Linear Regression with General Covariance

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

Outline

1. General covariance
   - Known Ω
   - Ω diagonal
     - $\omega_i = h(z_i, \alpha)$
     - t errors
   - Parsimonious Ω
2. Seemingly unrelated regressions

General covariance

Suppose that

$$y_i = x_i'\beta + \epsilon_i \qquad N(0, \sigma^2\omega_{ii})$$

and

$$cov(\epsilon_i, \epsilon_j) = \omega_{ij}$$

for $i, j = 1, \ldots, n$.

Therefore,

$$y = X\beta + \epsilon \qquad \epsilon \sim N(0, \sigma^2\Omega)$$

Known Ω

If $P\Omega P' = I_n$ then

$$\epsilon^* = P\epsilon \sim N(0, \sigma^2 I_n)$$

and

$$y^* = Py = PX\beta + \epsilon^* = X^*\beta + \epsilon^*$$

which is, given Ω, a standard linear regression model.

Posterior inference can be obtained by means of a standard Gibbs sampler when

$$p(\beta, \sigma^2|\Omega) = p(\beta)p(\sigma^2)$$

and

$$\beta \sim N(\beta_0, V_0) \quad \text{and} \quad \sigma^2 \sim IG(\nu_0/2, \nu_0 s_0^2/2)$$

Full conditionals

It is easy to see that

$$\sigma^2|\beta, \Omega, y \sim IG(\nu_1/2, \nu_1 s_1^2/2)$$

where

$$\begin{aligned} \nu_1 &= \nu_0 + n \\ \nu_1 s_1^2 &= \nu_0 s_0^2 + (y - X\beta)'\Omega^{-1}(y - X\beta) \end{aligned}$$

It is also easy to see that

$$\beta|\sigma^2, \Omega, y \sim N(\beta_1, V_1)$$

where

$$\begin{aligned} V_1^{-1} &= V_0^{-1} + \sigma^{-2} X'\Omega^{-1} X \\ V_1^{-1}\beta_1 &= V_0^{-1}\beta_0 + \sigma^{-2} X'\Omega^{-1} X y \end{aligned}$$

Ω diagonal

Let us consider two forms of

$$\Omega = \text{diag}(\omega_1, \ldots, \omega_n)$$

Case I: $\omega_i = h(z_i, \alpha)$

Case II: $\omega_i^{-1} \sim IG(\nu/2, \nu/2)$

Case I: $\omega_i = h(z_i, \alpha)$

One possible example is

$$h(z_i, \alpha) = (\alpha_1 z_{i1} + \cdots + \alpha_p z_{ip})^2$$

which is a function of p parameters.

Another (perhaps more) common example is

$$h(z_i, \alpha) = \exp\{\alpha_1 z_{i1} + \cdots + \alpha_p z_{ip}\}$$

In general, Metropolis-Hastings steps will be required to iteratively sample from

$$p(\alpha | y, X, Z, \beta, \sigma^2)$$

Simulated example

We simulated $n = 100$ observations based on $\beta = (2.0, 4.0)$, $\alpha = (1.0, 1.0)$ and 1's in the first column of x, such that $h(x_i, \alpha) = \exp\{\alpha_0 + \alpha_1 x_i\}$, for $x_i \sim N(0, 1)$. Because of the presence of α_0, we set $\sigma^2 = 1$ for identification purposes.

The prior hyperparameters are:

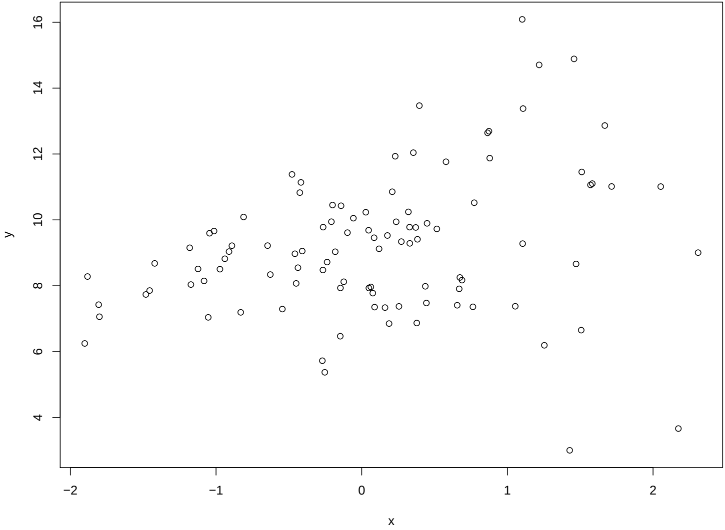
$$\beta_0 = 0 \quad V_0 = 100 I_2 \quad \alpha_0 = 0.0 \quad \text{and} \quad V_\alpha = 100$$

