Lecture 23 5/16/16 Math 341.03-08

Linear Regression Under OLS. Ridge priors but σ^2 unknown!

Lik: $Y_1 \sim N(\beta_0 + \beta_1 x_1, \sigma^2), \ldots, Y_n \sim N(\beta_0 + \beta_1 x_n, \sigma^2)$

$\beta_0, \beta_1 \overset{iid}{\sim} N\left(0, \frac{\sigma^2}{m}\right)$

$$Q := \sum \left(y_i^2 + \beta_0^2 + \beta_1^2 x_i^2 - 2y_i\beta_0 - 2\beta_1 y_i x_i + 2\beta_0\beta_1 x_i\right)$$

Now $\sigma^2 \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$

$$P(\beta_0, \beta_1, \sigma^2 \mid X, y) \propto P(Y \mid X, \beta_0, \beta_1, \sigma^2)\, P(\beta_0, \beta_1, \sigma^2)$$

$$\propto (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2} \underbrace{\frac{1}{\sqrt{2\pi\frac{\sigma^2}{m}}}}_{(\sigma^2)^{-\frac{1}{2}}} e^{-\frac{m}{2\sigma^2}\beta_0^2} \underbrace{\frac{1}{\sqrt{2\pi\frac{\sigma^2}{m}}}}_{(\sigma^2)^{-\frac{1}{2}}} e^{-\frac{m}{2\sigma^2}\beta_1^2} (\sigma^2)^{-\left(\frac{n_0}{2}+1\right)} e^{-\frac{n_0\sigma_0^2}{2\sigma^2}}$$

$$-\left(\frac{n+n_0}{2} + 1 + 1\right)$$

$$\propto (\sigma^2)^{-\frac{n}{2} - 1 - \frac{n_0}{2} - 1} \quad e^{-\frac{1}{2\sigma^2}\left((Q) + m\beta_0^2 + m\beta_1^2 + n_0\sigma_0^2\right)} \quad \text{MESS!}$$

But...

$$P(\sigma^2 \mid \beta_0, \beta_1, X, y) \propto \text{InvGamma}\left(\frac{n+n_0}{2} + 1, \frac{\sum(y_i - \beta_0 - \beta_1 x_i)^2 + m(\beta_0^2 + \beta_1^2) + n_0\sigma_0^2}{2}\right)$$

$$P(\beta_0 \mid \sigma^2, \beta_1, X, y) \propto e^{-\frac{1}{2\sigma^2}\left(n\beta_0^2 - 2\bar{y}n\beta_0 + 2\beta_0\beta_1\bar{x}n + m\beta_0^2\right)}$$

$$= e^{\underbrace{-\frac{n+m}{2\sigma^2}}_{a}\beta_0^2 + \underbrace{\frac{\bar{y}n - \beta_1\bar{x}n}{\sigma^2}}_{b}\beta_0} \propto N\left(\frac{\bar{y}n - \beta_1\bar{x}n}{n+m}, \frac{\sigma^2}{n+m}\right)$$

$$e^{-\frac{1}{2v}(\beta_0 - c)^2} \propto e^{-\frac{\beta_0^2}{2v}} e^{\frac{\beta_0 c}{v}}$$

$$a = -\frac{1}{2v} \Rightarrow v = -\frac{1}{2a} = -\frac{1}{2\left(-\frac{n+m}{2\sigma^2}\right)} = \frac{\sigma^2}{n+m}$$

$$b = \frac{c}{v} \Rightarrow c = bv = \frac{\bar{y}n + \beta_1\bar{x}n}{\sigma^2} \cdot \frac{\sigma^2}{n+m} = \frac{\bar{y}n - \beta_1\bar{x}n}{n+m}$$

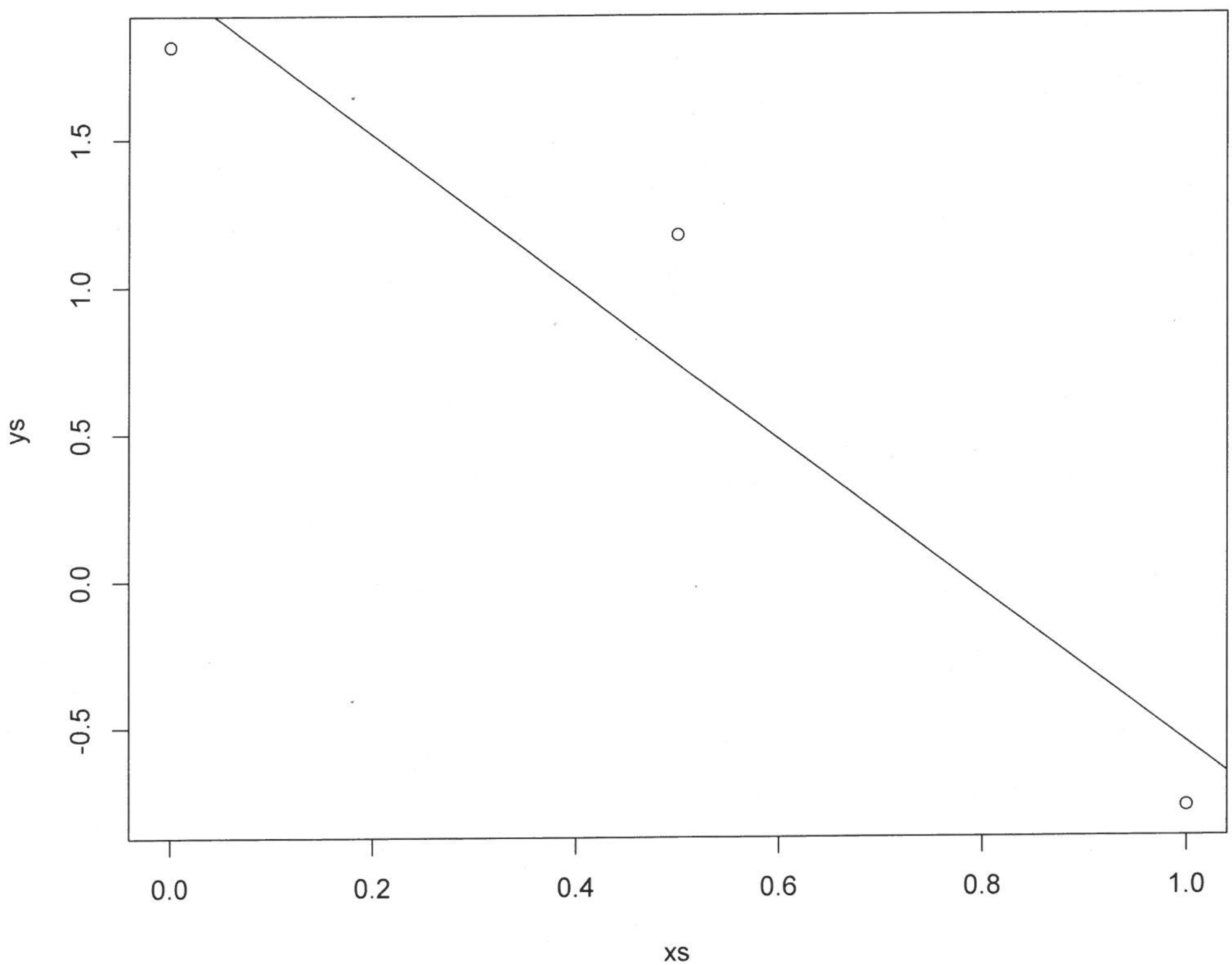
$$p(\beta_1 \mid \sigma^2, \beta_0, X, y) \propto e^{-\frac{1}{2\sigma^2}\left(\beta_1^2 \Sigma x_i^2 - 2\beta_1 \Sigma x_i y_i + 2\beta_1 \beta_0 \bar{x} n + n\beta_1^2\right)}$$

$$e^{-\frac{n+\Sigma x_i^2}{2\sigma^2}\beta_1^2 + \frac{\Sigma x_i y_i - \beta_0 \bar{x} n}{\sigma^2}\beta_1}$$

Same square completion as before...

$$\propto N\left(\frac{\Sigma x_i y_i - \beta_0 \bar{x} n}{\Sigma x_i^2 + n}, \frac{\sigma^2}{\Sigma x_i^2 + n}\right)$$

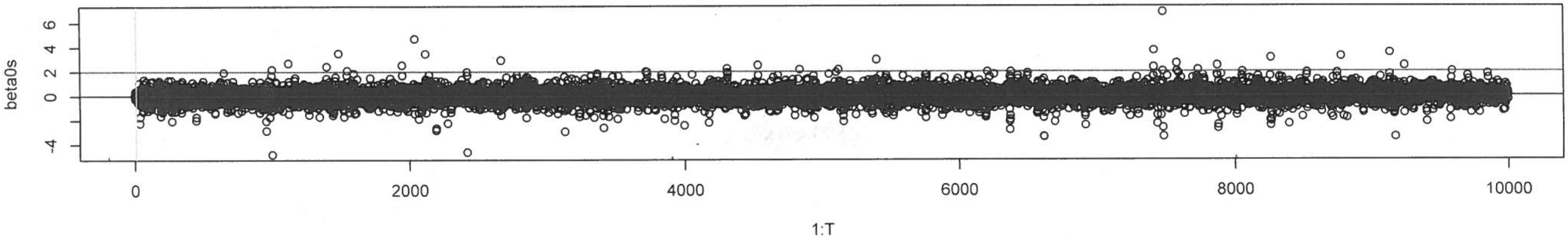
Nearly
Impossible $\Rightarrow$ possible!

