, \ldots, n$

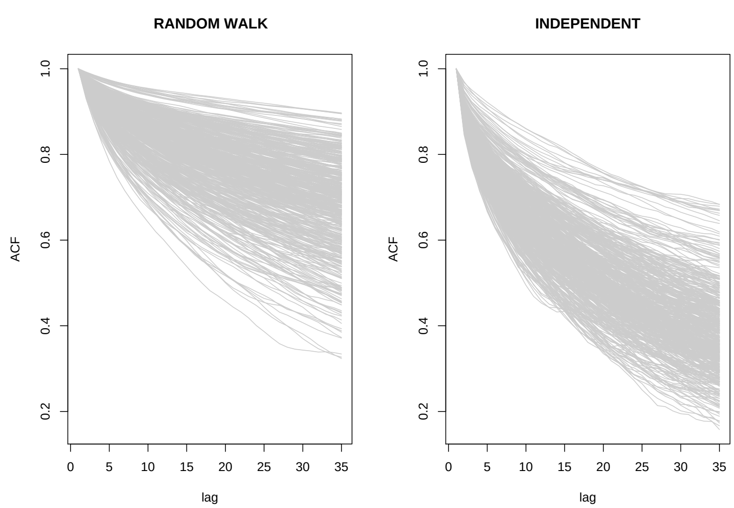
1. Current state: $h_t^{(j)}$
2. Sample h_t^* from $N(\tilde{\mu}_t, \nu_t^2)$
3. Compute the acceptance probability

$$\alpha = \min\left\{1, \frac{f_N(h_t^*; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^*})}{f_N(h_t^{(j)}; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^{(j)}})} \times \frac{f_N(h_t^{(j)}; \tilde{\mu}_t, \nu_t^2)}{f_N(h_t^*; \tilde{\mu}_t, \nu_t^2)}\right\}$$

4. New state:

$$h_t^{(j+1)} = \begin{cases} h_t^* & \text{w. p. } \alpha \\ h_t^{(j)} & \text{w. p. } 1-\alpha \end{cases}$$

ACF for both schemes



Sampling h^n - normal approximation and FFBS

Let $y_t^* = \log y_t^2$ and $\epsilon_t = \log \varepsilon_t^2$.

The SV-AR(1) is a DLM with nonnormal observational errors, i.e.

$$
\begin{aligned}
y_t^* &= h_t + \epsilon_t \\
h_t &= \mu + \phi h_{t-1} + \tau\eta_t
\end{aligned}
$$

where $\eta_t \sim N(0,1)$.

The distribution of ϵ_t is $\log\chi_1^2$, where

$$
\begin{aligned}
E(\epsilon_t) &= -1.27 \\
V(\epsilon_t) &= \frac{\pi^2}{2} = 4.935
\end{aligned}
$$

Normal approximation

Let ϵ_t be approximated by $N(\alpha, \sigma^2)$, $z_t = y_t^* - \alpha$, $\alpha = -1.27$ and $\sigma^2 = \pi^2/2$.

Then

$$
\begin{aligned}
z_t &= h_t + \sigma v_t \\
h_t &= \mu + \phi h_{t-1} + \tau\eta_t
\end{aligned}
$$

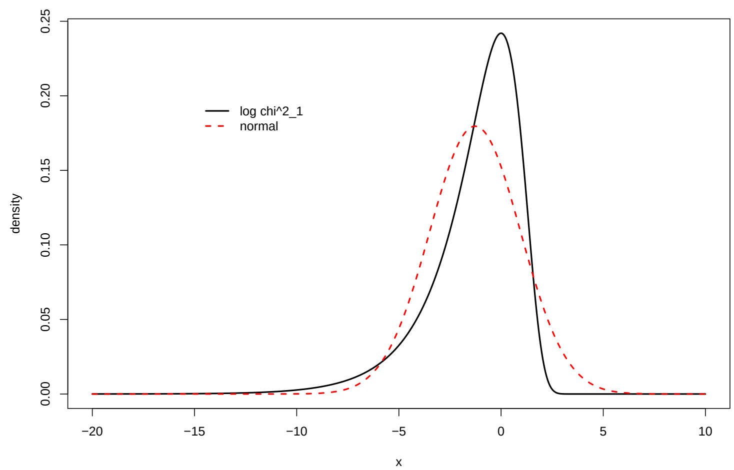
is a simple DLM where v_t and η_t are $N(0,1)$.

Sampling from

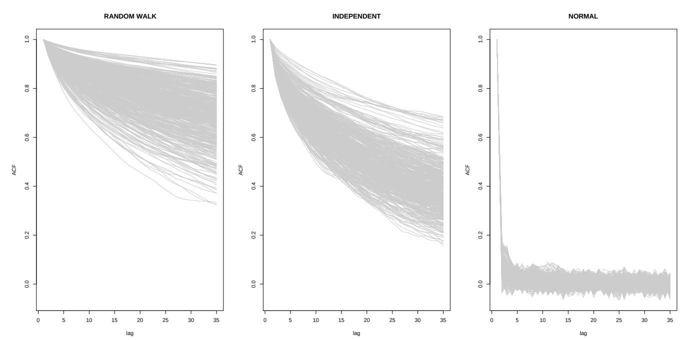
$$p(h^n|\theta, \tau^2, \sigma^2, z^n)$$

can be performed by the FFBS algorithm.

$\log\chi_1^2$ and $N(-1.27, \pi^2/2)$



ACF for the three schemes



Sampling h^n - mixtures of normals and FFBS

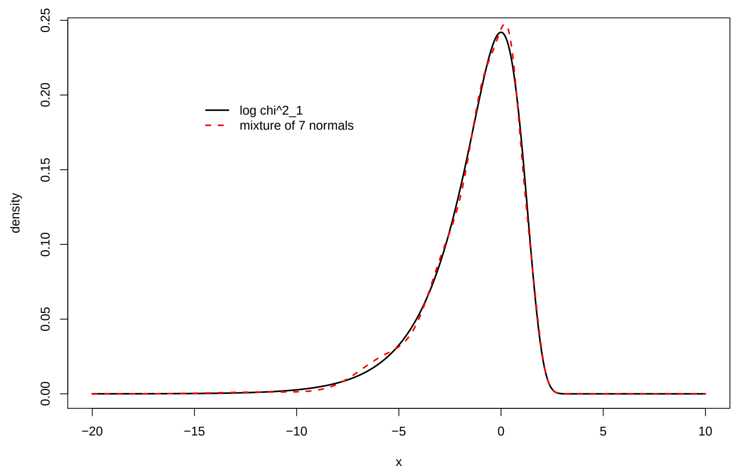
The $\log \chi_1^2$ distribution can be approximated by

$$\sum_{i=1}^{7} \pi_i N(\mu_i, \omega_i^2)$$

where

i	π_i	μ_i	ω_i^2
1	0.00730	-11.40039	5.79596
2	0.10556	-5.24321	2.61369
3	0.00002	-9.83726	5.17950
4	0.04395	1.50746	0.16735
5	0.34001	-0.65098	0.64009
6	0.24566	0.52478	0.34023
7	0.25750	-2.35859	1.26261

$\log \chi_1^2$ and $\sum_{i=1}^{7} \pi_i N(\mu_i, \omega_i^2)$



Mixture of normals

Using an argument from the Bayesian analysis of mixture of normal, let $z_1, \ldots, z_n$ be unobservable (latent) indicator variables such that $z_t \in \{1, \ldots, 7\}$ and $Pr(z_t = i) = \pi_i$, for $i = 1, \ldots, 7$.

Therefore, conditional on the z's, y_t is transformed into $\log y_t^2$,

$$\begin{aligned}
\log y_t^2 &= h_t + \log \varepsilon_t^2 \\
h_t &= \mu + \phi h_{t-1} + \tau_\eta \eta_t
\end{aligned}$$

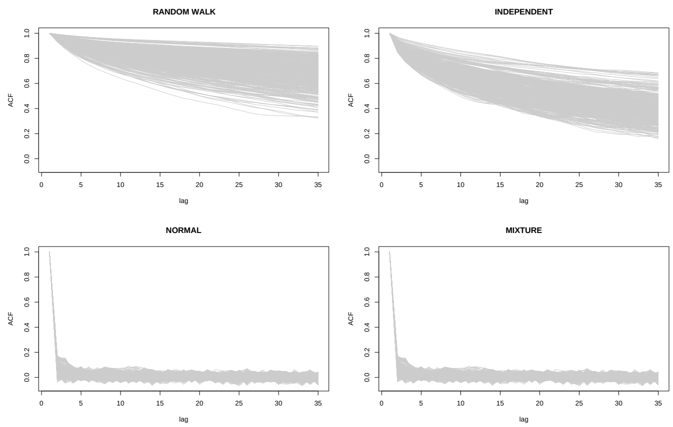
which can be rewritten as a normal DLM:

$$\begin{aligned}
\log y_t^2 &= h_t + v_t \qquad\qquad v_t \sim N(\mu_{z_t}, \omega_{z_t}^2) \\
h_t &= \mu + \phi h_{t-1} + w_t \qquad w_t \sim N(0, \tau_\eta^2)
\end{aligned}$$

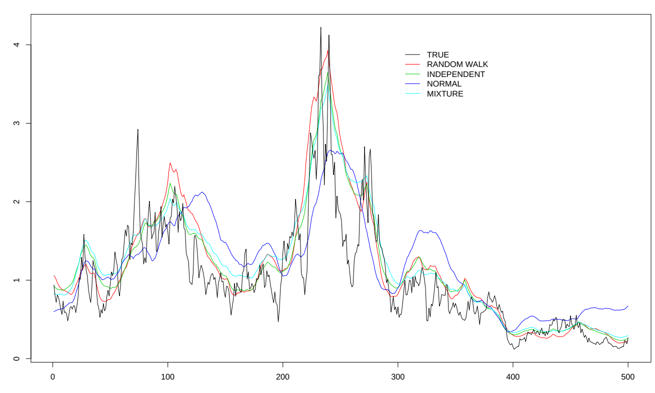
where μ_{z_t} and $\omega_{z_t}^2$ are provided in the previous table.

Then h^n is jointly sampled by using the the FFBS algorithm.

ACF for the four schemes



Posterior means: volatilities



Chapter 4

Bayesian VAR

4.1 VAR

VAR at a glance

Del Negro and Schorfheide (2011) says

"VARs appear to be straightforward multivariate generalizations of univariate autoregressive models. They turn out to be one of the key empirical tools in modern macroeconomics.

Sims (1980) proposed that VARs should replace large-scale macroeconometric models inherited from the 1960s, because the latter imposed incredible restrictions, which were largely inconsistent with the notion that economic agents take the effect of todays choices on tomorrows utility into account.

VARs have been used for macroeconomic forecasting and policy analysis to investigate the sources of business-cycle fluctuations and to provide a benchmark against which modern dynamic macroeconomic theories can be evaluated."

4.1.1 Model set up

Model set up

Let $y_t = (y_{t1}, \ldots, y_{tq})'$ contain q (macroeconomic) time series observed at time t.

The (basic) VAR(p) can be written as

$$y_t = B_1 y_{t-1} + \ldots + B_p y_{t-p} + u_t$$

where

$$u_t \sim \text{i.i.d. } N(0, \Sigma)$$

and

- $B_1, \ldots, B_p$ are $(q \times q)$ autoregressive matrices
- Σ is an $(q \times q)$ variance-covariance matrix

Multivariate regression

More compactly,

$$y_t = Bx_t + u_t \qquad u_t \sim N(0, \Sigma),$$

where

- $B = (B_1, \ldots, B_p)$ is the $(q \times qp)$ autoregressive matrix,
- $x_t = (y'_{t-1}, \ldots, y'_{t-p})'$ is qp-dimensional.

Matrix normal

Let

- $Y = (y_1, \ldots, y_n)'$ is $n \times q$,
- $X = (x_1, \ldots, x_n)'$ is $n \times qp$,
- $U = (y_1, \ldots, u_n)'$ is $n \times q$.

Then

$$Y = XB' + U \qquad U \sim N(0, I_n, \Sigma).$$

Multivariate normal

Let $y_i = (y_{1i}, \ldots, y_{ni})'$ be the n-dimensional vector with the observations for time series i, for $i = 1, \ldots, q$.

Let $y = \text{vec}(Y) = (y'_1, y'_2, \ldots, y'_q)'$.

Similarly, $\beta = \text{vec}(B')$ and $u = \text{vec}(U)$.

Then

$$y = (I_q \otimes X)\beta + u \qquad u \sim N(0, \Sigma \otimes I_n).$$

4.1.2 Forecasting

h-step ahead forecast

$$\begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \end{pmatrix} = \begin{pmatrix} B_1 & B_2 & \cdots & B_{p-1} & B_p \\ I_q & 0 & \cdots & 0 & 0 \\ 0 & I_q & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_q & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} u_t \\ u_{t-1} \\ u_{t-2} \\ \vdots \\ u_{t-p+1} \end{pmatrix}$$

or

$$y_t^* = Ay_{t-1}^* + u_t^*.$$

Therefore, the h-step ahead forecast is

$$y_t(h) = \int F(A, y_t^*, h)p(B, \Sigma|\text{data})d(B, \Sigma),$$

where the forecast function

$$F(A, y_t^*, h) = A^h y_t^*$$

is a highly nonlinear function of $B_1, \ldots, B_p$.