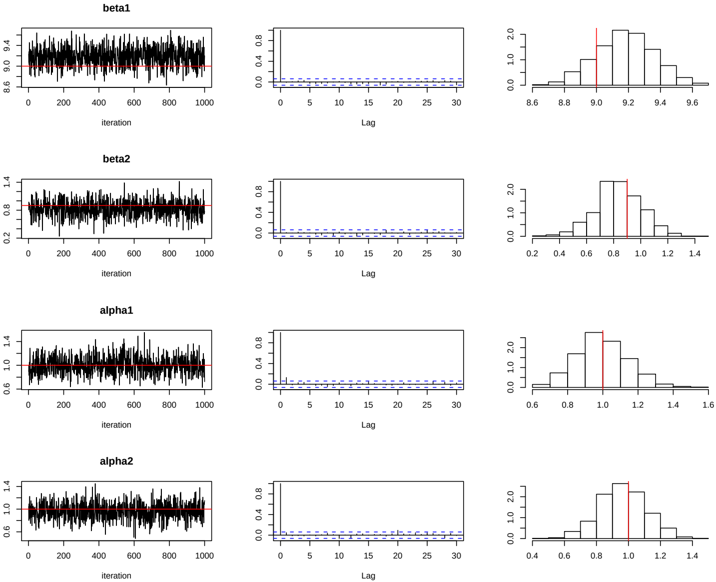
The initial values for the MCMC scheme are:

$$\beta = \hat{\beta}_{ols} = (9.12, 0.62) \quad \alpha = (-0.29, 0.99)^1$$

with burn-in of $M_0 = 1000$ and retaining every $k = 20^{th}$ draw until $M = 1000$ draws are obtained for posterior inference.

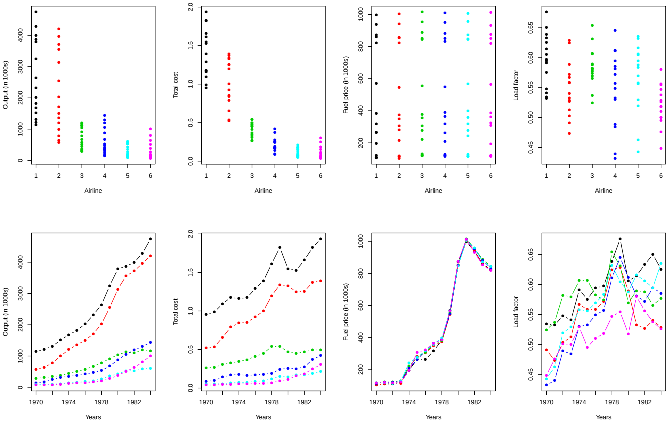
[1] Regression of $\log(y_i - x_i'\beta_{ols})^2$ on x_i.


16
14
12
10
8
6
4
y
−2
−1
0
1
2
x


beta1
iteration
Lag
beta2
iteration
Lag
alpha1
iteration
Lag
alpha2
iteration
Lag

Cost data for US Airlines

Example taken from Greene's book on Econometric Analysis (6th edition).
The data can be found in `http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm`
Cost Data for U.S. Airlines, 90 Oservations On 6 Firms For 15 Years, 1970-1984
Columns are: I = Airline, T = Year, Q = Output, in revenue passenger miles, index number,
C = Total cost, in \$1000, PF = Fuel price, LF = Load factor, the average capacity utilization of the fleet.