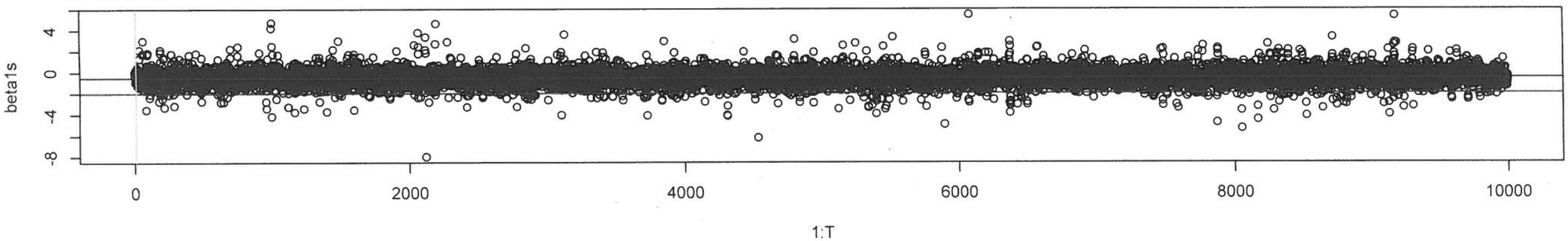


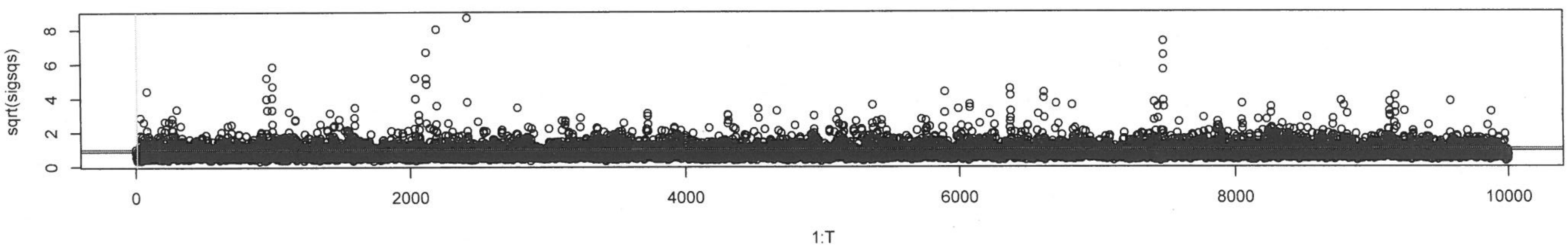
1.5
1.0
0.5
0.0
-0.5
ys
0.0
0.2
0.4
0.6
0.8
1.0
xs

$m=1,\ \nu_0=1,\ \sigma_0^2=1$ hyperparams $\rightarrow$ objective

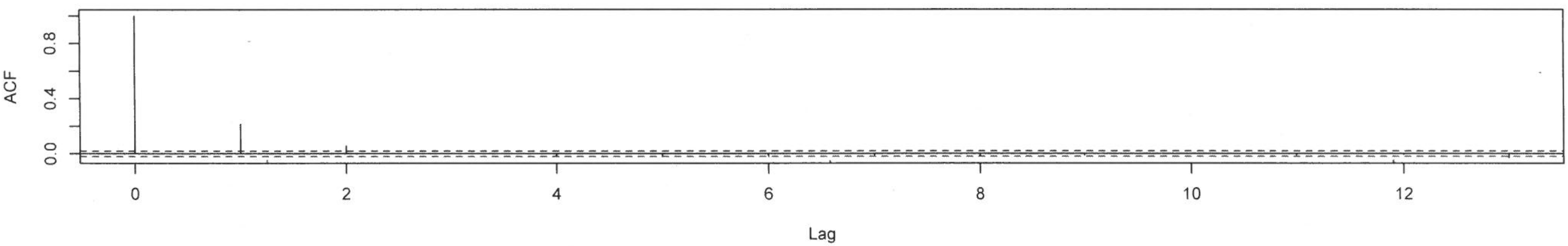
$n=3$ $\boxed{\beta_0=2,\ \beta_1=-2,\ \sigma^2=1}$ truth



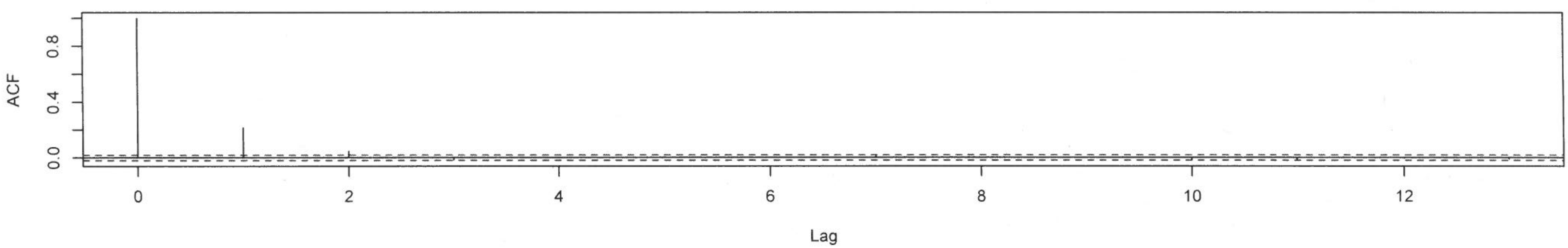




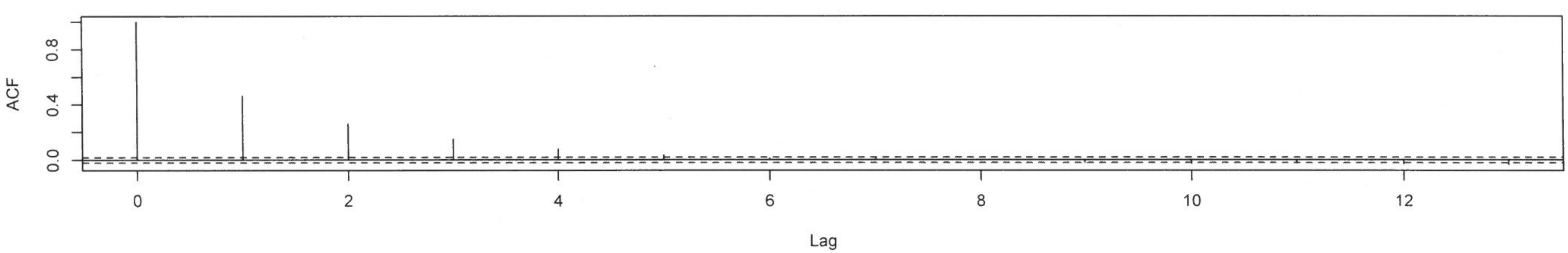
Series beta0s[B:T]

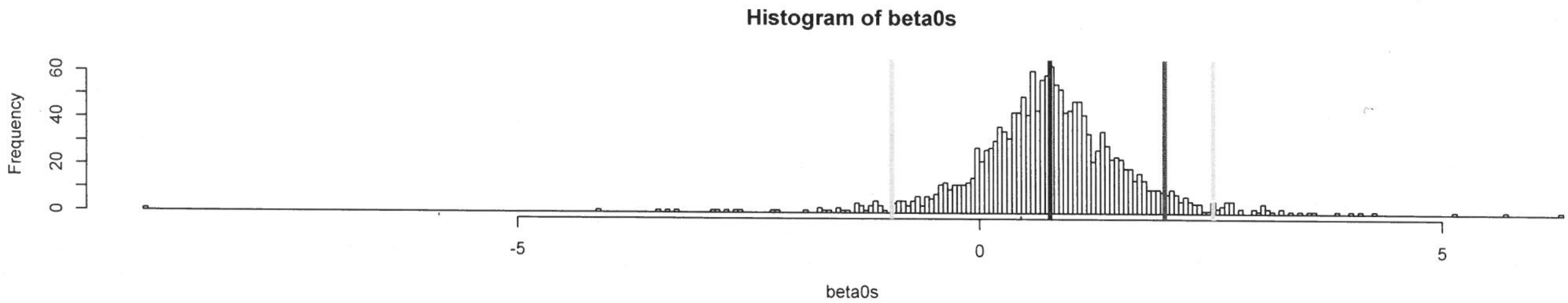


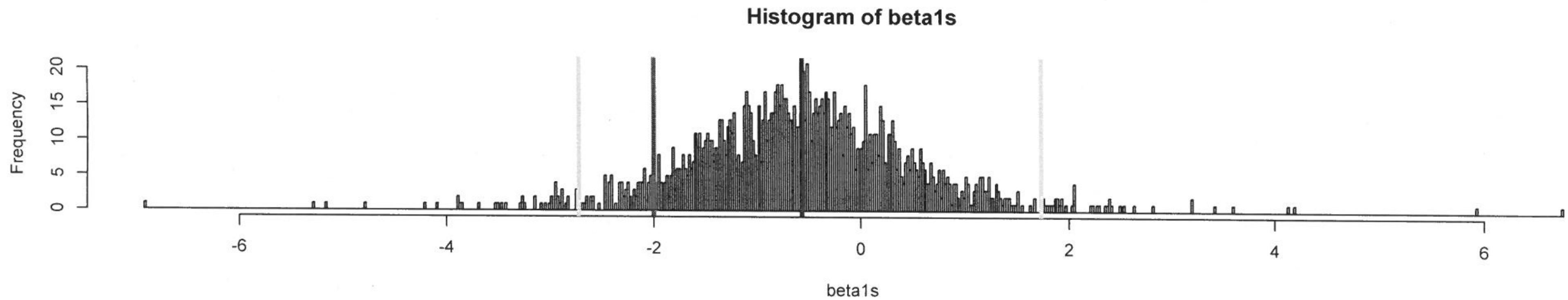
Series beta1s[B:T]

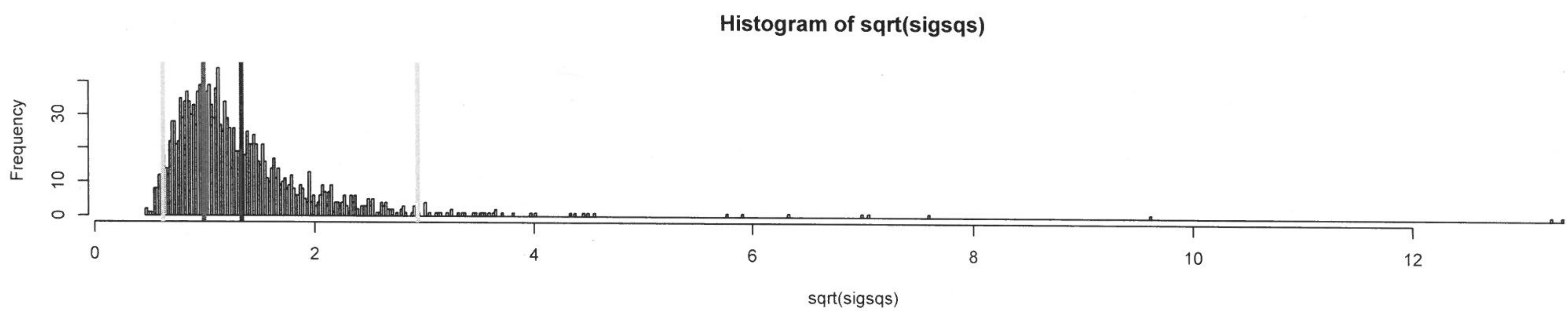


Series sqrt(sigsqs[B:T])