4.1.3 Stationarity

Stationarity

A VAR(p) is covariance-stationary if all values of z satisfying

$$|I_q - B_1 z - B_2 z^2 - \cdots - B_p z^p| = 0$$

lie outside the unit circle.

This is equivalent to all eigenvalues of A lying inside the unit circle.

4.1.4 VMA

Vector MA(∞) representation

If all eigenvalues of A lie inside the unit circle, then

$$y_t = \sum_{i=0}^{\infty} \Psi_i u_{t-i},$$

with

$$\begin{aligned} \Psi_0 &= I_q \\ \Psi_s &= B_1\Psi_{s-1} + B_2\Psi_{s-2} + \cdots + B_p\Psi_{s-p} \quad \text{for } s = 1, 2, \ldots \\ \Psi_s &= 0 \qquad \text{for } s < 0 \end{aligned}$$

4.1.5 VD

Variance Decomposition

The mean square error (MSE) of the h-step ahead forecast is

$$\Sigma + \Psi_1 \Sigma \Psi_1' + \cdots + \Psi_{h-1} \Sigma \Psi_{h-1}'.$$

The error $u_t \sim N(0, \Sigma)$ can be orthogonalized by

$$\varepsilon_t = A^{-1} u_t \sim N(0, D)$$

where $\Sigma = ADA'$ and D is diagonal (for instance, via singular value decomposition or Cholesky decomposition).

The MSE of the h-step ahead forecast can be rewritten as The conbribution of the jth orthogonalized innovation to the MSE is

$$d_j (a_j a_j' + \Psi_1 a_j a_j' \Psi_1' + \cdots + \Psi_{h-1} a_j a_j' \Psi_{h-1}')$$

4.1.6 IRF

Impulse-response function

The matrix Ψ_s has the interpretation

$$\frac{\partial y_{t+s}}{\partial u_s'} = \Psi_s,$$

that is, the (i, j) element of Ψ_s identifies the consequences of a one-unit increase in the innovation of variable j at time t for the value of variable i at time $t + s$, holding all other innovations at all dates constant.

A plot of the (i, j) element of Ψ_s,

$$\frac{\partial y_{t+s,i}}{\partial u_{s,i}},$$

as a function of s is called the *impulse-response function.*

Similar to the forecast function, the impulse-response function is also highly nonlinear on $B_1, \ldots, B_q$.

4.1.7 Examples

Simulated example[1]

Simulating 1000 observations from a trivariate VAR(2)

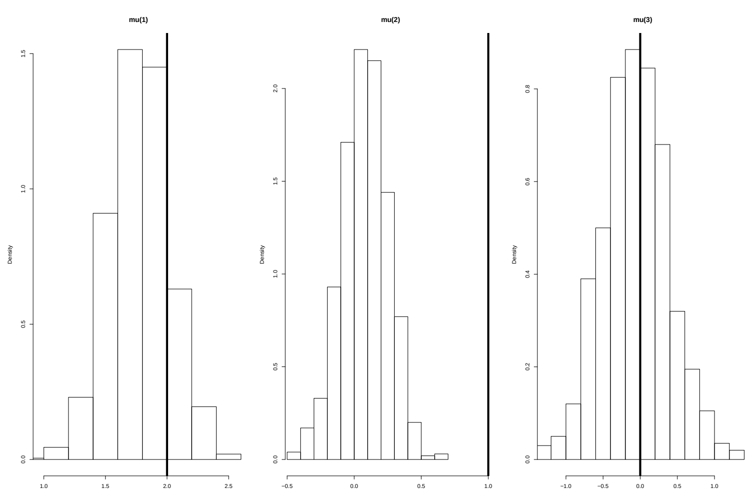
$$\mu = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \quad B_1 = \begin{pmatrix} 0.7 & 0.1 & 0.0 \\ 0.0 & 0.4 & 0.1 \\ 0.9 & 0.0 & 0.8 \end{pmatrix} \quad B_2 = \begin{pmatrix} -0.2 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.1 \\ 0.0 & 0.0 & 0.0 \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} 0.26 & 0.03 & 0.00 \\ 0.03 & 0.09 & 0.00 \\ 0.00 & 0.00 & 0.81 \end{pmatrix}.$$

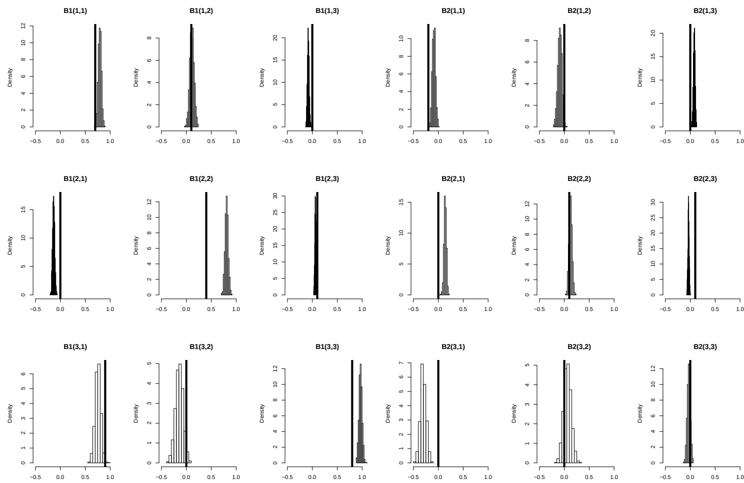
Posterior summaries based on 1000 draws.

$p(\mu|\mathbf{data})$

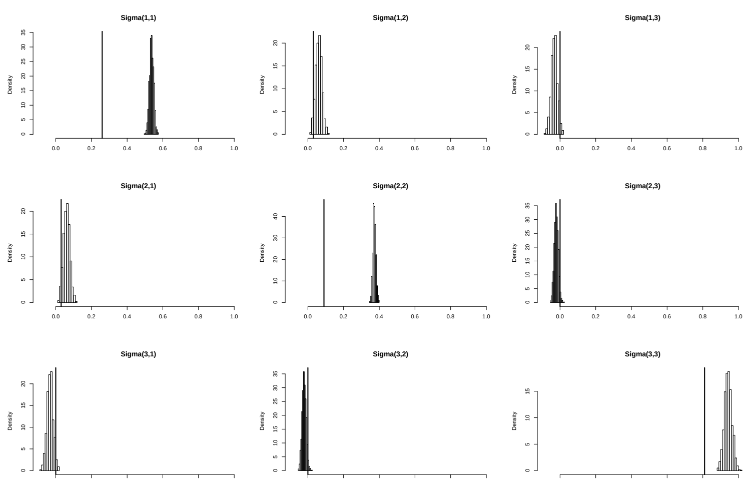


$p(B_i|\mathbf{data})$

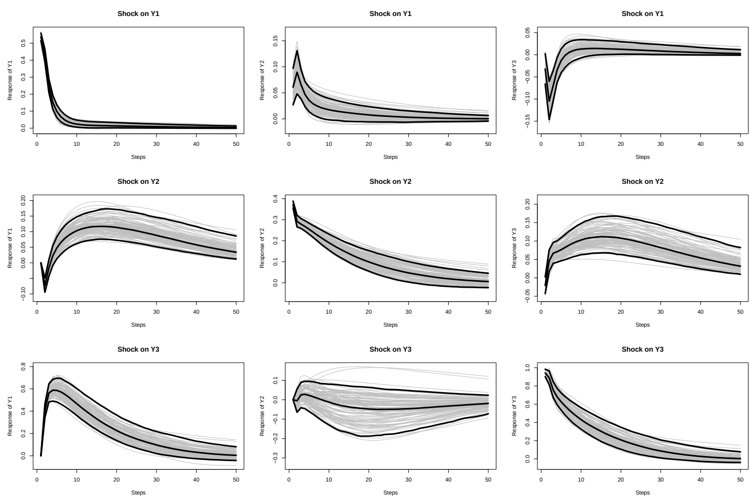
[1]Based on Lütkepohl's (2007) Problem 2.3 and using Sims' code (later slides).



$p(\Sigma|\mathbf{data})$

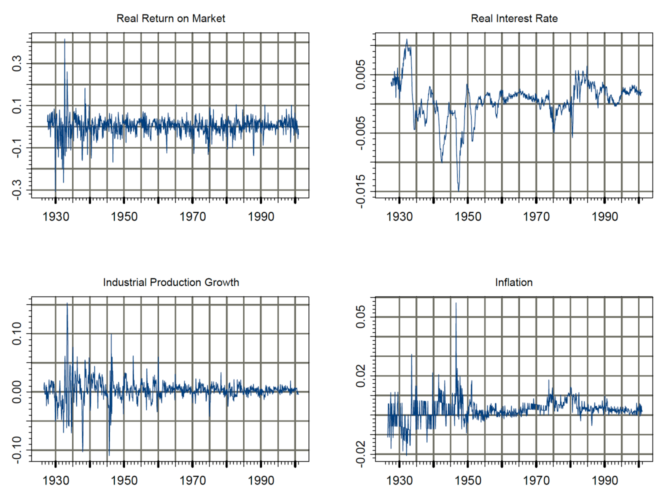


Impulse response



Real data example[2]

Monthly real stock returns, real interest rates, real industrial production growth and the inflation rate (1947.1 – 1987.12)



Posterior means

---

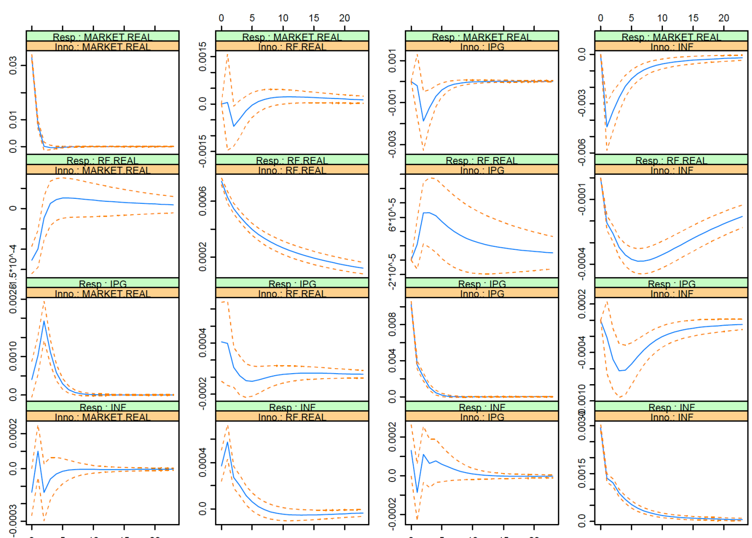
[2]Section 11.5 of Zivot and Wang (2003).

$$E(\mu|\text{data}) = \begin{pmatrix} 0.0074 & 0.0002 & 0.0010 & 0.0019 \end{pmatrix}'$$

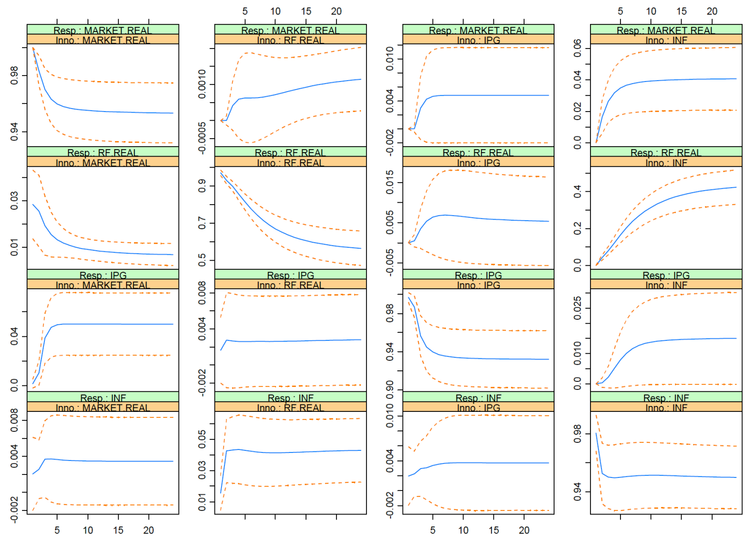
$$E(B_1|\text{data}) = \begin{pmatrix} 0.24 & 0.81 & -1.50 & 0.00 \\ 0.00 & 0.88 & -0.71 & 0.00 \\ 0.01 & 0.06 & 0.46 & -0.01 \\ 0.03 & 0.38 & -0.07 & 0.35 \end{pmatrix}$$

$$E(B_2|\text{data}) = \begin{pmatrix} -0.05 & -0.35 & -0.06 & -0.19 \\ 0.00 & 0.04 & 0.01 & 0.00 \\ -0.01 & -0.59 & 0.25 & 0.02 \\ 0.04 & -0.33 & -0.04 & 0.09 \end{pmatrix}$$

Impulse-response functions



Variance decomposition



4.1.8 Minnesota prior

Prior shrinkage

Recall that $\beta = vec(B')$.

Let β_i be column i of B': $B' = (\beta_1, \ldots, \beta_q)$.

Therefore,

$$\beta = (\beta_1', \ldots, \beta_q')'$$

with β_i a vector of dimension qp corresponding to all autoregressive coefficients from equation i.