General covariance

Known Ω

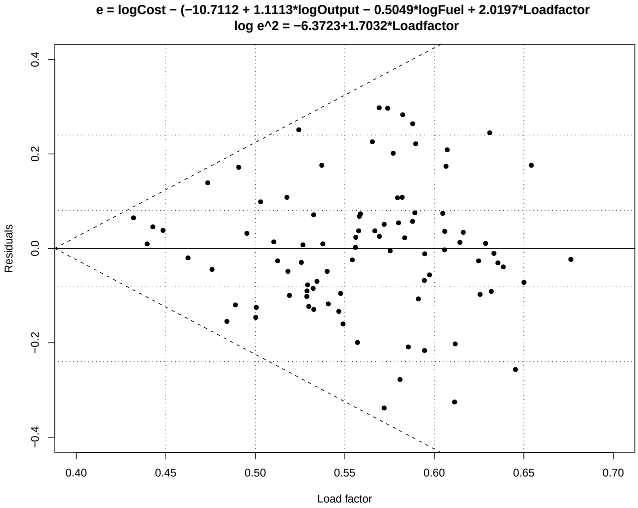
Ω diagonal

$\omega_i = h(z_i, \alpha)$

t errors

Parsimonious Ω

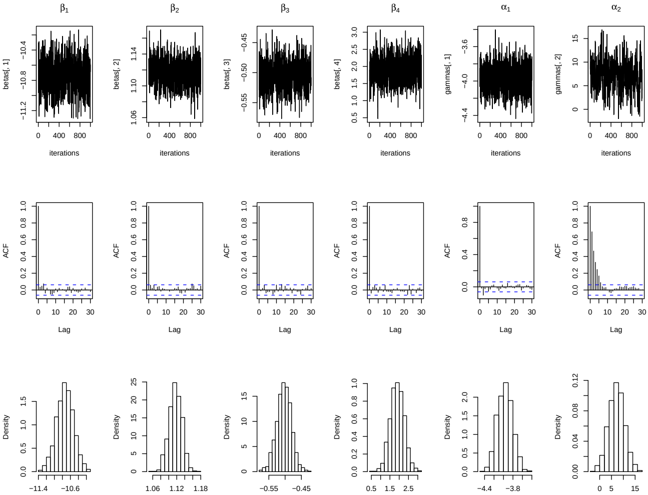
Seemingly unrelated regressions



Prior setup: $\beta_0 = 0_p$, $V_0 = 10000 I_p$, $\alpha_0 = 0_q$ and $V_\alpha = 10000 I_q$.

Random walk MH sampler

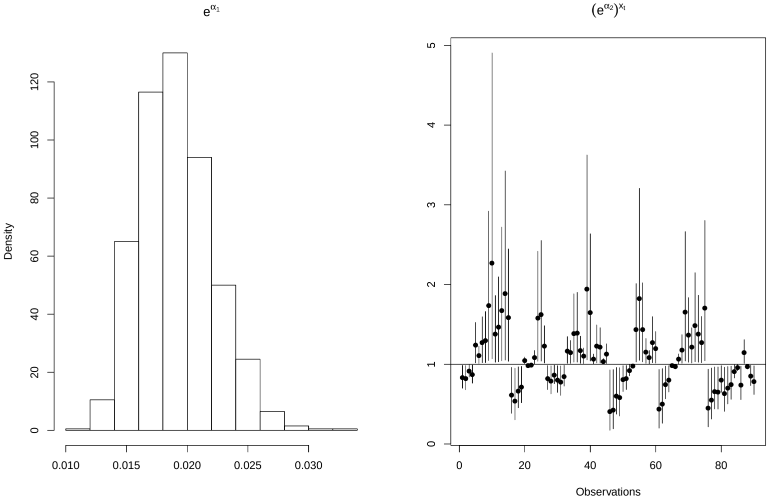
$(M0, LAG, M) = (1000, 100, 1000)$
$\beta^{(0)} = \hat{\beta} = (x'x)^{-1}x'y$ and $\alpha^{(0)} = (z'z)^{-1}z' \log(y - x'\hat{\beta})^2$
Random Walk variances: (0.5,0.5)



Posterior summary

			Percentiles		
θ	$E(\theta)$	$\sqrt{V(\theta)}$	2.5th	50th	97.5th
β_1	-10.733	0.211	-11.146	-10.730	-10.317
β_2	1.117	0.016	1.086	1.117	1.148
β_3	-0.504	0.023	-0.549	-0.503	-0.460
β_4	1.895	0.374	1.178	1.887	2.626
α_1	-3.968	0.157	-4.256	-3.964	-3.667
α_2	7.085	3.293	0.572	7.077	13.731

Variance components



General covariance
Known Ω
Ω diagonal
$\omega_i = h(z_i, \alpha)$
t errors
Parsimonious Ω
Seemingly unrelated regressions

Case II: $\omega_i \sim IG(\nu/2, \nu/2)$

Recall that

$$\epsilon_i | \omega_i \sim N(0, \sigma^2 \omega_i)$$

If, in addition,

$$\omega_i \sim IG(\nu/2, \nu/2),$$

it follows that

$$\epsilon_i \sim t_\nu(0, \sigma^2).$$

In other words, the Student t distribution is a scale mixture of normal distributions[2].