Histogram of beta0s
Frequency
0
20
40
60
-5
0
5
beta0s


Histogram of beta1s
Frequency
0
5
10
15
20
-6
-4
-2
0
2
4
6
beta1s


Histogram of sqrt(sigsqs)
Frequency
0
10
30
0
2
4
6
8
10
12
sqrt(sigsqs)

Another model:

Imagine

stopping pt where model changes

$$\underbrace{\_\ \_\ \_\ \_\ \cdots\ \underset{m}{\_}}\ \ \underbrace{\underset{m+1}{\_}\ \_\ \_\ \_\ \underset{n}{\_}}$$

Lik: iid Poisson(λ_1) iid Poisson(λ_2)

Priors: $\lambda_1, \lambda_2 \overset{iid}{\sim} \text{Gamma}(\alpha, \beta)$

$$P(\lambda_1 | X_1, \ldots, X_m) = \text{Gamma}\left(\sum_{i=1}^{m} X_i + \alpha, m + \beta\right)$$

$$P(\lambda_2 | X_{m+1}, \ldots, X_n) = \text{Gamma}\left(\sum_{i=m+1}^{n} X_i + \alpha, n-m + \beta\right)$$

What if stopping pt. unknown? $m \sim U(0, \ldots, n)$

$$P(\lambda_1, \lambda_2, m | X_1, \ldots, X_m, X_{m+1}, \ldots, X_n) \propto P(X_1, \ldots, X_n | \lambda_1, \lambda_2, m)\, P(\lambda_1, \lambda_2, m)$$

$$= P(X_1, \ldots, X_m | \lambda_1)\, P(X_{m+1}, \ldots, X_n | \lambda_2)\, P(\lambda_1)\, P(\lambda_2)\, P(m)$$

$$\propto \prod_{i=1}^{m} \frac{e^{-\lambda_1} \lambda_1^{X_i}}{X_i!} \prod_{i=m+1}^{n} \frac{e^{-\lambda_2} \lambda_2^{X_i}}{X_i!}\, \lambda_1^{\alpha-1} e^{-\beta\lambda_1}\, \lambda_2^{\alpha-1} e^{-\beta\lambda_2}\, \frac{1}{n+1}$$

$$\propto e^{-m\lambda_1} e^{-(n-m)\lambda_2 - \beta\lambda_1 - \beta\lambda_2}\, \lambda_1^{\alpha-1+\sum_{i=1}^{m} X_i}\, \lambda_2^{\alpha-1+\sum_{i=m+1}^{n} X_i}$$

$$p(\lambda_1 \mid X, \lambda_2, m) \propto e^{\overbrace{-m\lambda_1 - \beta\lambda_1}^{-(m+\beta)\lambda_1}} \lambda_1^{\alpha-1+\sum_{i=1}^{m} x_i} \propto \text{Gamm}\left(\alpha + \sum_{i=1}^{m} x_i, m+\beta\right)$$

$$p(\lambda_2 \mid X, \lambda_1, m) \quad - - - - \quad \propto \text{Gamm}\left(\alpha + \sum_{i=m+1}^{n} x_i, n-m+\beta\right)$$

$$p(m \mid X, \lambda_1, \lambda_2) \propto e^{-\overbrace{(m\lambda_1 + (n-m)\lambda_2)}^{m\lambda_1 + n\lambda_2 - m\lambda_2}} \lambda_1^{\sum x_i} \lambda_2^{\sum x_i} / \prod x_i!$$

$$\propto e^{m(\lambda_2 - \lambda_1)} \cdots$$

Not a std distr. but easily grid sampled since $\text{Supp}(m) = \{0, \dots, n\}$ with no error due to finite support.

~~Up next... (a) Em within Gibbs (b) Gibbs steps where you do not have the ability to sample $\Rightarrow$ need sol'n $\Rightarrow$ Metropolis Hastings rejection sampling (out of course)~~

~~Em-within-Gibbs~~

~~$$p(X \mid \dots) = \rho N(\theta_0, \sigma_0^2) + (1-\rho) N(\theta_1, \sigma_1^2)$$~~

~~$\theta_0 \sim N\left(\mu_0, \frac{\sigma_0^2}{m_0}\right)$~~

~~$\theta_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{m_1}\right)$~~

~~$\sigma_0^2 \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0 a^2}{2}\right)$~~

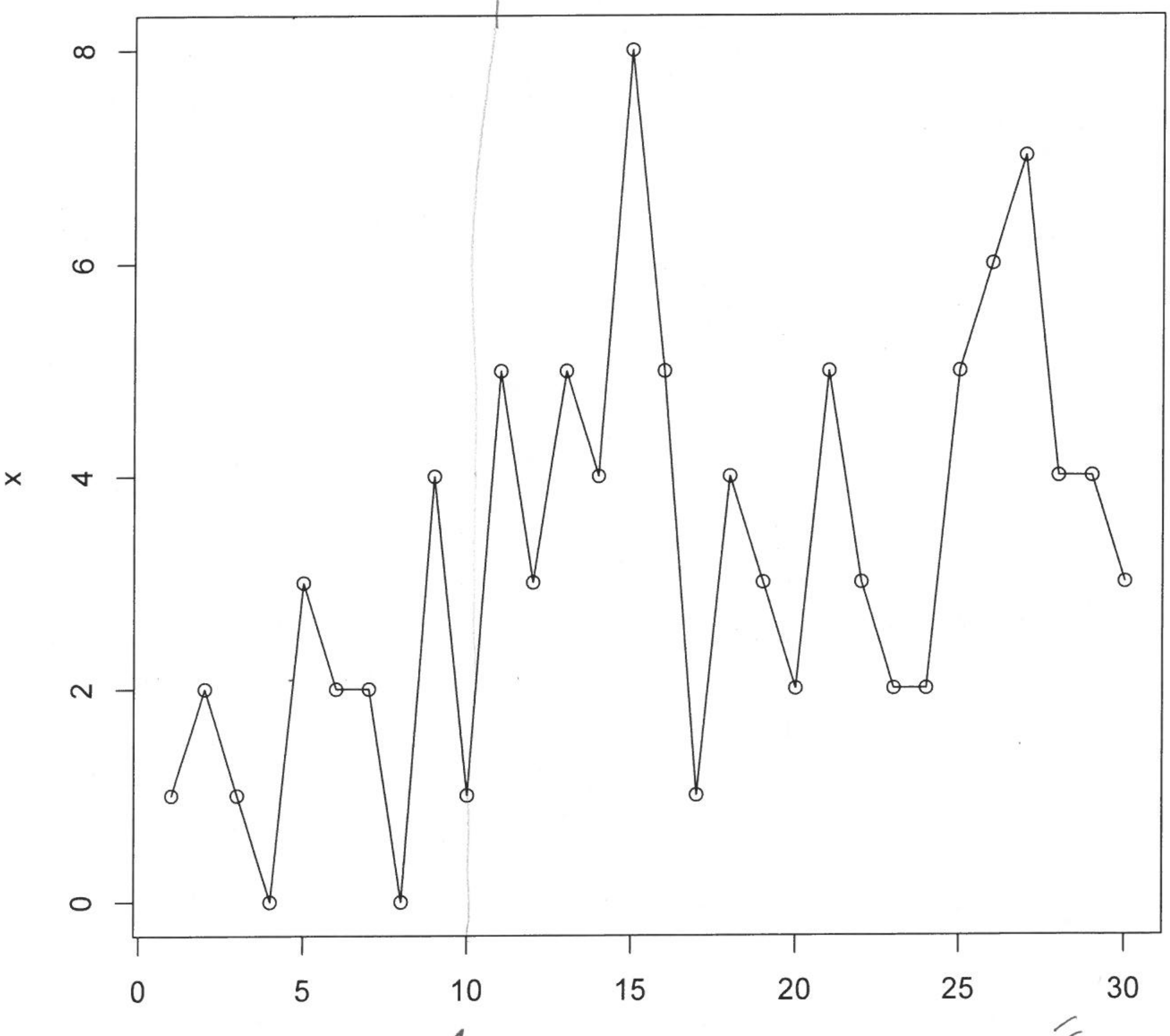
~~$\sigma_1^2 \sim \text{InvGamma}\left(\frac{n_1}{2}, \frac{n_1 b^2}{2}\right)$~~