Example: $q = 3$, $p = 2$

$$B = \left(\begin{array}{ccc|ccc} b_{1,11} & b_{1,12} & b_{1,13} & b_{2,11} & b_{2,12} & b_{2,13} \\ b_{1,21} & b_{1,22} & b_{1,23} & b_{2,21} & b_{2,22} & b_{2,23} \\ b_{1,31} & b_{1,32} & b_{1,33} & b_{2,31} & b_{2,32} & b_{2,33} \end{array} \right) = \left(\begin{array}{c} \beta_1' \\ \beta_2' \\ \beta_3' \end{array} \right)$$

Minnesota prior

Litterman's (1980,1986) Minnesota prior advocates that

$$\beta_i \sim N(\beta_{i0}, V_{i0})$$

with V_{i0} chosen to "center" the individual equations around the random walk with drift:

$$y_{ti} = \mu_i + y_{t-1,i} + u_{ti}.$$

This amounts to:

- Shrinking the diagonal elements of B_1 toward one,
- Shrinking the remaining coefficients of $B_1, \ldots, B_p$ toward zero,

In addition:

- Shrinking the number of lags p towards one,
- Own lags should explain more of the variation than the lags of other variables in the equation.

Specifying V_{i0}
V_{i0} is diagonal with elements

$$\frac{\theta^2 \lambda^2}{k^2} \times \frac{\sigma_i^2}{\sigma_j^2}$$

for coefficients on lags of variable $j \neq i$, and

$$\frac{\lambda^2}{k^2}$$

for coefficients on own lags.

The error matrix Σ is assumed to be diagonal, fixed and known

$$\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_q^2).$$

4.1.9 Other priors

Other priors on (β, Σ)
Kadiyala and Karlsson (1993,1997) extend the Minnesota prior:

- Normal-Wishart prior

$$N(\beta_0, \Sigma \otimes \Omega) IW(\Sigma_0, \alpha)$$

- Jeffreys' prior

$$p(\beta, \Sigma) \propto |\Sigma|^{-(q+1)/2}$$

- Normal-Diffuse prior

$$\beta \sim N(\beta_0, V_0) \quad p(\Sigma) \propto |\Sigma|^{-(q+1)/2}$$

- Extended Natural Conjugate (ENC) prior

Kadiyala and Karlsson (1993) use MC integration. Kadiyala and Karlsson (1997) use Gibbs sampler.

	Prior	Posterior
Minnesota	$\gamma_i \sim N(\tilde{\gamma}_i, \tilde{\Sigma}_i)$, Ψ fix and diagonal	$\gamma_i \mid \mathbf{y} \sim N(\bar{\gamma}_i, \bar{\Sigma}_i)$
Diffuse	$p(\gamma, \Psi) \propto \lvert\Psi\rvert^{-(m+1)/2}$	$\Gamma \mid \mathbf{y} \sim MT(\mathbf{Z}'\mathbf{Z}, (\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'$ $\times(\mathbf{Y} - \mathbf{Z}\hat{\Gamma}), \hat{\Gamma}, T - k)$
Normal-Wishart	$\gamma \mid \Psi \sim N(\tilde{\gamma}, \Psi \otimes \tilde{\Omega})$, $\Psi \sim iW(\tilde{\Psi}, \alpha)$	$\Gamma \mid \mathbf{y} \sim MT(\bar{\Omega}^{-1}, \bar{\Psi}, \bar{\Gamma}, T + \alpha)$
Normal-Diffuse	$\gamma \sim N(\tilde{\gamma}, \tilde{\Sigma})$, $p(\Psi) \propto \lvert\Psi\rvert^{-(m+1)/2}$	$p(\gamma \mid y) \propto \exp\{-(\gamma - \tilde{\gamma})'\tilde{\Sigma}^{-1}$ $\times(\gamma - \tilde{\gamma})/2\} \times \lvert(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})$ $+(\Gamma - \hat{\Gamma})'\mathbf{Z}'\mathbf{Z}(\Gamma - \hat{\Gamma})\rvert^{-T/2}$
Extended Natural Conjugate	$p(\Delta) \propto \lvert\tilde{\Psi} + (\Delta - \tilde{\Delta})'$ $\times\tilde{\mathbf{M}}(\Delta - \tilde{\Delta})\rvert^{-\alpha/2}$ or independent multivariate t's for each equation, $\Psi \mid \Delta \sim iW(\tilde{\Psi} + (\Delta - \tilde{\Delta})'$ $\times\tilde{\mathbf{M}}(\Delta - \tilde{\Delta}), \alpha)$	$p(\Delta \mid \mathbf{y}) \propto \lvert\bar{\Psi} + (\Delta - \bar{\Delta})'$ $\times\mathbf{M}(\Delta - \bar{\Delta})\rvert^{-(T+\alpha)/2}$ $\Psi \mid \Delta, \mathbf{y} \sim iW$ $(\tilde{\Psi} + (\Delta - \bar{\Delta})'\mathbf{M}(\Delta - \bar{\Delta}), T + \alpha)$

4.1.10 Example

Example

Kadiyala and Karlsson (1997) revisited Litterman (1986):

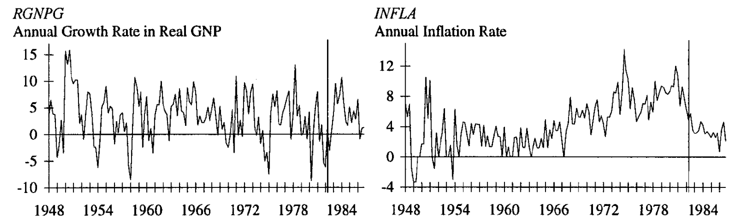
- Annual growth rates of real GNP (RGNPG),
- Annual inflation rates (INFLA),
- Unemployment rate (UNEMP)
- Logarithm of nominal money supply (M1)
- Logarithm of gross private domestic investment (INVEST),
- Interest rate on four- to six-month commercial paper (CPRATE)
- Change in business inventories (CBI).

Quarterly data from 1948:1 to 1981:1 (133 observations).

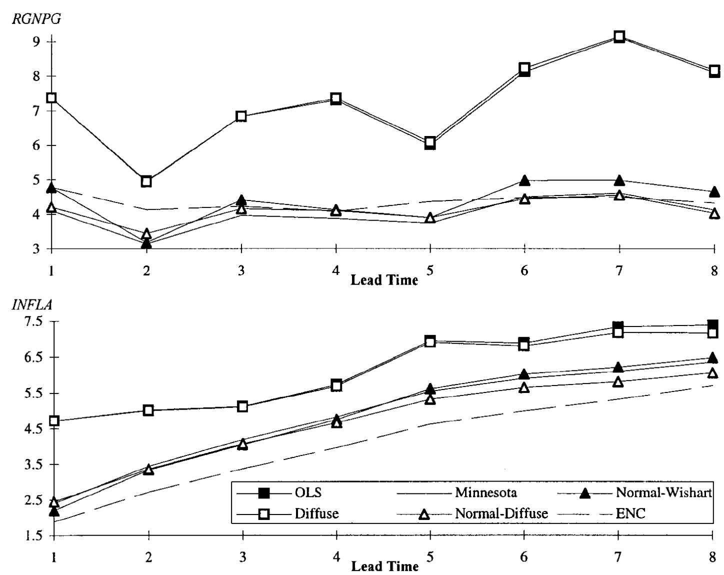
Out-of-sample forecast: 1980:2 to 1986:4

$q = 7$, $p = 6$ $\Rightarrow$ $7 + 6(7^2) = 301$ parameters.

Time series



Root mean square error



They argue that

> ... our preferred choice is the Normal-Wishart when the prior beliefs are of the Litterman type.

and

> For more general prior beliefs ... the Normal-Diffuse and EN priors are strong alternatives to the Normal-Wishart.

4.1.11 Structural VAR

Structural VAR

Rubio-Ramírez, Waggoner and Zha (2010) say that

> "Since the seminal work by Sims (1980), identification of structural vector autoregressions (SVARs) has been an unresolved theoretical issue.
>
> Filling this theoretical gap is of vital importance because impulse responses based on SVARs have been widely used for policy analysis and to provide stylized facts for dynamic stochastic general equilibrium (DSGE) models."

Example

Let

- $P_{c,t}$ is the price index of commodities.
- Y_t is output.
- R_t is the nominal short-term interest rate.

Trivariate SVAR(1) representation:

$$\begin{aligned}
a_{11}\Delta\log P_{c,t} + 0.0\log Y_t + a_{31}R_t &= c_1 + b_{11}\Delta\log P_{c,t-1} + b_{21}\Delta\log Y_{t-1} + b_{31}R_{t-1} + \varepsilon_{1,t} \\
a_{12}\Delta\log P_{c,t} + a_{22}\log Y_t + 0.0R_t &= c_2 + b_{12}\Delta\log P_{c,t-1} + b_{22}\Delta\log Y_{t-1} + b_{32}R_{t-1} + \varepsilon_{2,t} \\
a_{13}\Delta\log P_{c,t} + a_{23}\log Y_t + a_{33}R_t &= c_3 + b_{13}\Delta\log P_{c,t-1} + b_{23}\Delta\log Y_{t-1} + b_{33}R_{t-1} + \varepsilon_{3,t}
\end{aligned}$$

1st eq. monetary policy equation. 2nd eq. characterizes behaviour of finished-goods producers. 3rd eq. commodity prices are set in active competitive markets.

Model set up

The (basic) SVAR(p) can be written as

$$A_0 y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t \qquad u_t \sim \text{i.i.d. } N(0, I_q),$$

where

- $A = (A_1, \ldots, A_p)$
- $B_i = A_0^{-1}A_i \quad i = 1, \ldots, p$
- $B = A_0^{-1}A$
- $\Sigma = (A_0 A_0')^{-1}$

Much of the SVAR literature involves exactly identified models.

Exact identification

Define g such that $g(A_0, A) = (A_0^{-1}A, (A_0 A_0')^{-1})$.

Consider an SVAR with restrictions represented by R.

Definition: The SVAR is exactly identified if and only if, for almost any reduced-form parameter point (B, Σ), there exists a unique structural parameter point $(A_0, A) \in R$ such that $g(A_0, A) = (B, \Sigma)$.

Waggoner and Zha (2003) developed an efficient MCMC algorithm to generate draws from a restricted A_0 matrix.

Illustration 1

$$A_0 = \left(\begin{array}{cccccc} & PS & PS & MP & MD & Inf \\ \log Y & a_{11} & a_{12} & 0 & a_{14} & a_{15} \\ \log P & 0 & a_{22} & 0 & a_{14} & a_{25} \\ R & 0 & 0 & a_{33} & a_{34} & a_{35} \\ \log M & 0 & 0 & a_{43} & a_{44} & a_{45} \\ \log P_c & 0 & 0 & 0 & 0 & a_{15} \end{array}\right)$$

where

- $\log Y$: log gross domestic product (GDP)
- $\log P$: log GDP deflator
- R: nominal short-term interest rate
- $\log M$: log M3
- $\log P_c$: log commodity prices

and

- MP: monetary policy (central bank's contemporaneous behavior)
- Inf: commodity (information) market
- MD: money demand equation
- PS: production sector

Illustration 2

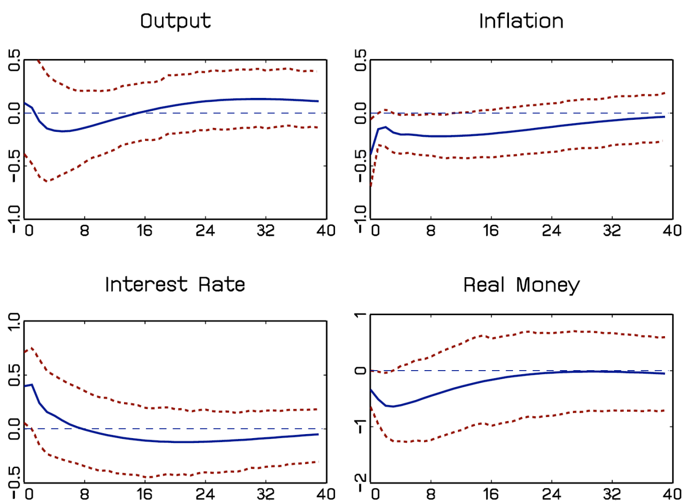
$$A_0 = \left(\begin{array}{ccccccc} & PCOM & M2 & R & Y & CPI & U \\ Inform & X & X & X & X & X & X \\ MP & 0 & X & X & 0 & 0 & 0 \\ MD & 0 & X & X & X & X & 0 \\ Prod & 0 & 0 & 0 & X & 0 & 0 \\ Prod & 0 & 0 & 0 & X & X & 0 \\ Prod & 0 & 0 & 0 & X & X & X \end{array}\right)$$

where

- PCOM: Price index for industrial commodities
- M2: Real money
- R: Federal funds rate (R)
- Y: real GDP interpolated to monthly frequency

- CPI: Consumer price index (CPI)
- U: Unemployment rate (U)
- Inform: Information market
- MP: Monetary policy rule
- MD: Money demand
- Prod: Production sector of the economy

Monetary Policy Shock



4.2 VAR-GARCH

VAR-GARCH

Pelloni and Polasek (2003) introduce the VAR model with GARCH errors as

$$y_t = \sum_{i=1}^{p} B_i y_{t-i} + u_t$$

where

$$u_t \sim N(0, \Sigma_t)$$

and

$$\text{vech}(\Sigma_t) = \alpha_0 + \sum_{i=1}^{r} A_i \text{vech}(\Sigma_{t-i}) + \sum_{i=1}^{s} \Theta_i \text{vech}(u_{t-i}u'_{t-i})$$

Example

German, U.S., and U.K. quarterly data sets over the period 1968-1998. Variables are logs of aggregate employment and of the employment shares of the manufacturing, finance, trade, and construction sectors for U.S. and U.K.