---

[2] See, amongst others, West (1987) On Scale Mixture of Normal Distributions. *Biometrika*, 74(3), 646-648, Geweke (1993) Bayesian Treatment of the Independent Student-t Linear Model. *Journal of Applied Econometrics*, 8, S19-S40, and Fernández and Steel (2000) Bayesian Regression Analysis with Scale Mixtures of Normals. *Econometric Theory*, 16(1), 80-101.

Prior of ν

A common choice is the exponential

$$\nu \sim G(1, 1/\nu^0) \equiv Exp(1/\nu^0)$$

such that $E(\nu) = \sqrt{V(\nu)} = \nu^0$.

For instance, $\nu^0 = 25$ allocates substantial prior on $\nu \leq 10$ (fat-tailed) and $\nu \geq 40$ (normality).

Note: Geweke (1993) showed that if $p(\beta) \propto 1$ then

$$E(\beta|y, X) \text{ does not exist unless } Pr\{\nu \in (0, 2]\} = 0$$

and

$$V(\beta|y, X) \text{ does not exist unless } Pr\{\nu \in (0, 4]\} = 0.$$

Medians and other quantiles will still exist.

The full conditional for ω_i, $i = 1, \ldots, n$ is

$$\omega_i | y, X, \beta, \sigma^2, \nu \sim IG\left(\frac{\nu+1}{2}, \frac{\nu + \epsilon_i^2/\sigma^2}{2}\right)$$

The full conditional for ν is

$$p(\nu|\omega) \propto \left[\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\right]^n \exp\left\{-\left[\frac{1}{\nu^0} + \frac{1}{2}\sum_{i=1}^{n}(\ln\omega_i + \omega_i^{-1})\right]\nu\right\}.$$

Sampling ν requires a Metropolis-Hastings step.

Simulated example

Here we simulated $n = 100$ observations based on: $\nu = 10$, $\sigma^2 = 1.0$ and $\beta = (0.5, 1.0, 2.0)$.

The prior hyperparameters are:

$$\beta_0 = 0 \quad V_0 = 10I_3 \quad \nu_0 = 5 \quad s_0^2 = 1 \quad \text{and} \quad \nu^0 = 25$$

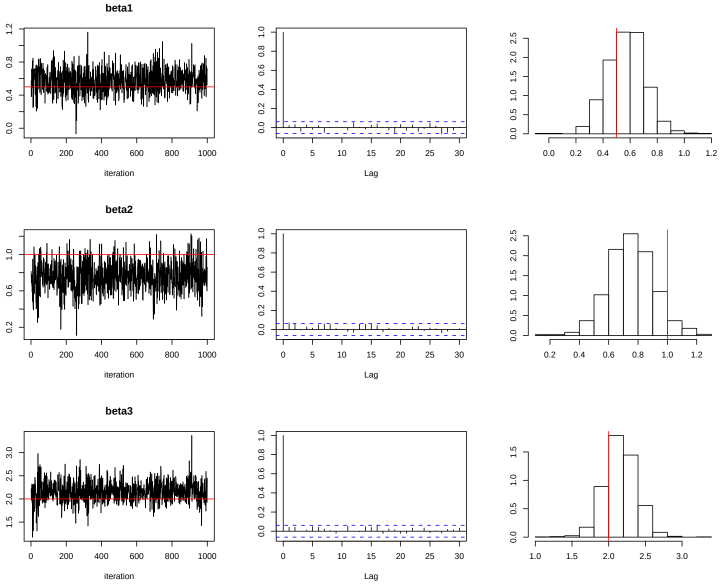
The initial values for the MCMC scheme are:

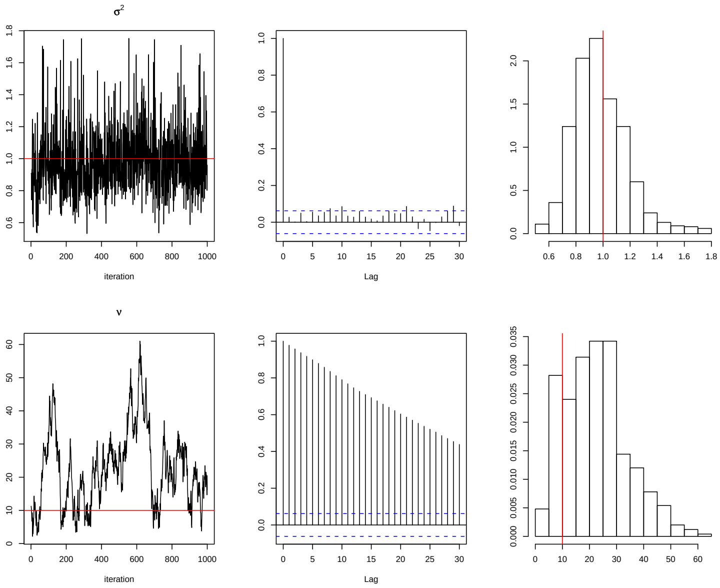
$$\beta = \hat{\beta}_{ols} \quad \sigma^2 = \hat{\sigma}^2_{ols} \quad \nu = 10 \quad \text{and} \quad \omega_i = 1 \ \forall i$$

with

$$q(\nu|\nu^{(j-1)}) = N(\nu^{(j-1)}, 0.25^2),$$

burn-in of $M_0 = 10000$ and retaining every $k = 100^{th}$ draw until $M = 1000$ draws are obtained for posterior inference.


beta1
beta2
beta3
iteration
Lag


σ²
ν
iteration
Lag
iteration
Lag

Parsimonious Ω

Let y_t follow a linear regression with autocorrelated errors, i.e.

$$y_t = x_t'\beta + \epsilon_t$$

where

$$\epsilon_t | \epsilon_{t-1} \sim N(\rho\epsilon_{t-1}, \sigma^2).$$

If $|\rho| < 1$, then

$$\begin{aligned} E(\epsilon_t) &= 0 && \forall t \\ Cov(\epsilon_t, \epsilon_{t+k}) &= \rho^k \left(\frac{\sigma^2}{1-\rho^2} \right) && \forall t, k. \end{aligned}$$

In this case

$$\epsilon|\sigma^2, \rho \sim N(0, \sigma^2\Omega_\rho),$$

where

$$\Omega_\rho = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-3} & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & \rho & 1 \end{pmatrix}.$$

and

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

The full conditional distributions of β and σ^2 are as before:
$\beta|\sigma^2, \Omega, y \sim N(\beta_1, V_1)$ and $\sigma^2|\beta, \Omega, y \sim IG(\nu_1/2, \nu_1 s_1^2/2)$, where
$V_1^{-1} = V_0^{-1} + \sigma^{-2}X'\Omega^{-1}X$ and $V_1^{-1}\beta_1 = V_0^{-1}\beta_0 + \sigma^{-2}X'\Omega^{-1}Xy$,
$\nu_1 = \nu_0 + n$ and $\nu_1 s_1^2 = \nu_0 s_0^2 + (y - X\beta)'\Omega^{-1}(y - X\beta)$,

Prior of ρ

By assuming that the prior of ρ is a truncated normal

$$\rho \sim N_{[-1,1]}(\rho_0, V_\rho),$$

then, its full conditional becomes

$$\begin{aligned} p(\rho|\epsilon) &\propto |\Omega_\rho|^{-1/2} \exp\left\{-0.5\epsilon'\Omega_\rho^{-1}\epsilon\right\} \\ &\times \exp\{-0.5(\rho^2 - 2\rho\rho_0)/V_\rho\} \\ &\times I(|\rho| < 1) \end{aligned}$$

Sampling ρ requires a Metropolis-Hastings step.