~~$\rho \sim \text{Beta}(\alpha, \beta)$~~

~~Assume $I_1, \dots, I_n$ known~~

COURSE OVER

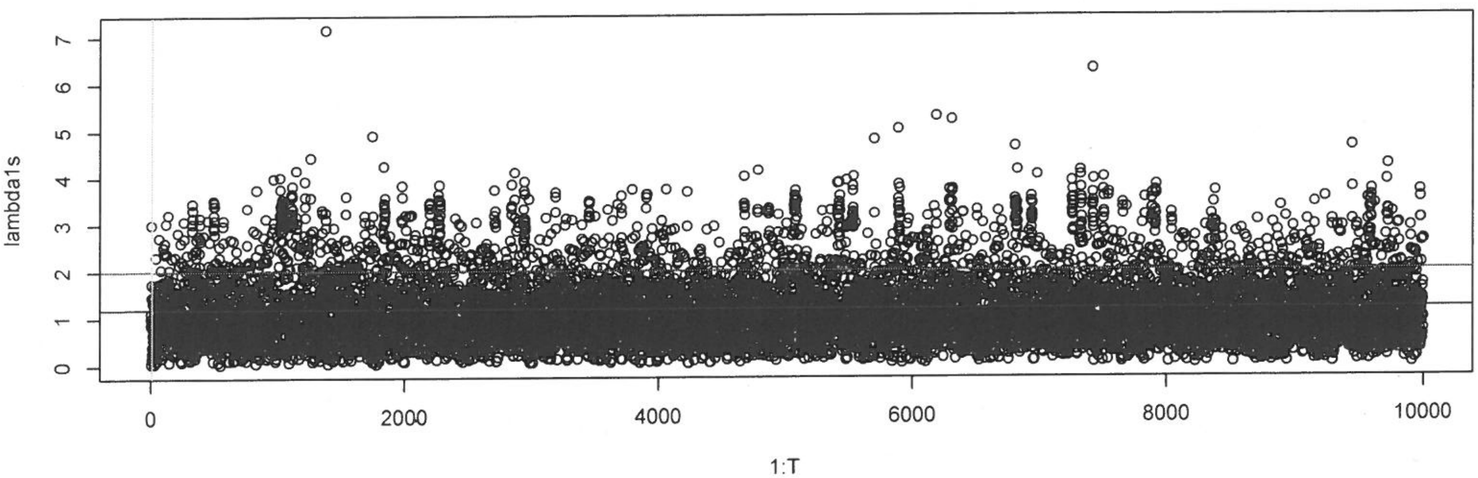
$\lambda_1 = 2$ $\lambda_2 = 4$

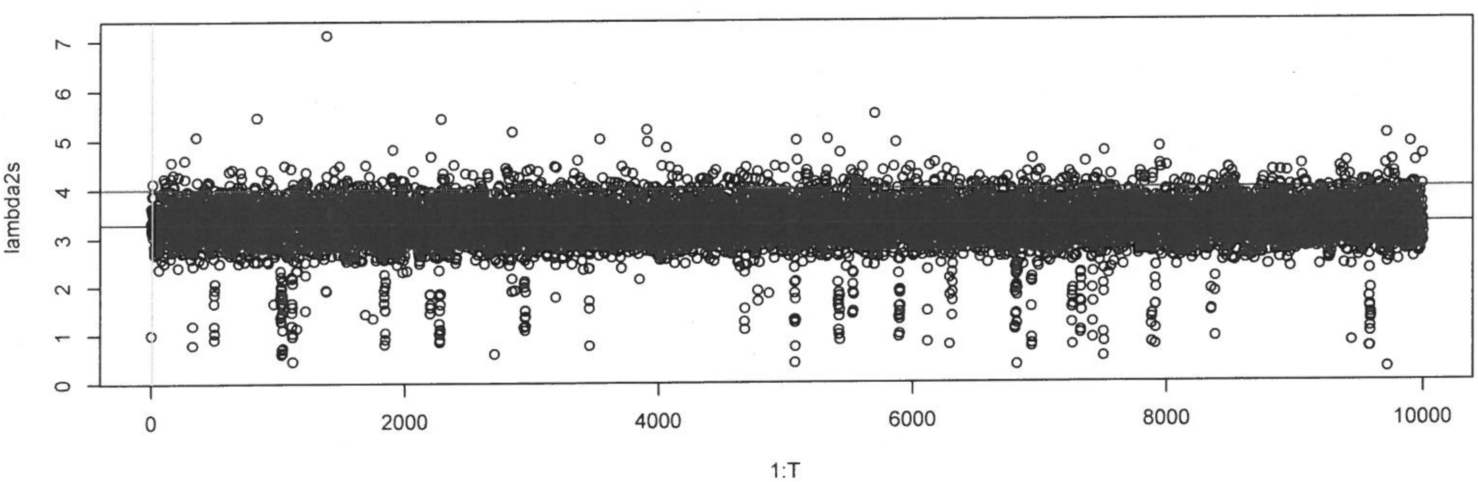


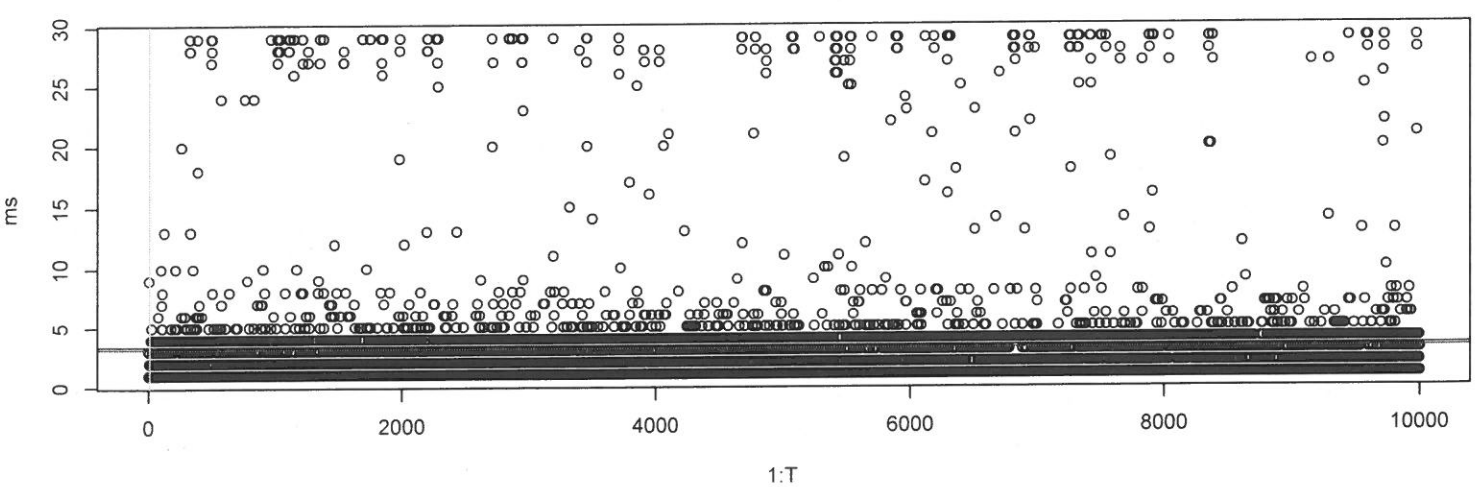
$\lambda =$

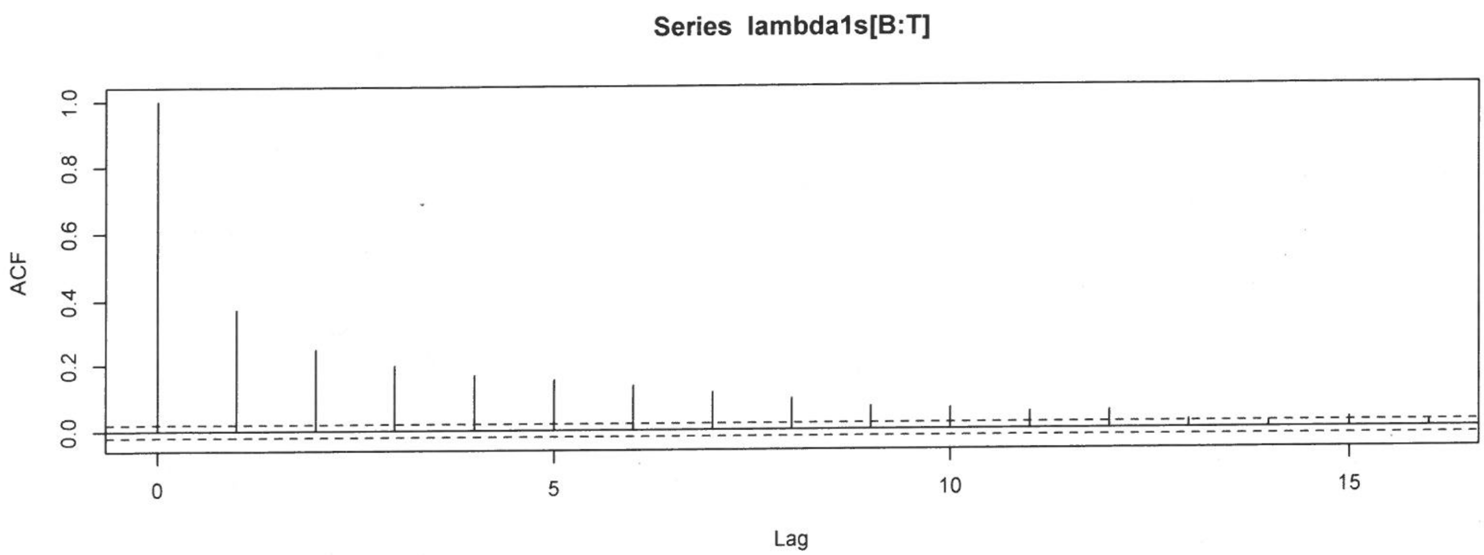
true change pt (m)

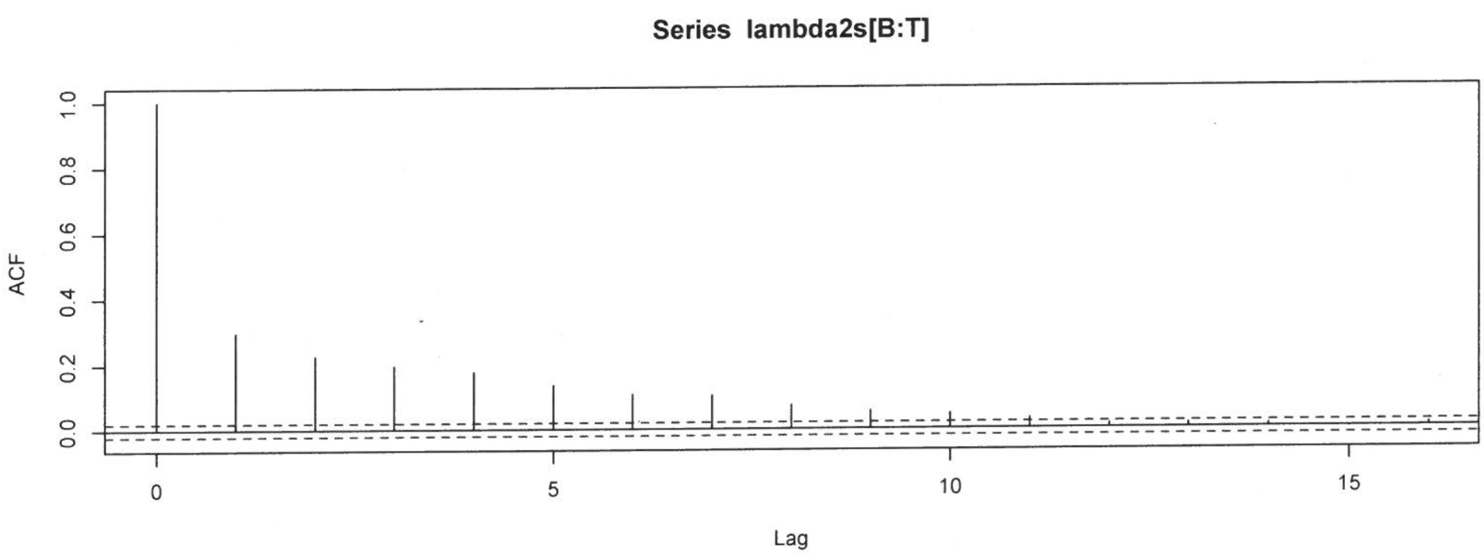
$\alpha = 1$, $\beta = 1$ } objective priors