Table IVa. Bayes factors for model selection using posterior log-marginal likelihoods.

$\log BF_{21}$	Country		
	Germany	U.K.	U.S.
VAR	239.04	353.67	100.09
EC-VARCH	3.15	2.83	4.39
CEC-VARCH	11.79	6.71	4.79
COIN-VARCH 1	2.41	9.70	2.36
COIN-VARCH 2	6.76	1.31	10.11

4.3 VAR-SV

VAR-SV

Uhlig (1997) introduced stochastic volatility (SV) for the error term in BVARs:

$$y_t = \sum_{i=1}^{p} B_i y_{t-i} + u_t,$$

where

$$u_t \sim N(0, \Sigma_t) \quad \text{and} \quad \Sigma_t^{-1} = L_t L_t',$$

and dynamics

$$\begin{aligned} \Sigma_{t+1}^{-1} &= \frac{L_t \Theta_t L_t'}{\lambda} \\ \Theta_t &\sim B_q\left(\frac{\nu + pq}{2}, \frac{1}{2}\right). \end{aligned}$$

4.4 TVP-VAR-SV

TVP-VAR-SV

Primiceri (2005) discusses VARs with time varying coefficients and stochastic volatility

$$y_t = \sum_{i=1}^{p} B_{it} y_{t-i} + u_t \qquad u_t \sim N(0, \Sigma_t)$$

with

$$\Sigma_t = (A_t)^{-1} D_t (A_t')^{-1},$$

and

$$A_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{21,t} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \alpha_{q1,t} & \cdots & \alpha_{q,q-1,t} & 1 \end{pmatrix} \quad D_t = \begin{pmatrix} \sigma_{1,t}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{2,t}^2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_{q,t}^2 \end{pmatrix}.$$

Dynamics

VAR coefficients:

$$B_t = B_{t-1} + \nu_t \qquad \nu_t \sim N(0, Q)$$

Cholesky coefficients:

$$\alpha_t = \alpha_{t-1} + \xi_t \qquad \xi_t \sim N(0, S)$$

Stochastic volatility:

$$\log \sigma_t = \log \sigma_{t-1} + \eta_t \qquad \eta_t \sim N(0, W)$$

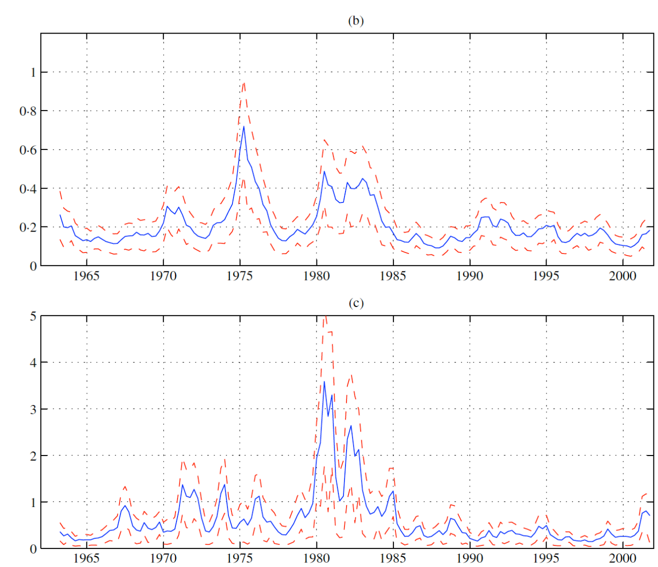


FIGURE 1

Posterior mean, 16-th and 84-th percentiles of the standard deviation of (a) residuals of the inflation equation, (b) residuals of the unemployment equation and (c) residuals of the interest rate equation or monetary policy shocks

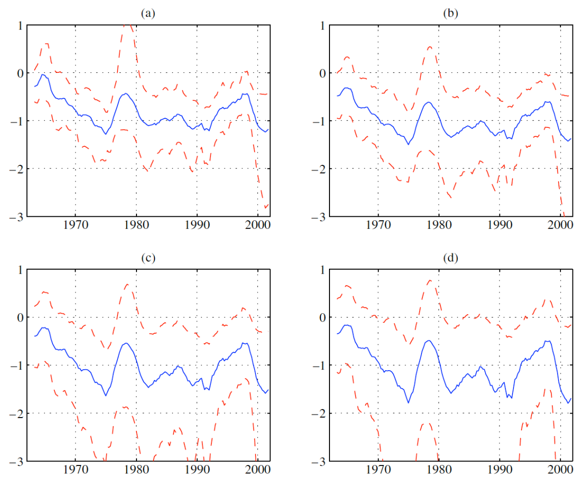


FIGURE 6

Interest rate response to a 1% permanent increase of unemployment with 16-th and 84-th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters

See Nakajima, Kasuya and Watanabe (2011) for an application to the Japanese economy.

4.5 Dimensionality

Curse of dimensionality

VAR(1) case[3]

Small: $q = 3 \;\Rightarrow$ 15 parameters

Medium: $q = 20 \;\Rightarrow$ 610 parameters

Large: $q = 131 \;\Rightarrow$ 25,807 parameters

References

Ahmadi and Uhlig (2009) Measuring the Dynamic Effects of Monetary Policy Shocks: A Bayesian FAVAR Approach with Sign Restriction. Technical Report.

Bańbura, Giannone and Reichlin (2010) Large Bayesian vector auto regressions. JAE, 25, 71-92.

Brandt and Freeman (2006) Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting, and Policy Analysis. *Political Analysis*, 14, 1-36.

Carriero, Kapetanios, Marcellino (2009) Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25, 400-417.

Del Negro and Schorfheide (2011) Bayesian Macroeconometrics. In *Handbook of Bayesian Econometrics*, Chapter 7, 293-387. Oxford University Press.

Doan, Litterman and Sims (1984) Forecasting and conditional projection using realistic prior distributions (with discussion). *Econometric Reviews*, 3, 1-144.

[3] Small, Medium and Large are based on the VAR specifications of Bańura, Giannone and Reichlin (2010).

Garratt, Koop, Mise and Vahey (2009) Real-Time Prediction With U.K. Monetary Aggregates in the Presence of Model Uncertainty. *Journal of Business & Economic Statistics*, 27, 480-491.

Jochmann, Koop, Leon-Gonzalez and Strachan (2011) Stochastic search variable selection in vector error correction models with application to a model of the UK macroeconomy. *JAE*.

Jochmann, Koop and Strachan (2010) Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks. *International Journal of Forecasting*, 26, 326-347.

Kadiyala and Karlsson (1993) Forecasting with generalized Bayesian vector autoregressions. *Journal of Forecasting*, 12, 365-78.

Kadiyala and Karlsson (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12, 99-132.

Refences (cont.)

Koop and Korobilis (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3, 267-358.

Koop and Korobilis (2012) Large Time-Varying Parameter VARs. Technical Report.

Korobilis (2008) Forecasting in vector autoregressions with many predictors. Technical Report, University of Strathclyde.

Korobilis (2012) VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, forthcoming.

Litterman (1980) A Bayesian procedure for forecasting with vector autoregressions. Technical Report, MIT.

Litterman (1986) Forecasting with Bayesian vector autoregressions – five years of experience. *Journal of Business & Economic Statistics*, 4, 25-38.

Lütkepohl (2007) *New Introduction to Multiple Time Series Analysis*. Springer.

Nakajima, Kasuya and Watanabe (2011) Bayesian analysis of time-varying parameter vector autoregressive model for the Japanese economy and monetary policy. *Journal of The Japanese and International Economies*, 25, 225-245.

Pelloni and Polasek (2003) Macroeconomic Effects of Sectoral Shocks in Germany, the U.K. and, the U.S.: A VAR-GARCH-M Approach. *Computational Economics*, 21, 65-85.

Pfaff (2008) VAR, SVAR and SVEC Models: Implementation within `R` Package `vars`. *Journal of Statistical Software*, 27, 1-32.

Primiceri (2005). Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies*, 72, 821-852.

Refences (cont.)

Rubio-Ramírez, Waggoner and Zha (2010) Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies*, 77, 665-696.

Sims (1980) Macroeconomics and Reality. *Econometrica*, 48, 1-48.

Sims (2007) Bayesian Methods in Applied Econometrics, or, Why Econometrics Should Always and Everywhere Be Bayesian. Technical Report, Princeton University.

Sims and Zha (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review*, 39, 949-968.

Stock and Watson (2005) Implications of dynamic factor models for VAR analysis. Technical Report.

Uhlig (1997) Bayesian Vector Autoregressions with Stochastic Volatility. *Econometrica*, 65, 59-73.

Villani (2009) Steady-state priors for vector autoregressions. *Journal of Applied Econometrics*, 24, 630-650.

Waggoner and Zha (2003) A Gibbs Sampler for Structural VARs. *Journal of Economic Dynamics and Control*, 28, 349-366.

Zha (1999) Block recursion and structural vector autoregressions. *Journal of Econometrics*, 90, 291-316.

Zivot and Wang (2003) *Modeling Financial Time Series with S-plus*. Springer.

Chapter 5

BVAR Extensions

5.1 Variable selection

Why variable selection in VAR?

VAR flexibility $\Rightarrow$ over-parameterization $\Rightarrow$ poor forecasts.

First solution: Minnesota prior of Doan *et al.* (1984).

Additional (more recent) solutions:

- Factor models (Stock and Watson, 2006)
- Variable selection priors (George *et al.*, 2008)
- Steady-state priors (Villani, 2009)
- Bayesian model averaging (Garratt *et al.*, 2009)
- Combination forecasts (Clark and McCracken, 2010)

Stochastic search

According to Korobilis (2011) "stochastic search"

> *... simply means that if the model space is too large to assess in a deterministic manner (i.e. enumerate and estimate all possible models, and decide on the best one using some goodness-of-fit measure), the algorithm will visit only the most probable models in a stochastic manner.*

5.1.1 ON/OFF switch

ON/OFF switch

Recall the basic VAR(p) model

$$y_t = B_1 y_{t-1} + \ldots + B_p y_{t-p} + u_t$$

where $u_t \sim$ i.i.d. $N(0, \Sigma)$ and $B = (B_1, \ldots, B_p)$.

Bayesian variable selection idea

$$B_{ij} = \begin{cases} B_{ij} = 0 & \gamma_{ij} = 0 \\ B_{ij} \neq 0 & \gamma_{ij} = 1 \end{cases}$$

Constrained VAR

The VAR(p) model can be rewritten as a SUR model as

$$y_t = z_t \beta + u_t$$

where $z_t = I_q \otimes (y'_{t-1}, \ldots, y'_{t-p})$ is $(q \times s)$, $\beta = \text{vec}(B)$ is $(s \times 1)$ and $s = q^2 p$.

Including the variable selection indicators:

$$y_t = z_t \Gamma \beta + u_t$$

where $\Gamma = \text{diag}(\gamma_{11}, \ldots, \gamma_{1,pq}, \ldots, \gamma_{q1}, \ldots, \gamma_{q,pq})$.

2^s VAR models $\Rightarrow$ variable selection = model selection.

Prior specification

VAR coefficients

$$\beta \sim N(\beta_0, V_0)$$

VAR error variance

$$p(\Sigma) \propto |\Sigma|^{(q+1)/2}$$

Variable selection indicators

$$\gamma_j | \gamma_{-j} \sim Bernoulli(\pi_{0j})$$

Full conditional of β

Let $\tilde{z}_t = z_t \Gamma$. Then,

$$p(\beta|\Sigma, \gamma, y) \propto \exp\left\{-0.5\left[\beta'(V_0 + A_1)^{-1}\beta - 2\beta'(V_0^{-1}\beta_0 + A_2)\right]\right\}$$

where

$$A_1 = \sum_{t=1}^{T} \tilde{z}_t' \Sigma^{-1} \tilde{z}_t \quad \text{and} \quad A_2 = \sum_{t=1}^{T} \tilde{z}_t' \Sigma^{-1} y_t.$$

Therefore,

$$\beta|\Sigma, \gamma, y \sim N(\beta_1, V_1)$$

where

$$\beta_1 = V_1(V_0^{-1}\beta_0 + A_2) \quad \text{and} \quad V_1^{-1} = V_0^{-1} + A_1$$

Full conditional of Σ

$$p(\Sigma|\beta, y) \propto |\Sigma|^{\frac{q+T+1}{2}} \exp\left\{-0.5 \sum_{t=1}^{T} (y_t - z_t\beta)' \Sigma^{-1} (y_t - z_t\beta)\right\}$$

or

$$\Sigma|\beta, y \sim IW(T, S)$$

where

$$S = \sum_{t=1}^{T} (y_t - z_t\beta)(y_t - z_t\beta)'$$

Full conditional of γ

It is easy to see that

$$p(\gamma_j = 1|\gamma_{-j}, \beta, \Sigma, y) \propto p(y|\theta_j, \Sigma, \gamma_{-j}, \gamma_j = 1)\pi_{0j} = A_j$$

and

$$p(\gamma_j = 0|\gamma_{-j}, \beta, \Sigma, y) \propto p(y|\theta_j, \Sigma, \gamma_{-j}, \gamma_j = 1)(1 - \pi_{0j}) = B_j$$

Therefore

$$(\gamma_j|\gamma_{-j}, \beta, \Sigma, y) \sim \text{Bernoulli}\left(\frac{A_j}{A_j + B_j}\right)$$

5.1.2 SSVS

SSVS

Jochmann *et al.* (2011) adapts George and McCulloch's (1993, 1997) stochastic search variable selection (SSVS) method to BVAR with structural breaks.