A small sin

Recall that

$$\epsilon_t = \rho\epsilon_{t-1} + u_t \qquad u_t \sim N(0, \sigma^2).$$

If we *pretend* that ϵ_0 is given and is equal to ϵ_1, and if we also let $\tilde{\epsilon} = (\epsilon_1, \ldots, \epsilon_{n-1})'$ and $\epsilon = (\epsilon_2, \ldots, \epsilon_n)'$, then

$$\rho | y, X, \beta, \sigma^2 \sim N_{[-1,1]}(\rho_1, \bar{V}_\rho)$$

where

$$\begin{aligned} \bar{V}_\rho &= (V_\rho^{-1} + \sigma^{-2}\tilde{\epsilon}'\tilde{\epsilon})^{-1} \\ \rho_1 &= \bar{V}_\rho^{-1}(V_\rho^{-1}\rho_0 + \sigma^{-2}\tilde{\epsilon}'\epsilon) \end{aligned}$$

Sampling from $Y \sim N_{[a,b]}(\mu, \sigma^2)$

Let $X \sim N(\mu, \sigma^2)$. Then, for $y \in [a, b]$, if follows that

$$
\begin{aligned}
u &= P(Y \leq y) = \frac{P(X \leq \frac{y-\mu}{\sigma}) - P(X \leq \frac{a-\mu}{\sigma})}{P(X \leq \frac{b-\mu}{\sigma}) - P(X \leq \frac{a-\mu}{\sigma})} \\
&= \frac{\Phi(\frac{y-\mu}{\sigma}) - A}{B - A}.
\end{aligned}
$$

Hence, y can be sampled from $N_{[a,b]}(\mu, \sigma^2)$ as a function of u sampled from $U(0, 1)$:

$$
y = \mu + \sigma\Phi^{-1}\left(u\Phi\left(\frac{b-\mu}{\sigma}\right) + (1-u)\Phi\left(\frac{a-\mu}{\sigma}\right)\right).
$$

Simulated example

Here we simulated $n = 200$ observations based on: $\rho = 0.9$, $\sigma^2 = 1.0$ and $\beta = (2.0, 4.0)$.

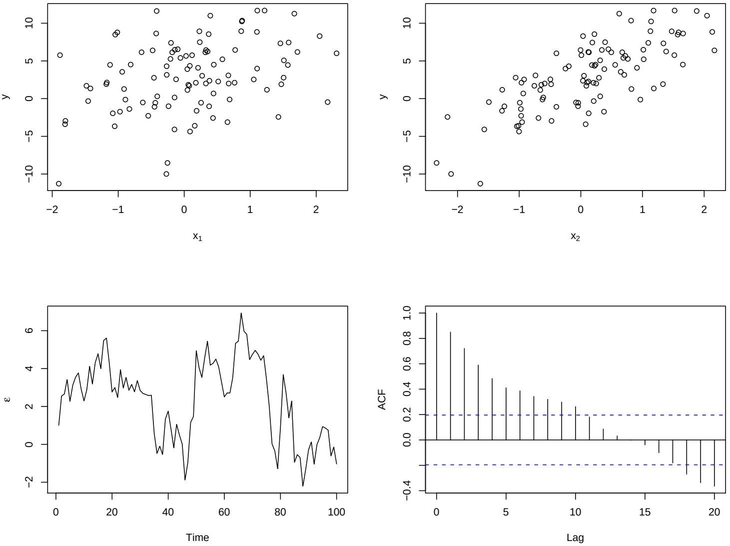
The prior hyperparameters are:

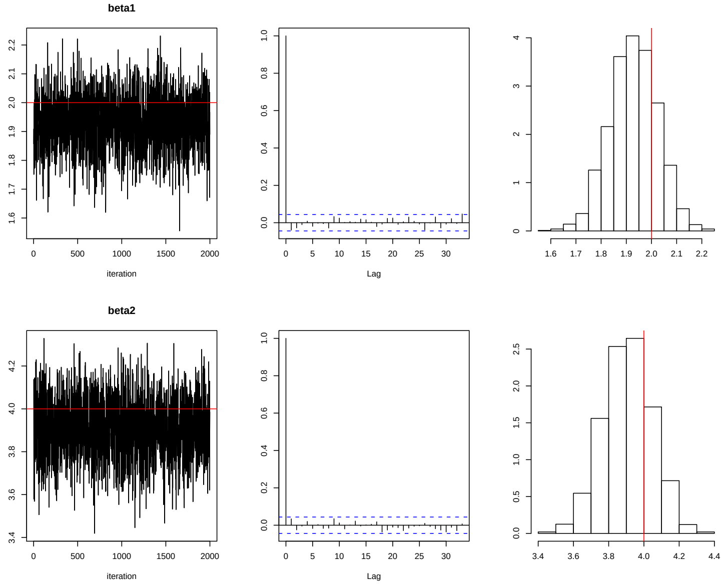
$$\beta_0 = 0 \quad V_0 = 100 I_2 \quad \nu_0 = 5 \quad s_0^2 = 1 \quad \rho_0 = 0.9 \quad \text{and} \quad V_\rho = 100$$