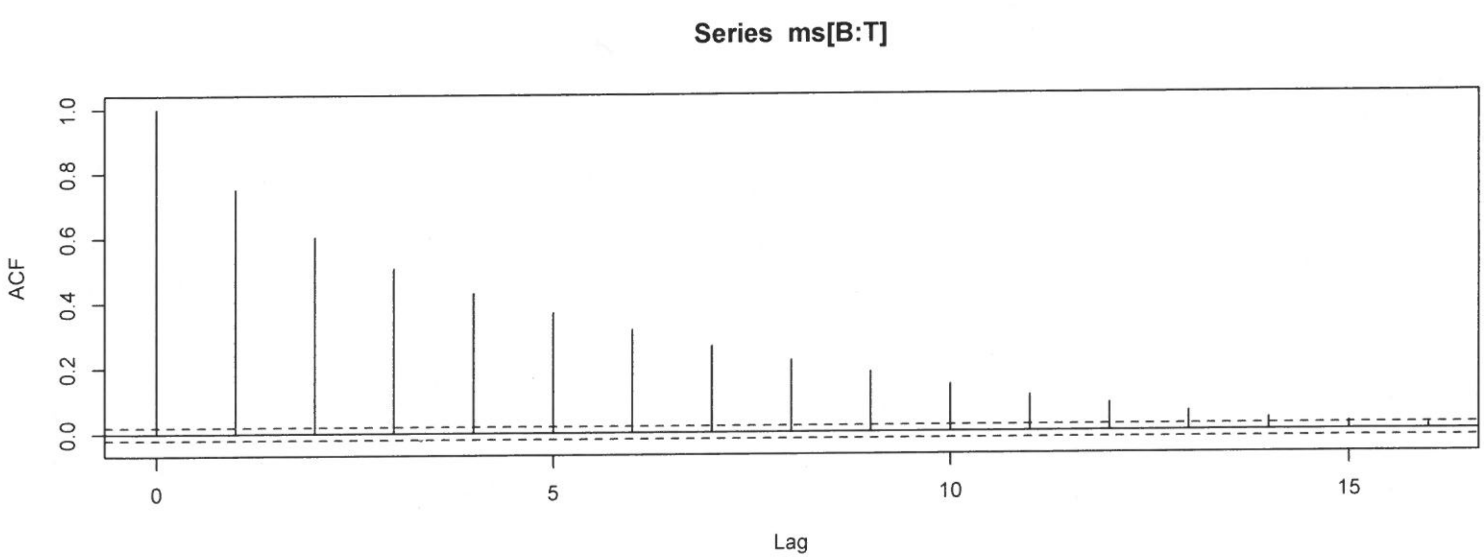

lambda1s
0
1
2
3
4
5
6
7
0
2000
4000
6000
8000
10000
1:T

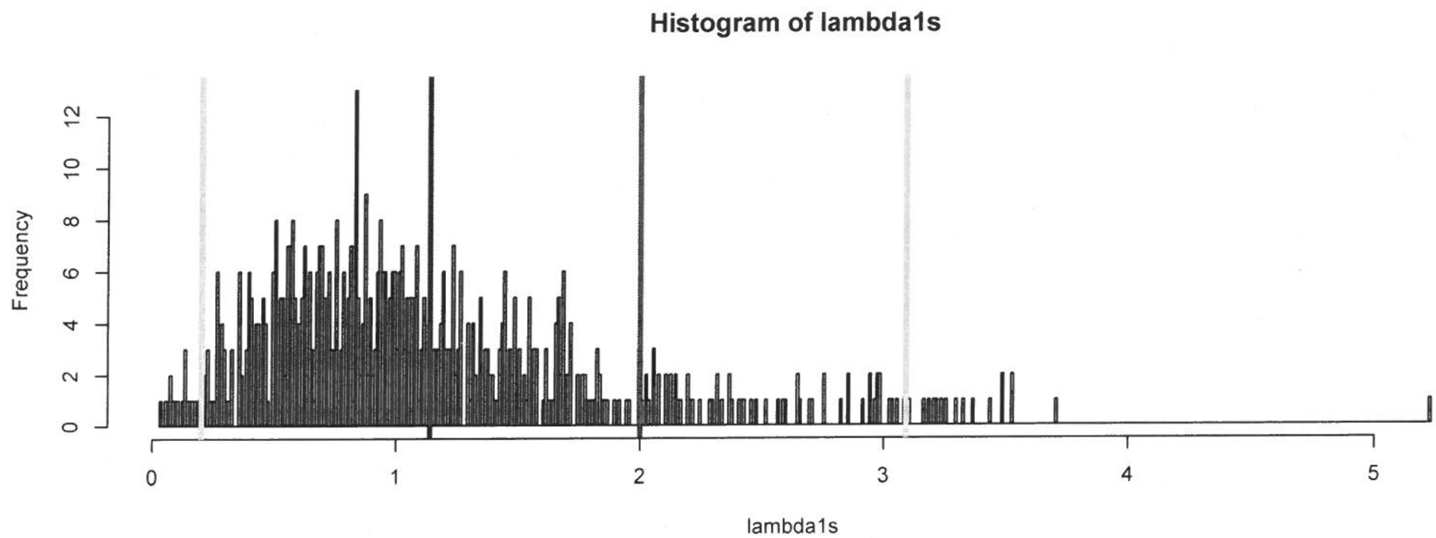

lambda2s
0
1
2
3
4
5
6
7
0
2000
4000
6000
8000
10000
1:T


ms
0
5
10
15
20
25
30
0
2000
4000
6000
8000
10000
1:T

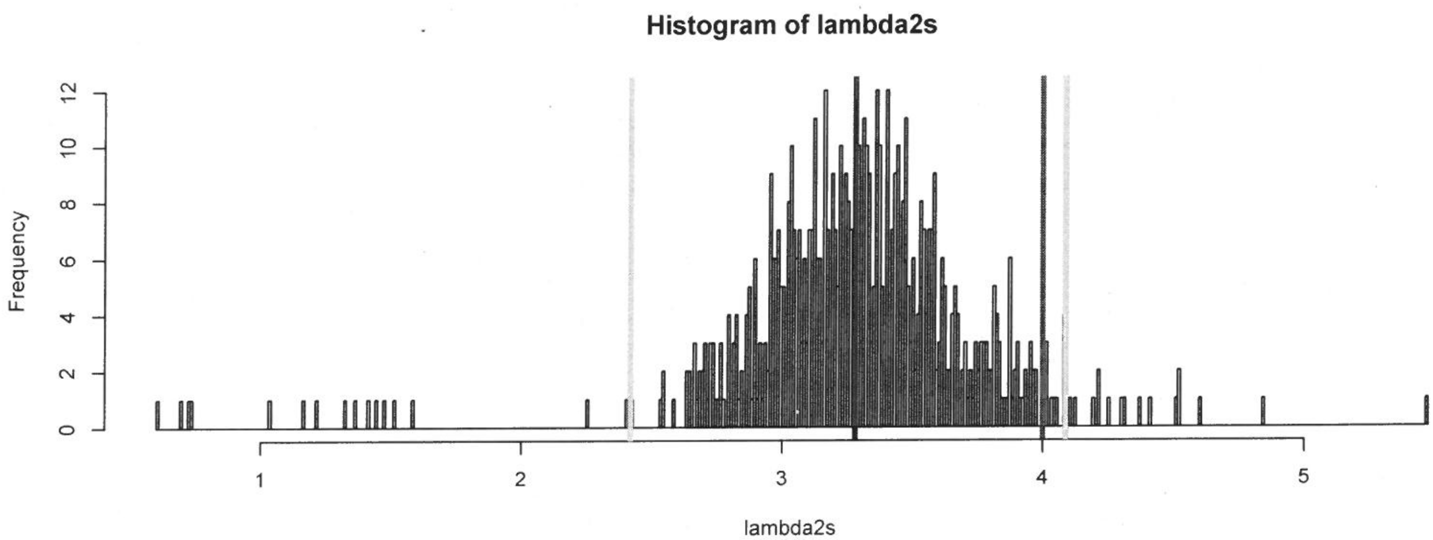

Series lambda1s[B:T]
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
Lag


Series lambda2s[B:T]
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
Lag

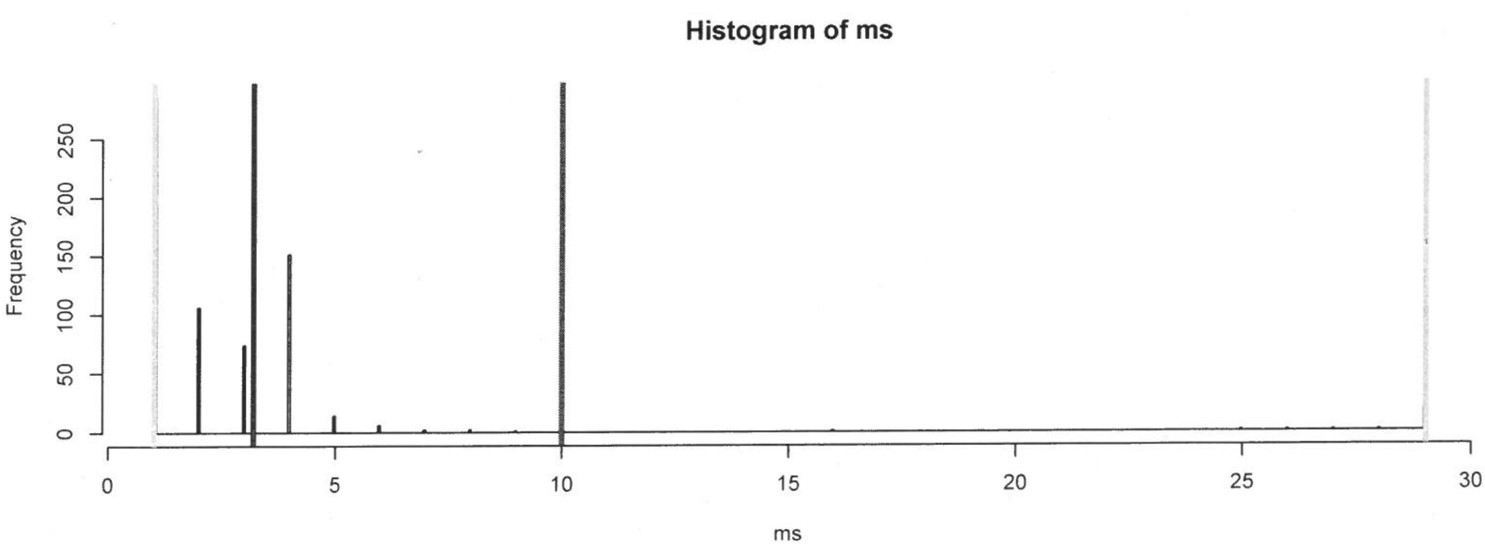

Series ms[B:T]
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
Lag



Histogram of lambda1s
Frequency
0
2
4
6
8
10
12
0
1
2
3
4
5
lambda1s



Histogram of lambda2s
Frequency
0
2
4
6
8
10
12
1
2
3
4
5
lambda2s



Histogram of ms
Frequency
0
50
100
150
200
250
0
5
10
15
20
25
30
ms

Imagine a model s.t.