The spike-slab prior for γ_j is given by

$$\beta_j|\gamma_j \sim (1-\gamma_j)N(0,\kappa_{0j}^2) + \gamma_j N(0,\kappa_{1j}^2).$$

SSVS aspect of the prior small κ_{0j}^2 (coefficient is virtually zero) large κ_{1j}^2 (relatively non-informative prior).

Indicator variable

$$\gamma_i \sim \text{Bernoulli}(q_j)$$

Default semi-automatic approach

George *et al.* (2008) suggests

$$\begin{aligned}\kappa_{0j} &= c_0\sqrt{\hat{\text{var}}(\beta_j)}\\ \kappa_{1j} &= c_1\sqrt{\hat{\text{var}}(\beta_j)},\end{aligned}$$

where $c_0 << c_1$.

$\hat{\text{var}}(\beta_j)$ is an estimate of the variance of the coefficient in an unrestricted VAR (e.g., the ordinary least squares quantity or an estimate based on a preliminary MCMC run of the VAR using a non-informative prior). Jochmann *et al.* (2011) set $c_0 = 0.1$ and $c_1 = 10$.

George (2000) reviews the variable selection problem.

Prior on Σ

Let Ψ be the lower-triangular Cholesky root of Σ^{-1}, i.e.

$$\Sigma^{-1} = \Psi\Psi'.$$

Diagonal components

$$\psi_{jj}^2 \sim G(a_j, b_j)$$

Off-diagonal components

$$\begin{aligned}\psi_{ij} &\sim (1-\omega_{ij})N(0,\kappa_{0ij}^2)+\omega_{ij}N(0,\kappa_{1ij}^2)\\ \omega_{ij} &\sim \text{Bernoulli}(q_{ij})\end{aligned}$$

Substantial progress has been made on Bayesian modeling by imposing restrictions on the Σ matrix (Pourahmadi, 1999, Sims and Zha, 1998, Smith and Kohn, 2002, Waggoner and Zha, 2003, Lopes, McCulloch and Tsay, 2012, Lopes and Polson, 2012).

5.2 Factor analysis

Bayesian factor analysis (BFA)[1]

- Pre-MCMC era
  - Press (1972)
  - Martin and McDonald (1975)
  - Bartholomew (1981, 1987)
  - Press and Shigemasu (1989)
- MCMC era
  - Geweke and Zhou (1996)
  - Aguilar and West (2000)
  - Lopes (2000)
  - Lopes and West (2004)
  - West (2003)
  - Lopes and Carvalho (2007)
  - Lopes, Salazar and Gamerman (2008)

5.2.1 Basic model

Basic model

For any specified positive integer $k \leq m$, the standard $k-$factor model relates each y_t to an underlying $k-$vector of random variables f_t, the common factors, via

$$\begin{aligned}y_t|f_t &\sim N(\beta f_t,\Sigma)\\ f_t &\sim N(0,I_k)\end{aligned}$$

where $\Sigma = \text{diag}(\sigma_1^2,\cdots,\sigma_m^2)$. Therefore,

[1] Lopes (2003) Factor models. *ISBA Bulletin*, 10(2), 7-10.

- $\text{var}(y_{it}|f) = \sigma_i^2$,
- $\text{cov}(y_{it}, y_{jt}|f) = 0$,
- $V(y_t|\beta, \Sigma) = \Omega = \beta\beta' + \Sigma$
- $\text{var}(y_{it}) = \sum_{l=1}^{k} \beta_{il}^2 + \sigma_i^2$,
- $\text{cov}(y_{it}, y_{jt}) = \sum_{l=1}^{k} \beta_{il}\beta_{jl}$.

Common factors explain all the dependence structure among the m variables.

Invariance

The model is invariant under transformations of the form $\beta^* = \beta P'$ and $f_t^* = P f_t$, where P is any orthogonal $k \times k$ matrix.

Classical approach: $\beta' \Sigma^{-1} \beta = I$.

Our approach: β is a block lower triangular.

$$\beta = \begin{pmatrix} \beta_{11} & 0 & 0 & \cdots & 0 \\ \beta_{21} & \beta_{22} & 0 & \cdots & 0 \\ \beta_{31} & \beta_{32} & \beta_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \beta_{k3} & \cdots & \beta_{kk} \\ \beta_{k+1,1} & \beta_{k+1,2} & \beta_{k+1,3} & \cdots & \beta_{k+1,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{m1} & \beta_{m2} & \beta_{m3} & \cdots & \beta_{mk} \end{pmatrix}$$

This form provides both identification and, often, useful interpretation of the factor model.

Number of parameters

The resulting factor form of Ω has

$$m(k+1) - k(k-1)/2$$

parameters, compared with the total

$$m(m+1)/2$$

in an unconstrained (or $k = m$) model, leading to the constraint that

$$m(m+1)/2 - m(k+1) + k(k-1)/2 \geq 0,$$

which provides an upper bound on k.

For example,

- $m = 6$ implies $k \leq 3$,

- $m = 12$ implies $k \leq 7$,
- $m = 20$ implies $k \leq 14$,
- $m = 50$ implies $k \leq 40$,

Even for small m, the bound will often not matter as relevant k values will not be so large.

Full-rank loading matrix

Geweke and Singleton (1980) show that, if β has rank $r < k$ there exists a matrix Q such that $\beta Q = 0$ and $Q'Q = I$ and, for any orthogonal matrix M,

$$\beta\beta' + \Sigma = (\beta + MQ')'(\beta + MQ') + (\Sigma - MM')$$

This translation invariance of Ω under the factor model implies lack of identification and, in application, induces symmetries and potential multimodalities in resulting likelihood functions.

This issue relates intimately to the question of uncertainty of the number of factors.

Ordering of the variables

Alternative orderings are trivially produced via Ay_t for some rotation matrix A.

The new rotation has the same latent factors but transformed loadings matrix $A\beta$.

$$Ay_t = A\beta f + \varepsilon_t$$

This new loadings matrix does not have the lower triangular structure.

However, we can always find an orthonormal matrix P such that $A\beta P'$ is lower triangular, and so simply recover the factor model with the same probability structure for the underlying latent factors Pf_t (Lopes and West, 2004).

The order of the variables in y_t is irrelevant assuming that k is properly chosen.

Prior specification

Loading matrix:

$$\begin{aligned} \beta_{ij} &\sim N(0, C_0) && \text{when} \quad i \neq j, \\ \beta_{ii} &\sim N(0, C_0)1(\beta_{ii} > 0) && \text{when} \quad i = 1, \ldots, k \end{aligned}$$

Idiosyncratic variances

$$\sigma_i^2 \sim IG(\nu/2, \nu s^2/2)$$

where s^2 is the prior mode of each σ_i^2 and ν the prior degrees of freedom hyper-parameter.

We eschew the use of standard improper reference priors $p(\sigma_i^2) \propto 1/\sigma_i^2$, since such priors lead to the Bayesian analogue of the so-called *Heywood problem* (Martin and McDonald, 1975, and Ihara and Kano, 1995).

Full conditional distributions

Factor scores

$$f_t \sim N(V_f \beta' \Sigma^{-1} y_t, V_f)$$

where $V_f = (I_k + \beta' \Sigma^{-1} \beta)^{-1}$.

Idiosyncrasies

$$\sigma_i^2 \sim IG((\nu + T)/2, (\nu s^2 + d_i)/2)$$

where $d_i = (y_i - f\beta_i')'(y_i - f\beta_i')$.

First k rows of β

$$\beta_i \sim N(M_i, C_i) 1(\beta_{ii} > 0)$$

where

$$\begin{aligned} M_i &= C_i \left(C_0^{-1} \mu_0 1_i + \sigma_i^{-2} f_i' y_i \right) \\ C_i^{-1} &= C_0^{-1} I_i + \sigma_i^{-2} f_i' f_i. \end{aligned}$$

Last $m - k$ rows of β

$$\beta_i \sim N(M_i, C_i)$$

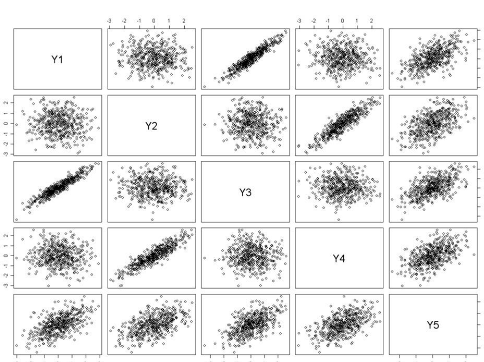
where

$$\begin{aligned} M_i &= C_i \left(C_0^{-1} \mu_0 1_k + \sigma_i^{-2} f' y_i \right) \\ C_i^{-1} &= C_0^{-1} I_k + \sigma_i^{-2} f' f. \end{aligned}$$

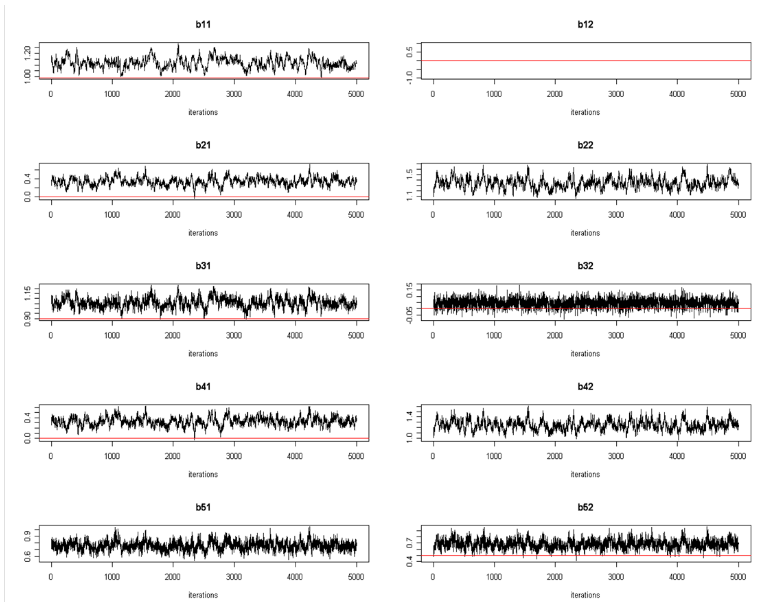
Simulation exercise

$T = 500$, $p = 5$ and $k = 2$.

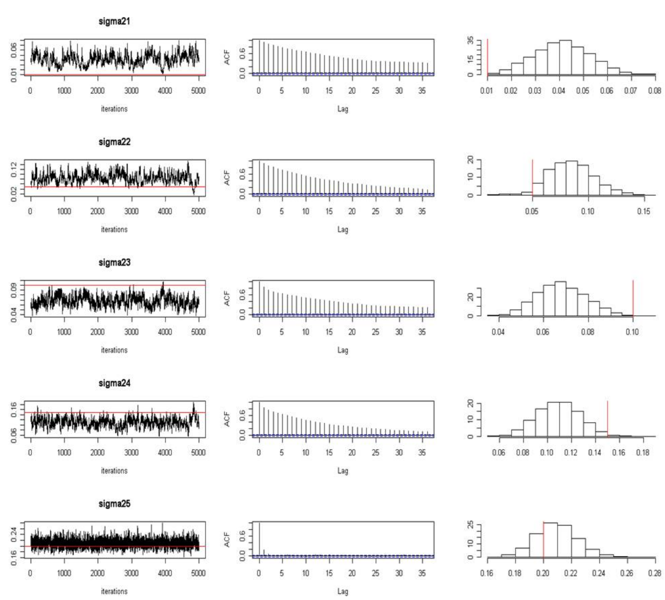
$$\beta = \begin{pmatrix} 0.99 & 0.00 \\ 0.00 & 0.99 \\ 0.90 & 0.00 \\ 0.00 & 0.90 \\ 0.50 & 0.50 \end{pmatrix} \qquad \Sigma = \text{diag} \begin{pmatrix} 0.01 \\ 0.05 \\ 0.10 \\ 0.15 \\ 0.20 \end{pmatrix}$$