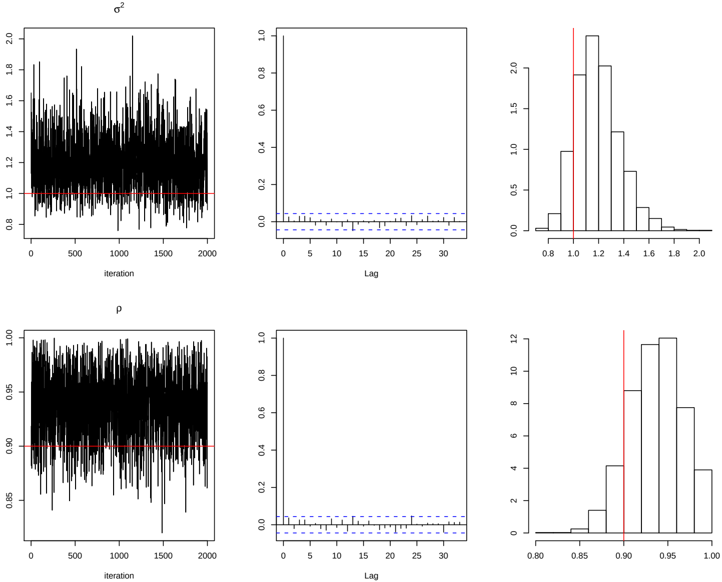
The initial values for the MCMC scheme are:

$$\beta = \hat{\beta}_{gls} = (1.94, 3.93) \quad \sigma^2 = \hat{\sigma}^2_{gls} = 1.12 \quad \rho = \rho_{gls} = 0.94$$

with burn-in of $M_0 = 1000$ and retaining every draw until $M = 2000$ draws are obtained for posterior inference.


y
x_1
y
x_2
ε
Time
ACF
Lag


beta1
iteration
Lag
beta2
iteration
Lag


σ²
iteration
Lag
ρ
iteration
Lag

Seemingly unrelated regressions

Assume that

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

for $m = 1, \ldots, M$ equations and $i = 1, \ldots, n$ observations.

Error structure: $\epsilon_{mi} \sim N(0, \sigma_i^2)$ and $Cov(\epsilon_{mi}, \epsilon_{\tilde{m}i}) = \sigma_{ij}$.

Let $y_i = (y_{1i}, \ldots, y_{Mi})'$, $\epsilon_i = (\epsilon_{1i}, \ldots, y_{Mi})'$, $\beta = (\beta_1', \ldots, \beta_M')'$, and $X_i = \text{diag}(x_{1i}', \ldots, x_{Mi}')$. Therefore

$$y_i = X_i\beta + \epsilon_i \qquad \epsilon_i \sim N(0, \Sigma)$$

and

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

Prior and full conditionals

The standard conditionally conjugate prior for (β, Σ) is given by $p(\beta, \Sigma) = p(\beta)p(\sigma)$, where

$$\beta \sim N(\beta_0, V_0) \quad \text{and} \quad \Sigma \sim IW(\nu_0, \Sigma_0)$$

The full conditional distributions are

$$\begin{aligned} \beta|y, X, \Sigma &\sim N(\beta_1, V_1) \\ \Sigma|y, X, \beta &\sim IW(\nu_1, \Sigma_1) \end{aligned}$$

where $\nu_1 = n + \nu_0$,

$$\begin{aligned} V_1^{-1} &= V_0^{-1} + \sum_{i=1}^{n} X_i' \Sigma X_i \\ V_1^{-1}\beta_1 &= V_0^{-1}\beta_0 + \sum_{i=1}^{n} X_i' \Sigma^{-1} y_i \\ \Sigma_1^{-1} &= \Sigma_0^{-1} + \sum_{i=1}^{n} (y_i - X_i\beta)(y_i - X_i\beta)' \end{aligned}$$