t	y
1	6
1.3	0
2	3
1	7
3.1	1
4	8
⋮	⋮

$$y_t = \text{Poisson}(\lambda = \beta_0 + \beta_1 t)$$

(generalized linear model with Poisson link function / Poisson regression)

Lik:

$$P(Y|\beta_0,\beta_1) = \prod_{i=1}^{n} \frac{(\beta_0+\beta_1 t_i)^{y_i} e^{-(\beta_0+\beta_1 t_i) y_i}}{y!}$$

Assume flat prior $P(\beta_0) \propto 1, P(\beta_1) \propto 1$

$$\Rightarrow P(\beta_0,\beta_1|Y) \propto \prod_{i=1}^{n} (\beta_0+\beta_1 t_i)^{y_i} e^{-(\beta_0+\beta_1 t_i) y_i}$$

which is by no means a std distr.

$P(\beta_0|\beta_1,Y) \propto$ ditto, $P(\beta_1|\beta_0,Y) \propto$ ditto

What to do?

w/ Edward Teller!!!

Use "Metropolis Hastings" Algorithm (Metropolis, 1953 and Hastings, 1970)

Step 1: Initialize $\beta_{0,0}, \beta_{0,1}$

free var, previous value, other params

Step 2: Draw $\beta_{1,0}$ from $q(\beta_{1,0}; \beta_{0,0}, \theta)$

"candidate density" e.g. $N(\beta_{0,0}, \sigma^2=1)$

posterior likelihood ratio of the transition

Step 3: Calc:

$$r := \frac{P(\beta_0=\beta_{1,0}, \beta_1=\beta_{0,1}|Y) \,/\, q(\beta_{1,0}; \beta_{0,0}, \sigma^2)}{P(\beta_0=\beta_{0,0}, \beta_1=\beta_{0,1}) \,/\, q(\beta_{0,0}; \beta_{1,0}, \sigma^2)}$$

← prob of transitioning forward

← prob. of transitioning backward

Step 4: Accept prob w/ r. (If $r \geq 1 \Rightarrow$ Accept)

Step 5-7 Ditto with β_1

Step 8: Iterate steps 2-7 until convergence.

How to pick q? ≈ 50 yr of research on this

Metropolis-within-Gibbs

$P(\theta_1 | \theta_{-1}) \sim$ sample from known distr.

$P(\theta_2 | \theta_{-2}) \sim$ sample " " " "

$P(\theta_3 | \theta_{-3})$ cannot be sampled, so use M-H here

$P(\theta_4 | \theta_{-4}) \sim$ sample " " " "

You can also do E-M within Gibbs

Ex. Gibbs sampling is a special case of M-H. Imagine θ_1, θ_2

$$P(\theta_1, \theta_2 | X) \propto P(X | \theta_1, \theta_2) P(\theta_1, \theta_2)$$

$$r = \frac{P(\theta_1 = \theta_{1,1}, \theta_2 = \theta_{0,2} | X) \;/\; P(\theta_1 = \theta_{1,1} | \theta_2 = \theta_{0,2}, X)}{P(\theta_1 = \theta_{0,1}, \theta_2 = \theta_{0,2} | X) \;/\; P(\theta_1 = \theta_{0,1} | \theta_2 = \theta_{0,2}, X)}$$

q

Candidate density is cond. density and doesn't use prev value

L7

$$r = \frac{\cancel{P(\theta_1=\theta_{1,1} \mid \theta_2=\theta_{0,2}, X)}\ \cancel{P(\theta_2=\theta_{0,2} \mid X)} \ \big/\ \cancel{P(\theta_1=\theta_{1,1} \mid \theta_2=\theta_{0,2}, X)}}{\cancel{P(\theta_1=\theta_{0,1} \mid \theta_2=\theta_{0,2}, X)}\ \cancel{P(\theta_2=\theta_{0,2} \mid X)} \ \big/\ \cancel{P(\theta_1=\theta_{0,1} \mid \theta_2=\theta_{0,2}, X)}} = 1$$

$\Rightarrow$ Accept all the time $\Rightarrow$ exactly what a Gibbs sampler does...