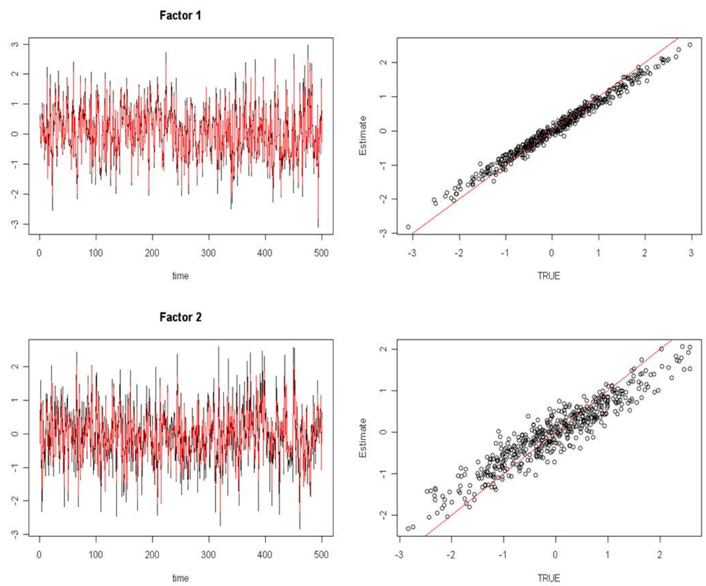
Factor loadings



Idiosyncratic variances



Common factors



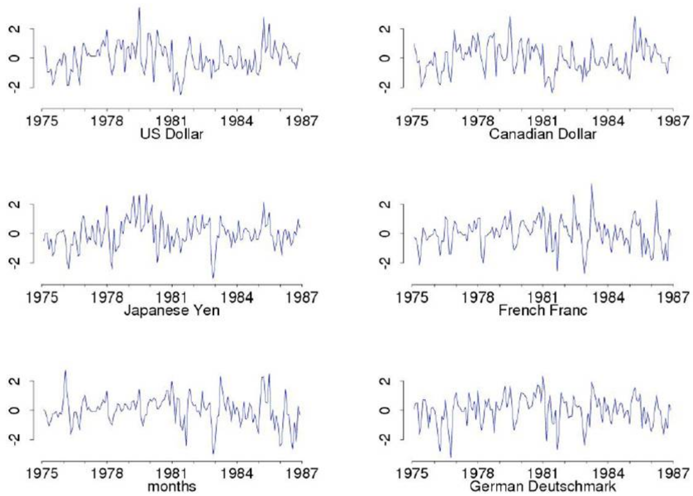
Exchange rates[2]

- Monthly international exchange rates.
- The data span the period from 1/1975 to 12/1986.
- Time series are the exchange rates in British pounds of
  - US dollar (US)
  - Canadian dollar (CAN)
  - Japanese yen (JAP)
  - French franc (FRA)
  - Italian lira (ITA)
  - (West) German (Deutsch)mark (GER)

Data

Standardized first differences of monthly log exchange rates

[2]Lopes and West (2004)



Standardized first differences of monthly observed exchange rates.

Posterior means

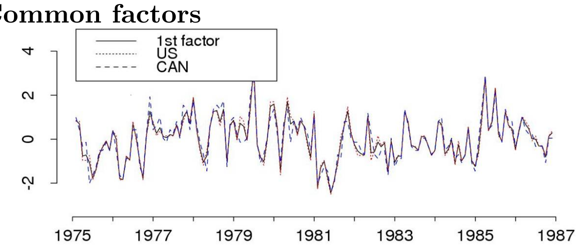
1st ordering

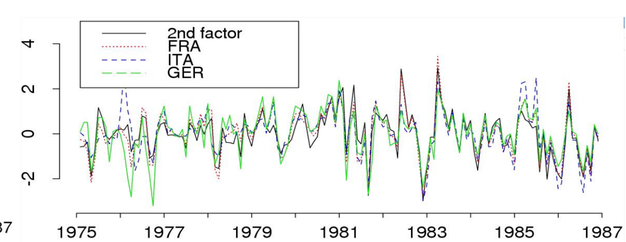
$$E(\beta|y) = \begin{pmatrix} \text{US} & 0.99 & 0.00 \\ \text{CAN} & 0.95 & 0.05 \\ \text{JAP} & 0.46 & 0.42 \\ \text{FRA} & 0.39 & 0.91 \\ \text{ITA} & 0.41 & 0.77 \\ \text{GER} & 0.40 & 0.77 \end{pmatrix} \quad E(\Sigma|y) = \text{diag}\begin{pmatrix} 0.05 \\ 0.13 \\ 0.62 \\ 0.04 \\ 0.25 \\ 0.28 \end{pmatrix}$$

2nd ordering

$$E(\beta|y) = \begin{pmatrix} \text{US} & 0.98 & 0.00 \\ \text{JAP} & 0.45 & 0.42 \\ \text{CAN} & 0.95 & 0.03 \\ \text{FRA} & 0.39 & 0.91 \\ \text{ITA} & 0.41 & 0.77 \\ \text{GER} & 0.40 & 0.77 \end{pmatrix} \quad E(\Sigma|y) = \text{diag}\begin{pmatrix} 0.06 \\ 0.62 \\ 0.12 \\ 0.04 \\ 0.25 \\ 0.26 \end{pmatrix}$$

Common factors





5.2.2 PBFA

Parsimonious BFA[3]

Frühwirth-Schnatter and Lopes (2009) revisited LW2004. We...

- Lay down a new and general set of identifiability conditions that handles the ordering problem present in most of the current literature,

- Introduce a new strategy for searching the space of parsimonious/sparse factor loading matrices,

- Designed a highly computationally efficient MCMC scheme for posterior inference which makes several improvements over the existing alternatives,

for the important class of Gaussian factor models.

Identification issues

- Block lower triangular Generalized lower triangular alternative

- Rank deficiency If β is rank-deficient, then $\exists Q$ such that

  $$\beta\beta' = (\beta + MQ')(\beta + MQ')' + (\Sigma - MM').$$

  for some orthogonal M with $\beta Q = 0$ and $Q'Q = I$. We use this "deficiency" in our model search strategy.

Generalized lower triangular

$$\begin{pmatrix} \beta_{11} & 0 & 0 & 0 \\ \beta_{21} & \beta_{22} & 0 & 0 \\ \beta_{31} & \beta_{32} & \beta_{33} & 0 \\ \beta_{41} & \beta_{42} & \beta_{43} & \beta_{44} \\ \beta_{51} & \beta_{52} & \beta_{53} & \beta_{54} \\ \beta_{61} & \beta_{62} & \beta_{63} & \beta_{64} \\ \beta_{71} & \beta_{72} & \beta_{73} & \beta_{74} \end{pmatrix} \Longrightarrow \begin{pmatrix} \beta_{11} & 0 & 0 & 0 \\ \beta_{21} & 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 & 0 \\ \beta_{41} & \beta_{42} & 0 & 0 \\ \beta_{51} & \beta_{52} & 0 & 0 \\ \beta_{61} & \beta_{62} & 0 & \beta_{64} \\ \beta_{71} & \beta_{72} & 0 & \beta_{74} \end{pmatrix}$$

Birth/death of loadings Birht/death of columns.

[3] Frühwirth-Schnatter and Lopes (2009) Parsimonious Bayesian factor analysis when number of factors is unknown. *Journal of Econometrics*.

Other contributions

- Our approach provides a principled way for inference on the number of factors, as opposed to previous work (Carvalho et al., 2008; Bhattacharya and Dunson, 2009).

- Our prior specification on Σ properly addresses Heywood problems

- Our fractional-like prior on β is more robust than the existing ones (Lopes and West, 2004, Ghosh and Dunson, 2009)

- Efficient (and correct) parameter expansion where the prior is unchanged (as opposed to GD2009).

The British Cohort Study[4]

A survey of all babies born (alive or dead) after the 24th week of gestation from 00.01 hours on Sunday, 5th April to 24.00 hours on Saturday, 11 April, 1970 in England, Scotland, Wales and Northern Ireland.

Follow-ups (so far): 1975, 1980, 1986, 1996, 2000, 2004, 2008.

Background characteristics:

- Cognitive, mental, physical health measurements (age 10)

- Education and adult outcomes (age 30)

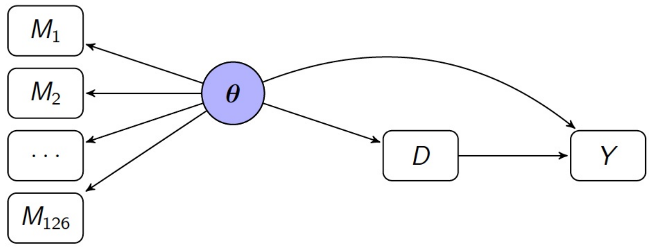
Sample size: 5,105 women and 5,420 men.

Graphical model

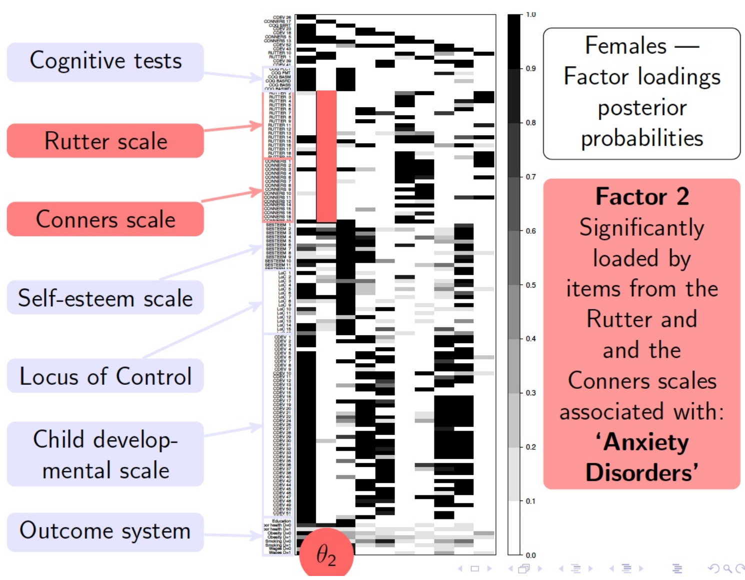
2nd factor

3rd factor

[4]Conti, Heckman, Frühwirth-Schnatter, Lopes and Piatek (2011) Constructing economically justified aggregates: an application to the early origins of health. *Journal of Econometrics*.



Variables		Definitions
Education	D	Observed: achieving A-level or higher
Outcomes	Y	Observed in one state only!
		Poor health, Obesity, Smoking, Wage
Measurements:	M_j	Observed: 126 items (binary and cont.)
Cognitive skills	θ	**Unobserved dimensions**
Personality		



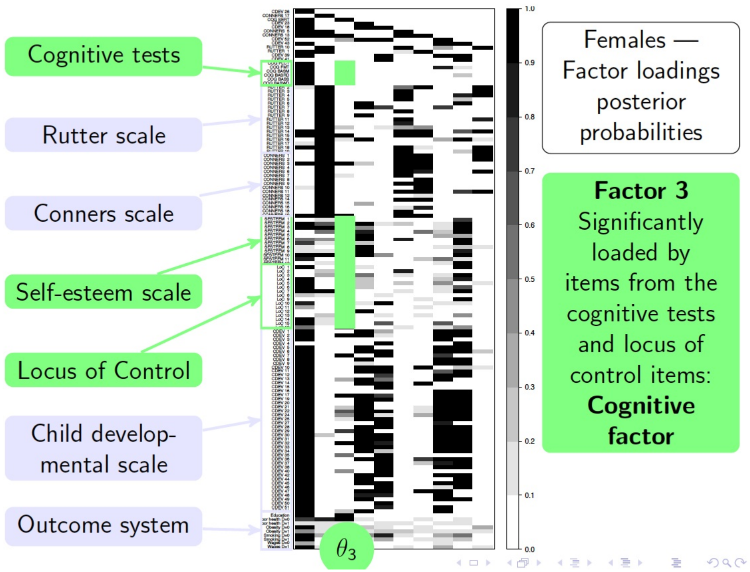
5.2.3 FSV

Factor stochastic volatility

Factor structure: y_t follows k-order factor stochastic volatility model[5] when

$$\begin{aligned} y_t|f_t &\sim N(\beta f_t, \Sigma_t) \\ f_t &\sim N(0, H_t) \end{aligned}$$

[5] Pitt and Shephard (1999).



and

$$\begin{aligned}\Sigma_t &= \text{diag}(\sigma^2_{1t}, \ldots, \sigma^2_{pt})\\ H_t &= \text{diag}(\sigma^2_{p+1,t}, \ldots, \sigma^2_{p+k,t}).\end{aligned}$$

Dynamic evolution: Let $\eta_{it} = \log(\sigma^2_{it})$ and $\lambda_{it} = \log(\sigma^2_{it})$. Then,

$$\begin{aligned}\eta_{it} &\sim N(\alpha_i + \gamma_i \eta_{i,t-1}, \xi^2_i)\\ \lambda_{jt} &\sim N(\mu_j + \phi_j \lambda_{j,t-1}, \tau^2_j).\end{aligned}$$

Correlated factor volatilities

Aguilar and West (2000) introduce contemporaneous covariation in the common factor log-volatilities.

Let

$$\lambda_t = (\sigma^2_{p+1,t}, \ldots, \sigma^2_{p+k,t})',$$

$$\mu = (\mu_1, \ldots, \mu_k)$$

and

$$\Phi = \text{diag}(\phi_1, \ldots, \phi_k).$$

Then

$$\lambda_t \sim N(\alpha + \phi\lambda_{t-1}, U)$$

where U is a full covariance matrix.

