Modeling Univariate Distributions

Professor: Hammou El Barmi
Columbia University

Introduction

- Usually the marginal distributions of financial time series are not well fit by normal distributions
- In this chapter we will look at other suitable distributions such as t-distribution and its skewed versions
- The parameters of these distributions are usually estimated using the maximum likelihood approach

- In parametric models, the distribution of the data is completely specified except for a finite number of parameters
- A model should have only as many parameters as needed to capture the important features of the data
- Each unknown parameter is another quantity to estimate and another source of estimation error

- Parameters are ofter classified as location, scale or shape
- A location parameter is a parameter that shift the distribution to the right or to the left without changing its shape. for example, if $f(y)$ is a fixed density, then $f(y-\mu)$ is a family of distributions with location parameter μ. As we change the values of μ, we shift the density but we do not change its shape.
- A parameter is a scale parameter if when the data is multiplied by $|a|$, it is also multiplied by $|a|$. The effect of the scale parameter is to stretch or squeeze the pdf.
- if f is a fixed density, then $\theta^{-1}f(y/\theta), \theta > 0$, is a scale family
- if f is a fixed density, the $\{f((y-\mu)), \quad -\infty < \mu < \infty, \quad \theta > 0\}$ is a family of distributions with location μ and scale θ.
- A shape parameters is defined as any parameter that is not changed by location and scale changes
- If Y has pdf f and $Z = \mu + \theta X$, then the pdf of Z is given by $\theta^{-1}f(\theta^{-1}(y-\mu))$
- $Y \sim N(\mu, \sigma^2)$, Then $Y = \mu + \sigma Z$ where $Z \sim N(0,1)$. μ is therefore a location parameter and σ is a scale parameter.
- The degrees of freedom parameter in a t-distribution is a shape parameter

- Skewness and kurtosis help characterize the shape of a probability distribution.
- Skewness measures the degree os asymmetry, with symmetry implying zero skewness, positive skewness indicating a relatively long right tail compared to the left tail and negative skewness indicating the opposite.
- Kurtosis indicates the extent to which the distribution is concentrated in the center and especially the tail of the distribution.
- The skewness of a random variable Y is defined as

$$Sk = E\left\{\frac{Y-\mu}{\sigma}\right\}^3 = \frac{E(Y-\mu)^3}{\sigma^3}$$

- For example it can be shown that for a Binomial distribution with parameters n and p

$$Sk = \frac{1-2p}{\sqrt{np(1-p)}}$$

  so

  - The skewness is positive if $p < 0.5$
  - The skewness is 0 if $p = 0.5$
  - The skewness is negative if $p > 0.5$

- The Kurtosis of a random variable Y is defined as

$$Kur = E\left\{\frac{Y-\mu}{\sigma}\right\}^4 = \frac{E(Y-\mu)^4}{\sigma^4}$$

- For example the kurtosis of a normal distribution is 3 (can show this using the moment generating function)
- The smallest the kurtosis can be is 1.
- it can be shown that for a Binomial distribution with parameters n and p

$$Kur = 3 + \frac{1-6p(1-p)}{np(1-p)}$$

- The kurtosis of a t-distribution with ν degrees of freedom is

$$Kur = 3 + \frac{6}{\nu-4}$$

So for $t_5, Kur = 9$

- Estimation on skewness and kurtosis of a distribution is relatively easy.
- Suppose $Y_1, Y_1, \ldots, Y_n$ is a random sample from this distribution.
- The sample skewness is denoted by $\hat{Sk}$ and is given by

$$\hat{Sk} = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \bar{Y})^3}{s^3}$$

- The sample kurtosis is denoted by $\hat{Kur}$ and is given by

$$\hat{Kur} = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \bar{Y})^4}{s^4}$$

- A test for normality can be based on the skewness and kurtosis coefficient.
- This test is called the Jarque-Bera test and is given by

$$JB = n\{\hat{SK}^2/6 + (\hat{Kur} - 3)^2/24\}$$

- This test compares the sample skewness and kurtosis to 0 and 3, their values under normality.
- In R, the test statistic and its p-value can be computed using **jarque.bera.test()**
- **A large sample approximation is used to compute the p-value.**
- **Under the null hypothesis (data from a normal distribution), JS converges in distribution to a chi-square with 2 degrees of freedom, so**

$$\textbf{p-value} = P(\chi_2^2 > JB)$$

- Distributions with higher tail probabilities compared to a normal distribution are called heavy-tailed distributions
- Heavy tailed distributions are very important in finance because equity returns usually have heavy tails
- We are concerned when the return distribution has heavy tails because of the possibility of an extremely large negative return which could entirely deplete the capital reserves of a company
- If one sells short then large positive returns are also worrisome