$\sigma^2 = \sqrt{0.5}$

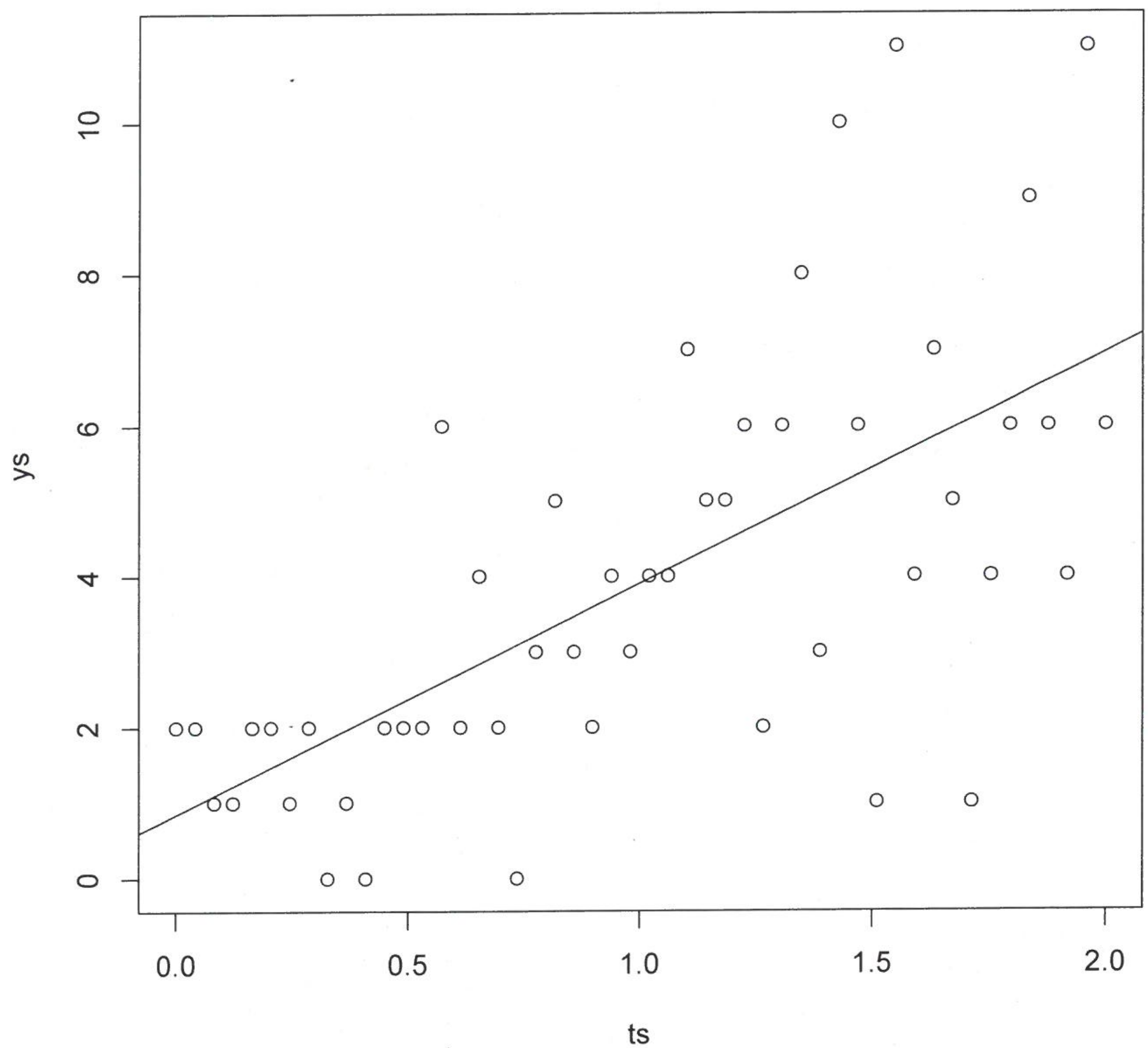
$\beta_0 = 1$

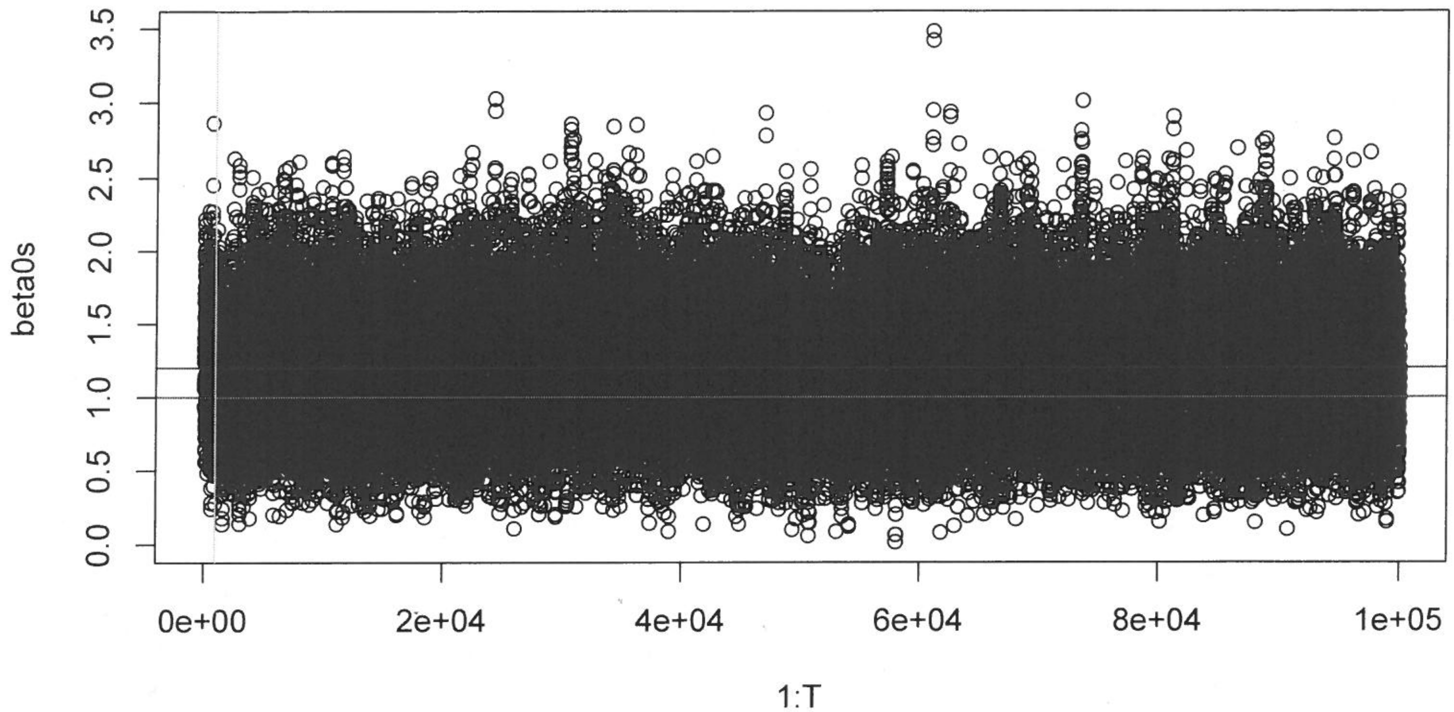
$\beta_1 = 3$

$n = 50$

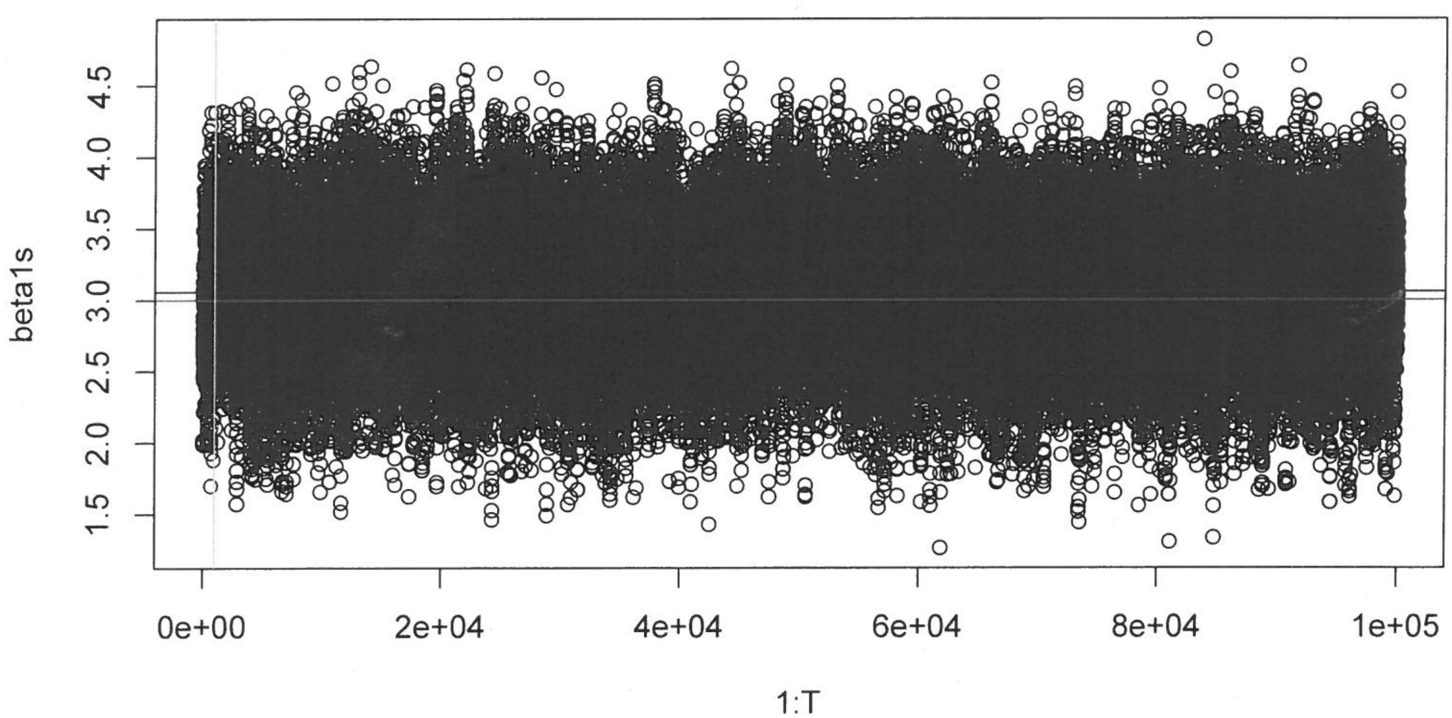
$\beta_{0,0} = 2$

$\beta_{0,1} = 2$

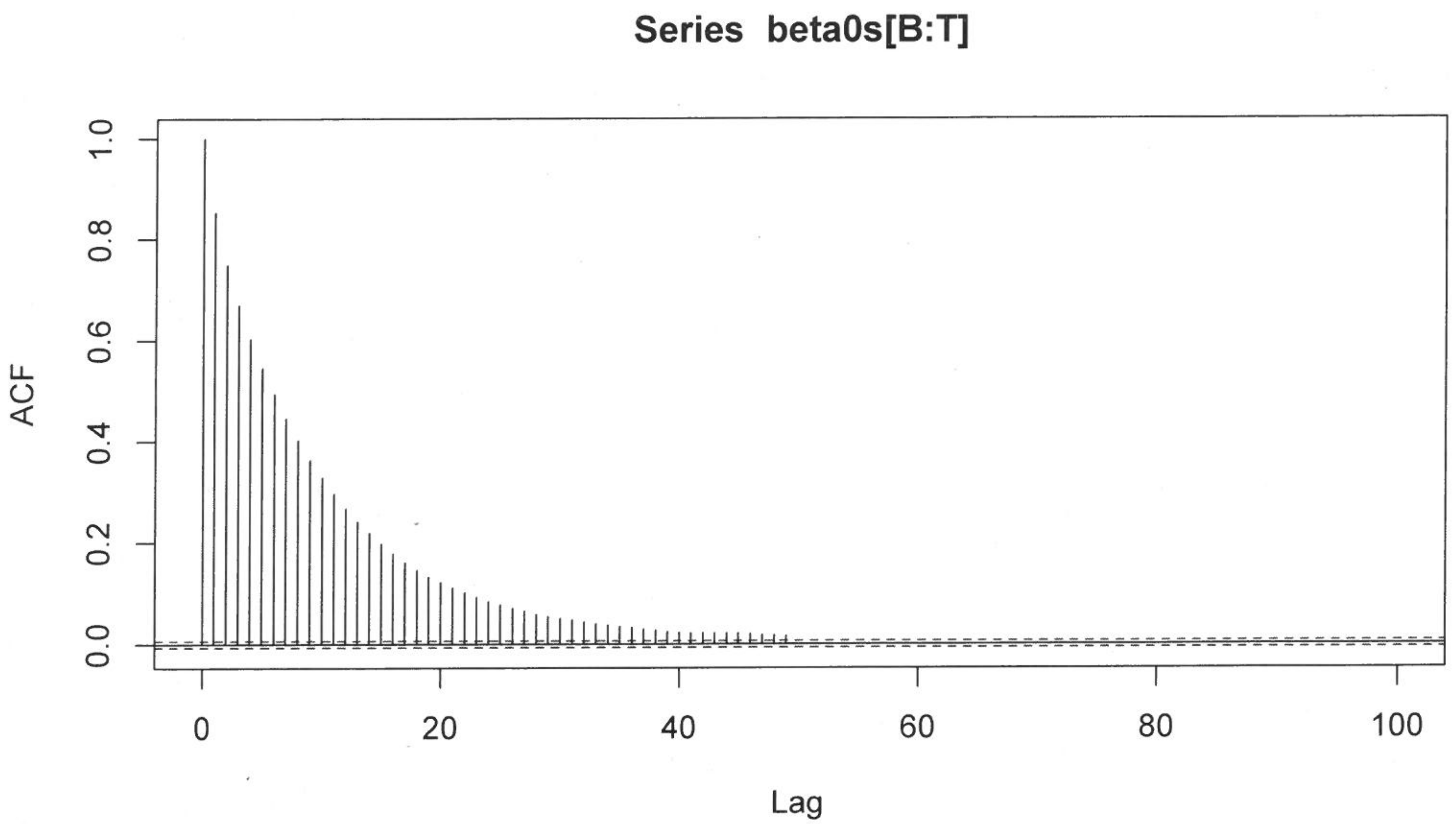


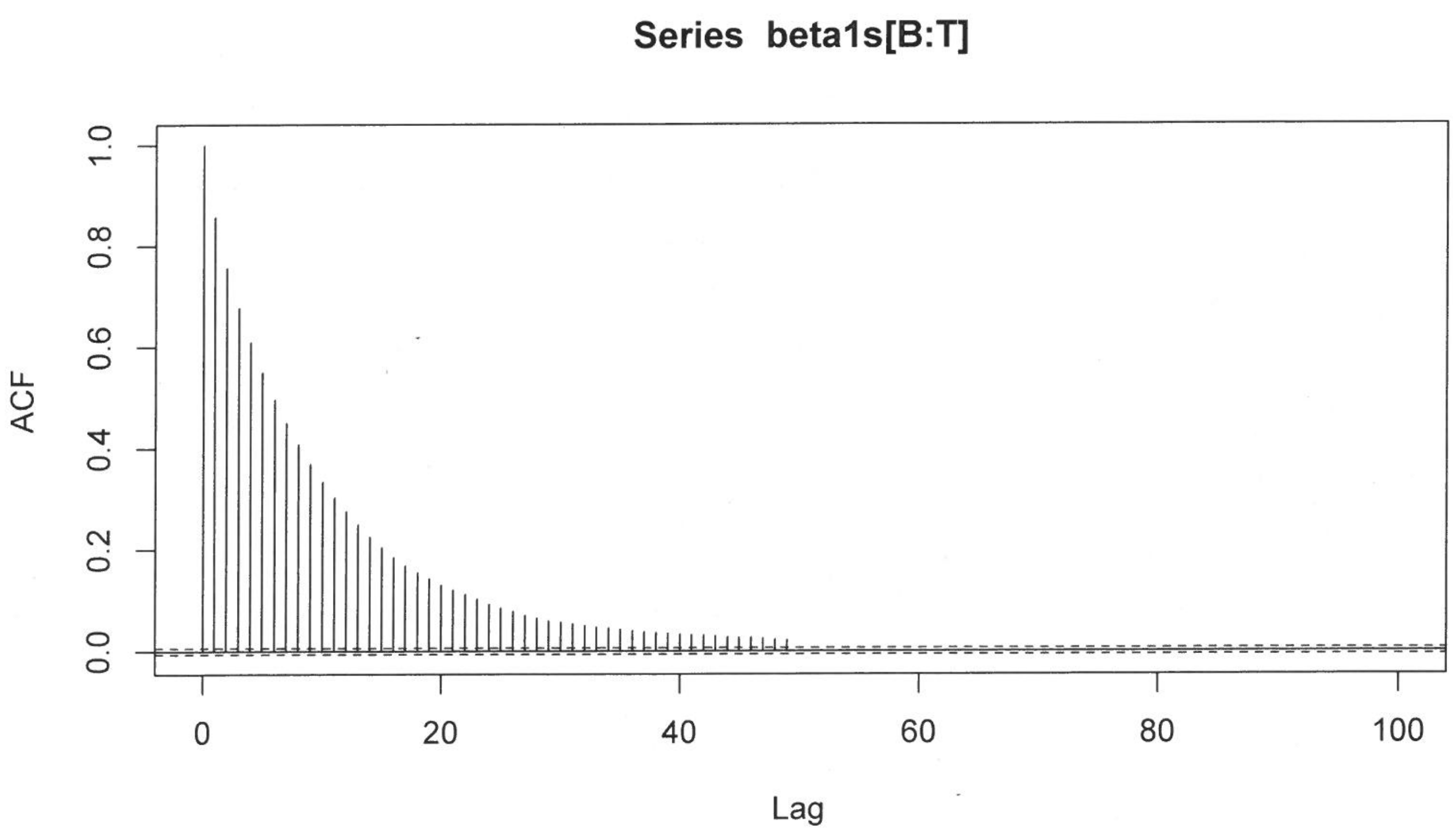


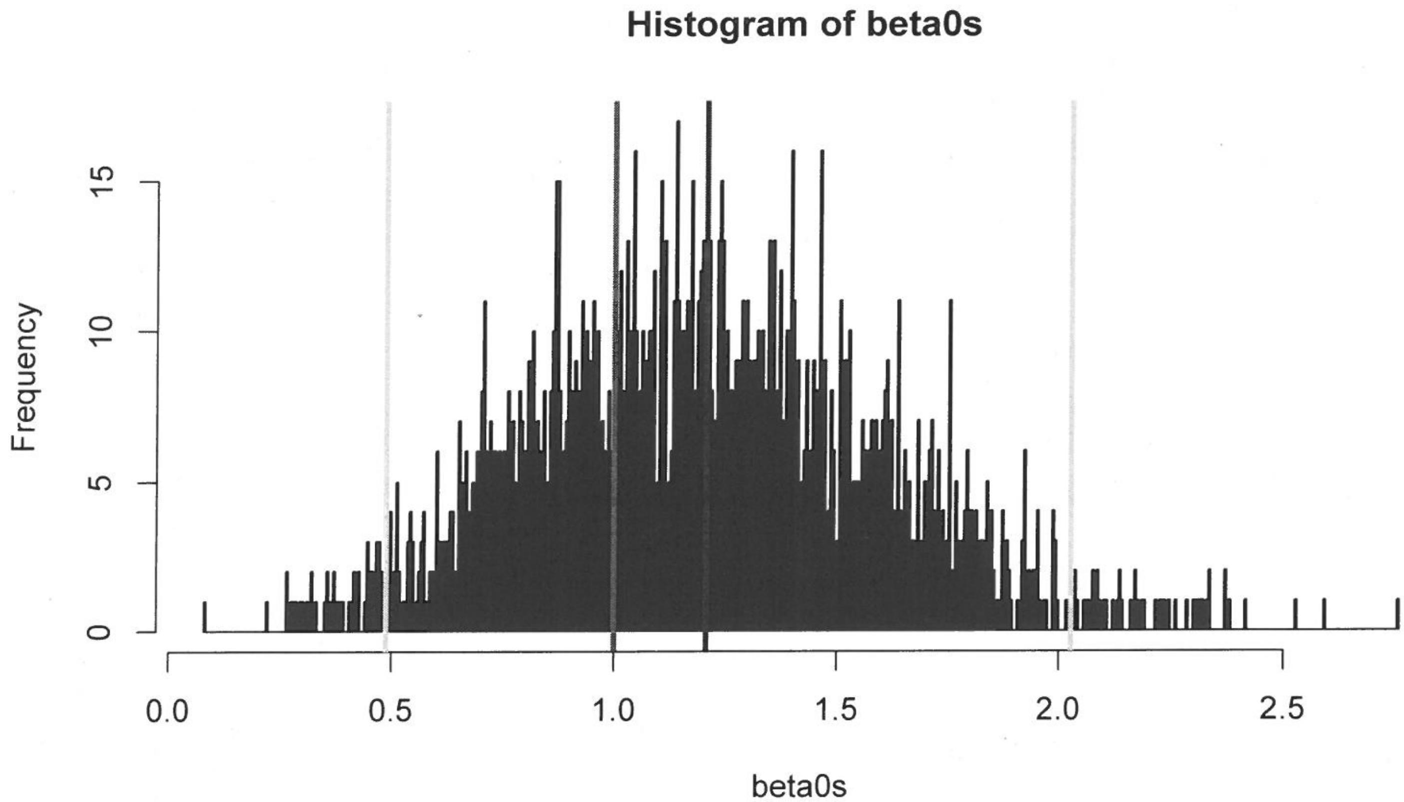
beta0s
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
1:T