Time-varying loadings

Lopes and Carvalho (2007) introduce time-varying loadings, β_t.

The $d = pk - k(k-1)/2$ unconstrained elements of β_t, namely

$$\beta_{21,t}, \beta_{31,t}, \dots, \beta_{p,k,t},$$

are modeled by simple first order autoregressive models, ie.

$$\beta_{ijt} \sim N(\zeta_{ij} + \Theta_{ij}\beta_{ij,t-1}, \omega_{ij}^2)$$

for $i = 2, \dots, p$ and $j = 1, \dots, \min(i-1, k)$.

Daily exchange rate

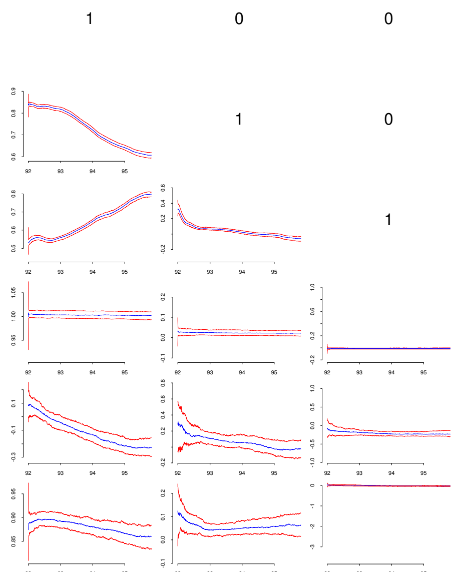
Returns on weekday closing spot prices for six currencies relative to the US dollar.

The data span the period from 1/1/1992 to 10/31/1995 inclusive.

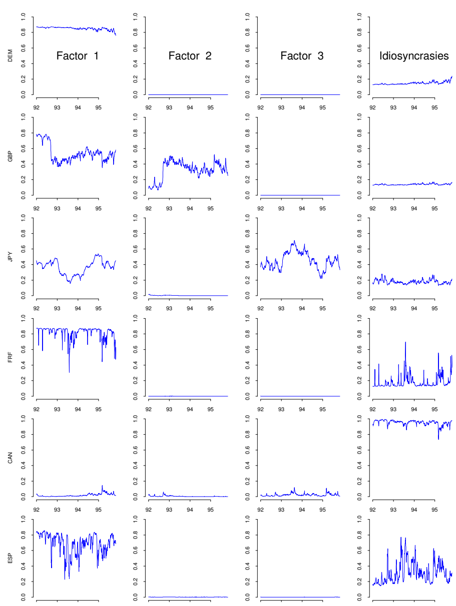
- German Mark(DEM)
- British Pound(GBP)
- Japanese Yen(JPY)
- French Franc(FRF)
- Canadian Dollar(CAD)
- Spanish Peseta(ESP)

A $k = 3$ factor stochastic volatility model with time-varying loadings was implemented with relatively vague priors for all model parameters.

Factor loadings



Variance decomposition



5.2.4 SDFM

Spatial dynamic factor models

Lopes, Salazar and Gamerman (2008) introduces the following spatio-temporal model for measurements on m spatial locations and over T time periods.

Factor structure:

$$y_t \sim N(\beta f_t, \Sigma)$$

Dynamic factors:

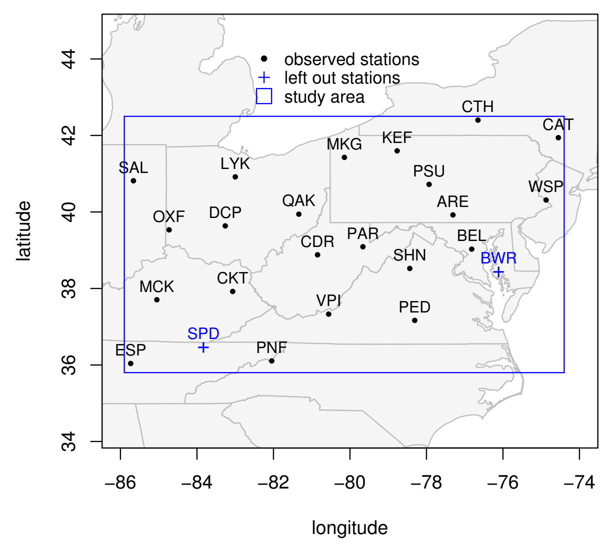
$$f_t \sim N(\Gamma f_{t-1}, \Gamma)$$

Spatial covariation:

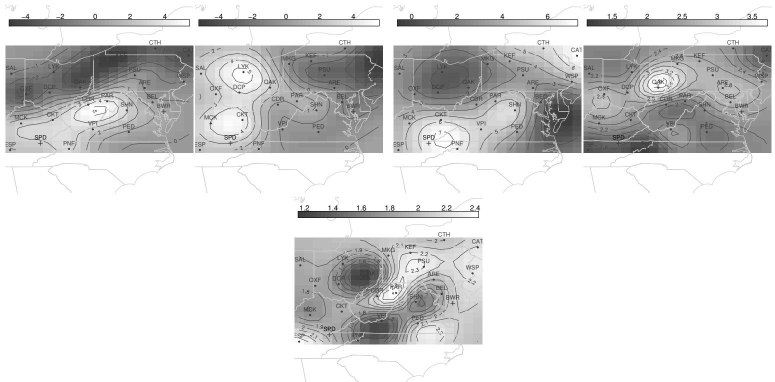
$$\beta_j \sim GP(\mu_j, \tau_j^2 R_{\phi_j})$$

where $\beta = (\beta_1, \ldots, \beta_k)$ and R_{ϕ_j} spatial correlation matrix.

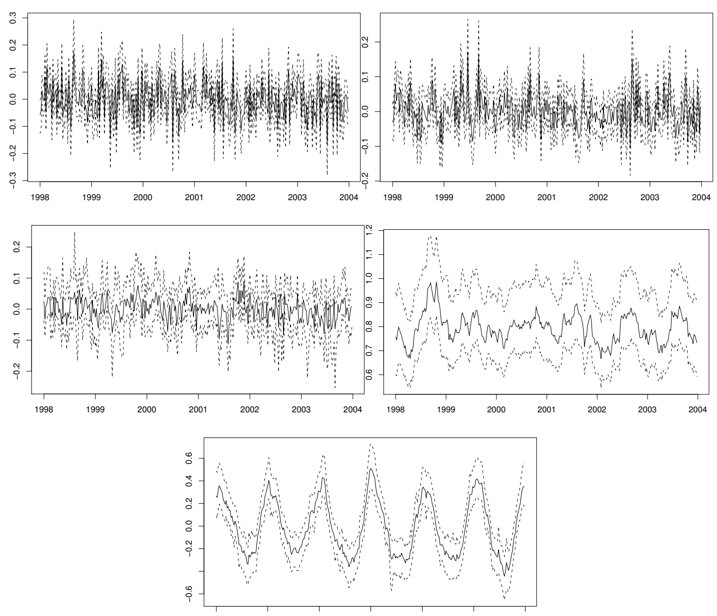
SO_2 in the US



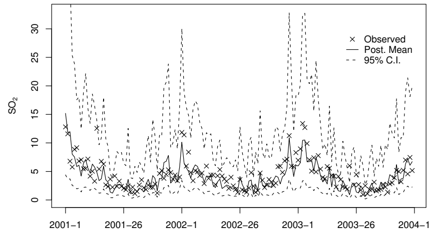
Spatial factor loadings

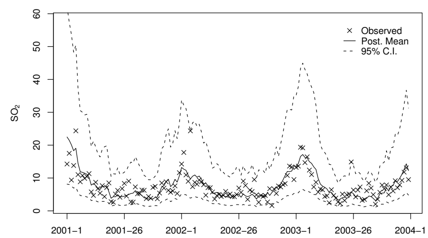


Dynamic factors



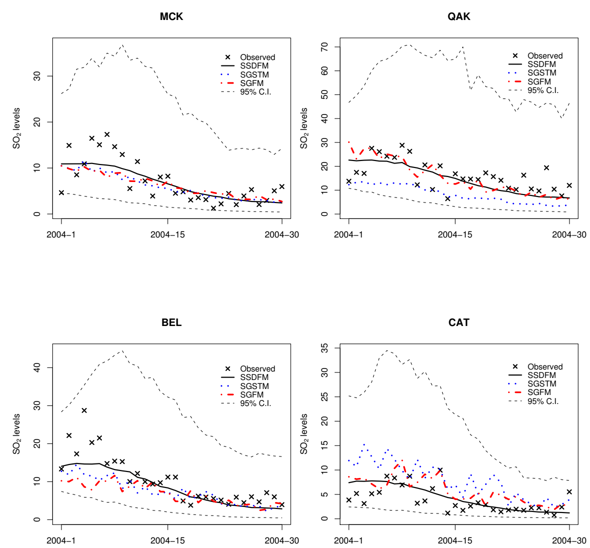
Spatial interpolation





Interpolated values at stations SPD and BWR.

Forecasting



5.3 Forecasting with large q

Forecasting with large BVAR

Carriero *et al.* (2009) forecasts $q = 33$ exchange rates, vis-a-vis the US dollar, with a Bayesian VAR model.

They find that the BVAR can lead to gains in forecast accuracy for the large majority of the exchange rates under analysis.

The forecast gains are typically in the range of 2%-3%, but in some relevant cases, such as the Euro-Dollar and the GBP-Dollar, they can reach 6%-9%.

A simple trading strategy based on the BVAR forecasts provides positive returns which are higher than those from RW forecasts.

In the post-2000 period the information in the exchange rates of emerging countries seems to matter for forecasting those of the developed countries more than the reverse.

5.4 Forecasting with many predictors

Forecasting in VAR with many predictors

Korobilis (2008) addresses the issue of improving the forecasting performance of VARs when the set of available predictors is inconveniently large to handle with methods and diagnostics used in traditional small-scale models.

Assume we have available observations $x_t = (x_{t1}, \ldots, x_{tr})'$ on some macroeconomic quantities, where r is large (in the order of hundreds).

Factor analysis

A popular and simple method to incorporate into an econometric model all the information inherent in a large set of variables, is to reduce their dimension into a lower-dimensional vector of $k << n$ latent factors and insert these in the VAR model as explanatory variables:

$$\begin{aligned}(y_t|y_{t-1:t-p}, \Theta) &\sim N\left(\sum_{i=1}^{p} B_i y_{t-i} + \sum_{i=0}^{s} A_i f_{t-i}, \Sigma\right)\\ (x_t|f_t, \Theta) &\sim N(\beta f_t, \Omega),\end{aligned}$$

where $\Theta = (B_1, \ldots, B_p, A_0, \ldots, A_s, \Sigma, \beta, \Omega)$.

This model is closely related o the factor augmented VAR (FAVAR) of Bernanke *et al.* (2005).

Data

Korobilis (2008) uses the Stock and Watson (2005) data set, with $q = 132$ US variables (JAN60-DEC03).

real output and income employment and hours real retail, manufacturing, and trade sales consumption housing starts and sales real inventories orders stock prices exchange rates interest rates and spreads money and credit quantity aggregates price indexes average hourly earnings miscellaneous

RMSFE

An appropriate common way to quantify out-of-sample forecasting performance is to compute the root mean square forecast error (RMSFE) statistic for each forecast horizon h:

$$\mathrm{RMSFE}_{ij}^{h} = \sqrt{\sum_{t=1982:12}^{2003:12-h} (y_{i,t+h}^{*} - \tilde{y}_{i,t+h,j})^2}$$

where $y_{i,t+h}^{*}$ is the realized (observed) value of y at time $t+h$ for the i-th series, and $\tilde{y}_{i,t+h,j}$ is the mean of the posterior predictive density at time $t+h$, for the i-th series, from the j-th forecasting model.

Relative RMFSE

Table C3. Forecast Comparison - relative RMSFE

	PI	IP	EMP	UR	TBILL	PPI	CPI	PCED
BVAR with factors (Bayesian Model Averaging)								
$h = 1$	0.94	1	0.9	0.96	1.08	0.88	0.95	1.09
$h = 4$	1.06	0.96	0.93	0.94	0.95	0.92	1.05	0.94
$h = 12$	0.97	0.92	0.99	1.02	0.98	0.92	0.95	0.96
BVAR with factors (Model Selection)								
$h = 1$	0.86	0.98	0.87	0.96	1.06	0.91	0.93	0.91
$h = 4$	0.9	0.97	0.85	0.92	0.94	0.94	0.98	0.93
$h = 12$	0.87	0.99	0.91	0.98	0.89	0.87	0.99	0.96
VAR with factors (BIC Selection)								
$h = 1$	0.92	0.99	0.94	0.99	1.22	0.99	1.01	0.97
$h = 4$	0.93	0.97	0.94	0.94	1.12	0.97	1.06	0.94
$h = 12$	0.97	1.04	0.98	1.05	0.99	0.9	1.1	0.95

Note: The variables of interest are: PI: Personal Income (A0M052), IP: Industrial Production(IPS10), EMP:Employment Rate (CES002), UR: Unemployment Rate (LHUR), TBILL: 3-month Treasury Bill Rate (FYGM3), PPI: Producer Price Index (PWFSA), CPI: Consumer Price Index (PUNEW), andPCED: PCE Deflator (GMDC)

5.5 Large BVAR & BFAVAR

Large BVAR & BFAVAR

Bańbura *et al.* (2010)show that VAR with Bayesian shrinkage is an appropriate tool for large dynamic models.