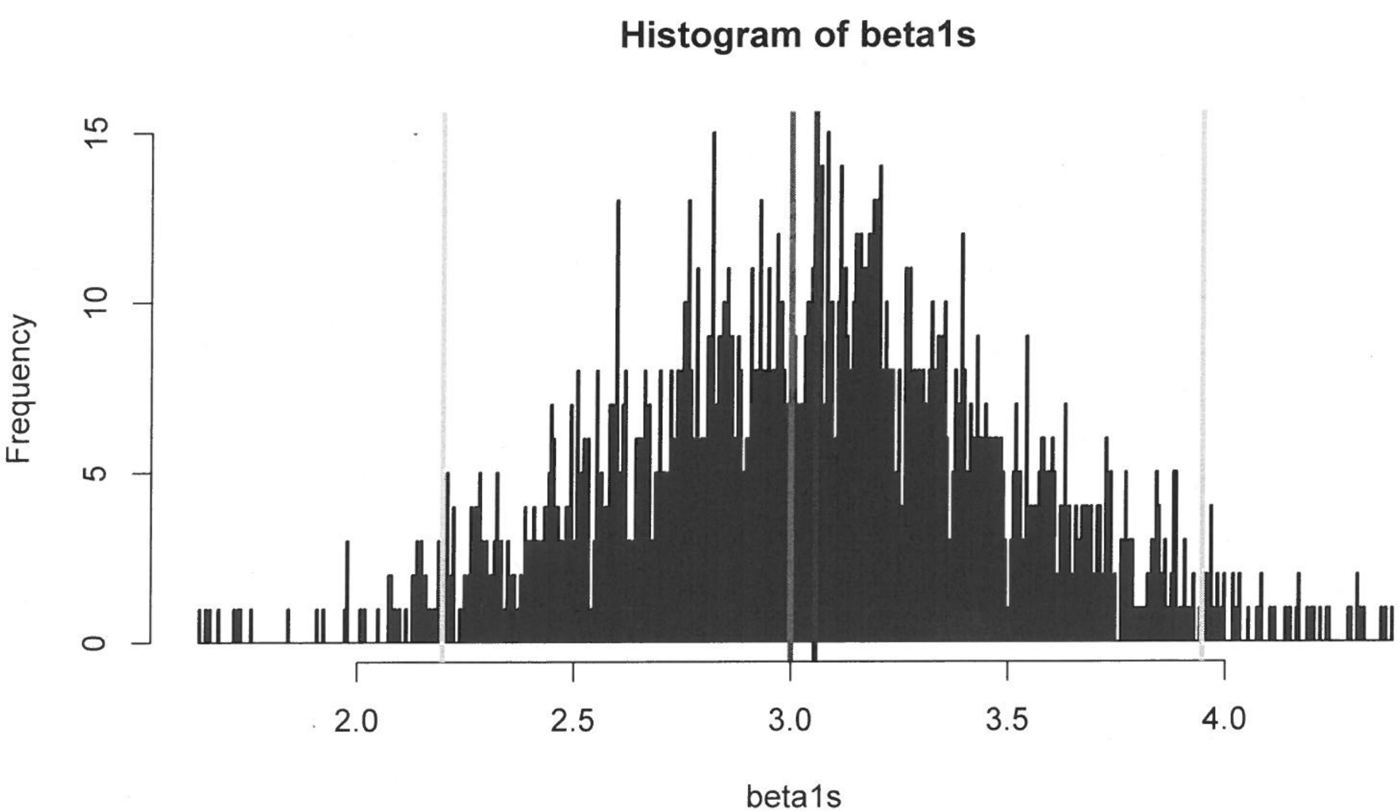


beta1s
1.5
2.0
2.5
3.0
3.5
4.0
4.5
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
1:T


Series beta0s[B:T]
1.0
0.8
0.6
0.4
0.2
0.0
ACF
0
20
40
60
80
100
Lag


Series beta1s[B:T]
1.0
0.8
0.6
0.4
0.2
0.0
ACF
0
20
40
60
80
100
Lag


Histogram of beta0s
Frequency
0
5
10
15
0.0
0.5
1.0
1.5
2.0
2.5
beta0s


Histogram of beta1s
Frequency
0
5
10
15
2.0
2.5
3.0
3.5
4.0
beta1s