When the degree of shrinkage is set in relation to the cross-sectional dimension, the forecasting performance of small monetary VARs can be improved by adding additional macroeconomic variables and sectoral information.

Large VARs with shrinkage produce credible impulse responses and are suitable for structural analysis.

Data sets

SMALL: Small monetary VAR with 3 variables.

CEE: Monetary model of Christiano et al. (1999), or SMALL + 4 variables.

MEDIUM: CEE + 13 variables.

LARGE: All 131 macroeconomic indicators from Stock and Watson (2005).

Table II. OLS and BVAR, relative MSFE, 1971–2003

		SMALL			*CEE*			*LARGE*
		$p = 13$	$p = \text{BIC}$	BVAR	$p = 13$	$p = \text{BIC}$	BVAR	BVAR
$h = 1$	EMPL	1.14	0.73	1.14	7.56	0.76	0.67	0.46
	CPI	0.89	0.55	0.89	5.61	0.55	0.52	0.50
	FFR	1.86	0.99	1.86	6.39	1.21	0.89	0.75
$h = 3$	EMPL	0.95	0.76	0.95	5.11	0.75	0.65	0.38
	CPI	0.66	0.49	0.66	4.52	0.45	0.41	0.40
	FFR	1.77	1.29	1.77	6.92	1.27	1.07	0.94
$h = 6$	EML	1.11	0.90	1.11	7.79	0.78	0.78	0.50
	CPI	0.64	0.51	0.64	4.80	0.44	0.41	0.40
	FFR	2.08	1.51	2.08	15.9	1.48	1.30	1.29
$h = 12$	EMPL	1.02	1.15	1.02	22.3	0.82	1.21	0.78
	CPI	0.83	0.56	0.83	21.0	0.53	0.57	0.44
	FFR	2.59	1.59	2.59	47.1	1.62	1.71	1.93

Notes: The table reports MSFE relative to that from the benchmark model (random walk with drift) for employment (EMPL), CPI and federal funds rate (FFR) for different forecast horizons h and different models. *SMALL*, *CEE* refer to the VARs with 3 and 7 variables, respectively. Those systems are estimated by OLS with number of lags fixed to 13 or chosen by the BIC. For comparison, the results of Bayesian estimation of the two models and of the large model are also provided.

Table III. FAVAR, relative MSFE, 1971–2003

		FAVAR 1 factor			FAVAR 3 factors			*LARGE*
		$p = 13$	$p = \text{BIC}$	BVAR	$p = 13$	$p = \text{BIC}$	BVAR	BVAR
$h = 1$	EMPL	1.36	0.54	0.70	3.02	0.52	0.65	0.46
	CPI	1.10	0.57	0.65	2.39	0.52	0.58	0.50
	FFR	1.86	0.98	0.89	2.40	0.97	0.85	0.75
$h = 3$	EMPL	1.13	0.55	0.68	2.11	0.50	0.61	0.38
	CPI	0.80	0.49	0.55	1.44	0.44	0.49	0.40
	FFR	1.62	1.12	1.03	3.08	1.16	0.99	0.94
$h = 6$	EMPL	1.33	0.73	0.87	2.52	0.63	0.77	0.50
	CPI	0.74	0.52	0.55	1.18	0.46	0.50	0.40
	FFR	2.07	1.31	1.40	3.28	1.45	1.27	1.29
$h = 12$	EMPL	1.15	0.98	0.92	3.16	0.84	0.83	0.78
	CPI	0.95	0.58	0.70	1.98	0.54	0.64	0.44
	FFR	2.69	1.43	1.93	7.09	1.46	1.69	1.93

Notes: The table reports MSFE for the FAVAR model relative to that from the benchmark model (random walk with drift) for employment (EMPL), CPI and federal funds rate (FFR) for different forecast horizons h. FAVAR includes 1 or 3 factors and the three variables of interest. The system is estimated by OLS with number of lags fixed to 13 or chosen by the BIC and by applying Bayesian shrinkage. For comparison the results from large Bayesian VAR are also provided.

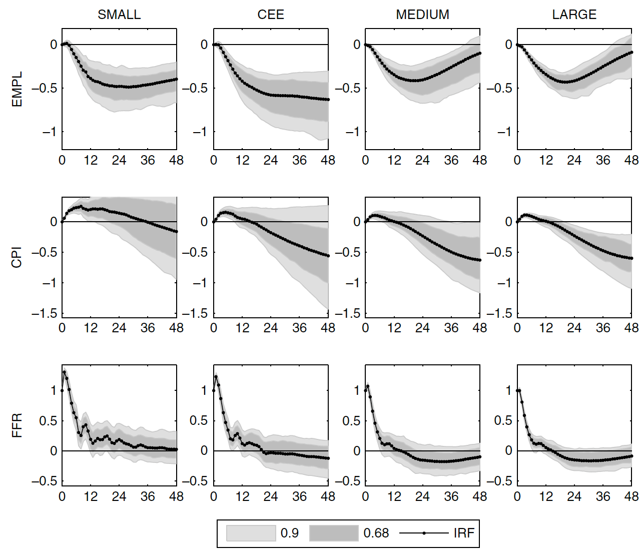


Figure 1. BVAR, impulse response functions. The figure presents the impulse response functions to a monetary policy shock and the corresponding posterior coverage intervals at 0.68 and 0.9 level for employment (EMPL), CPI and federal funds rate (FFR). *SMALL*, *CEE*, *MEDIUM* and *LARGE* refer to VARs with 3, 7, 20 and 131 variables, respectively. The prior on the sum of coefficients has been added with the hyperparameter $\tau = 10\lambda$.

5.6 Large TVP-BVAR & BFAVAR

Large TVP-BVAR & BFAVAR

Koop and Korobilis (2012) develop methods for estimation and forecasting in large time-varying parameter vector autoregressive models (TVP-VARs).

They use the SUR representation of the VAR(p) model with time-varying parameters

$$\begin{aligned} y_t &= z_t\beta_t + u_t \qquad u_t \sim N(0, \Sigma) \\ \beta_t &= \beta_{t-1} + v_t \qquad v_t \sim N(0, V_t) \end{aligned}$$

They use the old *discount* strategy of West and Harrison (1997) to avoid MCMC altogether.

See Ahmadi and Uhlig (2009) for a fully Bayesian FAVAR alternative.

Review papers

Two recent review chapters in the same volume are

Geweke and Whiteman (2006) Bayes Forecasting. In *Handbook of Economic Forecasting*, 3-80.

Stock and Watson (2006) Forecasting with Many Predictors. In *Handbook of Economic Forecasting*, 515-554.

5.7 References

References

Aguilar and West(2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics*, 18, 338-357.

Ahmadi and Uhlig (2009). Measuring the Dynamic Effects of Monetary Policy Shocks: A Bayesian FAVAR Approach with Sign Restriction. Technical Report.

Bańbura, Giannone and Reichlin (2010). Large Bayesian vector auto regressions. JAE, 25, 71-92.

Bartholomew, D. J. (1981). Posterior analysis of the factor model. *The British Journal of Mathematical and Statistical Psychology*, 34, 93-99.

Bartholomew, D. J. (1987) *Latent Variable Models and Factor Analysis*. London: Charles Griffin.

Bernanke, Boivin and Eliasz (2005) Measuring monetary policy: a factor augmented autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120, 387-422.

Brandt and Freeman (2006) Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting, and Policy Analysis. *Political Analysis*, 14, 1-36.

Carriero, Kapetanios, Marcellino (2009) Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25, 400-417.

Del Negro and Schorfheide (2011) Bayesian Macroeconometrics. In *Handbook of Bayesian Econometrics*, Chapter 7, 293-387. Oxford University Press.

Doan, Litterman and Sims (1984) Forecasting and conditional projection using realistic prior distributions (with discussion). *Econometric Reviews*, 3, 1-144.

Garratt, Koop, Mise and Vahey (2009) Real-Time Prediction With U.K. Monetary Aggregates in the Presence of Model Uncertainty. *Journal of Business & Economic Statistics*, 27, 480-491.

References (cont.)

George (2000) The variable selection problem. *Journal of the American Statistical Association*, 95, 1304-1308.

George, Sun and Ni (2008) Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142, 553-580.

George and McCulloch (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88, 881-889.

George and McCulloch (1997) Approaches to Bayesian variable selection. *Statistica Sinica*, 7, 339-379.

Geweke and Singleton (1980). Interpreting the likelihood ratio statistic in factor models when sample size is small. *Journal of the American Statistical Association*, 75, 133-137.

Geweke and Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies*, 9, 557-587.

Ihara and Kano (1995) Identifiability of full, marginal, and conditional factor analysis model. *Statistics and Probability Letters*, 23, 343-50.

Jochmann, Koop, Leon-Gonzalez and Strachan (2011) Stochastic search variable selection in vector error correction models with application to a model of the UK macroeconomy. *Journal of Applied Econometrics*.

Jochmann, Koop and Strachan (2010) Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks. *International Journal of Forecasting*, 26, 326-347.

Kadiyala and Karlsson (1993) Forecasting with generalized Bayesian vector autoregressions. *Journal of Forecasting*, 12, 365-78.

Kadiyala and Karlsson (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12, 99-132.

References (cont.)

Koop and Korobilis (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3, 267-358.

Koop and Korobilis (2012) Large Time-Varying Parameter VARs. Technical Report.

Korobilis (2008) Forecasting in vector autoregressions with many predictors. In *Bayesian Econometrics (Advances in Econometrics)*, 23, 403-431.

Korobilis (2012) VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, forthcoming.

Litterman (1980) A Bayesian procedure for forecasting with vector autoregressions. Technical Report, MIT.

Litterman (1986) Forecasting with Bayesian vector autoregressions – five years of experience. *Journal of Business & Economic Statistics*, 4, 25-38.

Lopes (2000) Bayesian Analysis in Latent Factor and Longitudinal Models. Unpublished PhD. Thesis, Institute of Statistics and Decision Sciences, Duke University.

Lopes (2003) Factor models. *ISBA Bulletin*, 10(2), 7-10. `http://faculty.chicagobooth.edu/hedibert.lopes/research/pdf/factormodels-lopes.pdf`

Lopes and Carvalho (2007) Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*, 137, 3082-3091.

Lopes, McCulloch and Tsay (2012) Cholesky stochastic volatility models for high-dimensional time series. Technical Report, Chicago Booth.

Lopes and Polson (2012) Bayesian instrumental variables: likelihoods and priors. *Econometric Reviews*.

References (cont.)

Lopes and West (2004) Bayesian model assessment in factor analysis. *Statistica Sinica*, 14, 41-67.

Lopes, Salazar and Gamerman (2008) Spatial dynamic factor models. *Bayesian Analysis*,

Lütkepohl (2007) *New Introduction to Multiple Time Series Analysis.* Springer.

Martin and McDonald (1975) Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, 40, 505-517.

Nakajima, Kasuya and Watanabe (2011) Bayesian analysis of time-varying parameter vector autoregressive model for the Japanese economy and monetary policy. *Journal of The Japanese and International Economies*, 25, 225-245.

Pelloni and Polasek (2003) Macroeconomic Effects of Sectoral Shocks in Germany, the U.K. and, the U.S.: A VAR-GARCH-M Approach. *Computational Economics*, 21, 65-85.

Pfaff (2008) VAR, SVAR and SVEC Models: Implementation within `R` Package `vars`. *Journal of Statistical Software*, 27, 1-32.

Pitt and Shephard (1999) Time varying covariances: A factor stochastic volatility approach. In *Bayesian statistics 6* (ed. J.M. Bernardo et al). London: Oxford University Press.

Pourahmadi (1999) Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* 86, 677-690.

Press (1982) *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference* (2nd edition). New York: Krieger.

Press and Shigemasu (1989) Bayesian inference in factor analysis. In *ASA Proceedings of Social Statistics Section*, 292-294.

References (cont.)

Primiceri (2005). Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies*, 72, 821-852.

Rubio-Ramírez, Waggoner and Zha (2010) Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies*, 77, 665-696.

Sims (1980) Macroeconomics and Reality. *Econometrica*, 48, 1-48.

Sims (2007) Bayesian Methods in Applied Econometrics, or, Why Econometrics Should Always and Everywhere Be Bayesian. Technical Report, Princeton University.

Sims and Zha (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review*, 39, 949-968.

Smith and Kohn (2002) Parsimonious covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association*, 97, 1141-1153.

Stock and Watson (2005) Implications of dynamic factor models for VAR analysis. Technical Report.

Uhlig (1997) Bayesian Vector Autoregressions with Stochastic Volatility. *Econometrica*, 65, 59-73.

Villani (2009) Steady-state priors for vector autoregressions. *Journal of Applied Econometrics*, 24, 630-650.

Waggoner and Zha (2003) A Gibbs Sampler for Structural VARs. *Journal of Economic Dynamics and Control*, 28, 349-366.

West (2003) Bayesian factor regression models in the "large p, small n" paradigm". in *Bayesian Statistics 7*, 723-732.

Zha (1999) Block recursion and structural vector autoregressions. *Journal of Econometrics*, 90, 291-316.

Zivot and Wang (2003) *Modeling Financial Time Series with S-plus*. Springer.