- The tails of the normal distribution drop at an exponential rate when $|y| \to \infty$if $\mu = 0$ and $\sigma^2 = 1$, then $f(y) \sim e^{-y^2/2}$
- If $f(y) \sim A|y|^{-(1+\alpha)}$ as $y \to \infty$ then is heavy tailed
- For example the t-distribution with ν degrees of freedom has density

$$f_{t,\nu}(y) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)} \frac{1}{\{1 + y^2/\nu\}^{(1+\nu)/2}} \sim \frac{1}{|y|^{\nu+1}} \quad \text{as} \quad y \to \infty$$

- recall that if $Z \sim N(0,1)$ and $W \sim \chi^2_\nu$ and Z and W are independent, then

$$t = \frac{Z}{\sqrt{W/\nu}}$$

- The variance of the t-distribution with ν degrees of freedom, t_ν, is finite and equal to $\nu/(\nu-2)$ is $\nu > 2$. If $0 < \nu \leq 1$, then the mean of t_ν does not exist.If $1 < \nu \leq 2$, then the mean is finite but the variance is infinite.
- If $Y \sim t_\nu$, then

$$V = \mu + \lambda Y$$

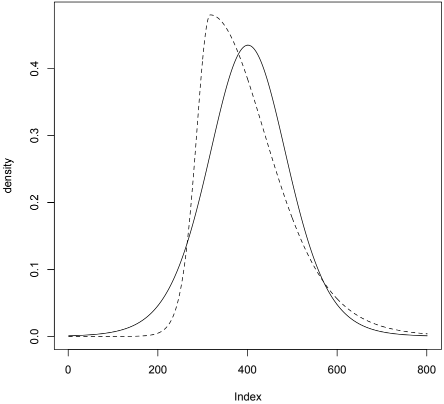
  is said to have a $t_\nu(\mu, \lambda^2)$ (μ is the location parameter and λ is the scale parameter)
- If $\nu > 1$, then $E(V) = \mu$ and if $\nu > 2$ then $\sigma^2(Y) = \lambda^2\nu/(\nu-2)$.
- $t_\nu = t_\nu(0,1)$
- The standardized t-distribution is defined as $t_\nu(0,(\nu-2)/\nu)$. Note that for $\nu > 2$, the mean of the standardized t-distribution is 0 and its variance is 1.
- Notice that $t_\nu(0,(\nu-2)/\nu)$ distribution with $\nu > 2$ has mean 0 and variance equal to 1. It is called the standardized t-distribution and is denoted by t_ν^{std}.
- In general $t_\nu^{std}(\mu,\sigma^2) = t_\nu(\mu, \{(\nu-2)/\nu\}\sigma^2)$

- Financial market data typically have no natural lower bound of upper bounds hence why we need to use models with support $(-\infty, \infty)$
- If the data are symmetric on can use the normal, t or generalized distributions as models. What if the data is skewed.
- Many of the well known distribution such as gamma and log-normal disturbutions has support $[0, \infty)$ and are therefore not suitable
- One can use instead the skewed normal or t distributions.
- (Fernandez and Steel, 1998) Let ξ be a positive constant and f a symmetric density about 0 and define

$$f(y|\xi) \begin{cases} cf(y\xi), & y < 0 \\ cf(y/\xi) & y \geq 0 \end{cases}$$

  where $c = 2\xi/(1+\xi^2)$
- if $\xi > 1$, the distribution is skewed to the right and if $xi < 1$, the distribution is skewed to the left
- If f is a t-distribution, then $f(y|\xi)$ is a skewed t-distribution

Figure: Symmetric (solid) and skewed (dotted) t-densities bothe with mean 0 and standard deviation 1, $\nu = 10$ and $\xi = 2$ in the skewed density



Generalized Error Distributions

- The generalized standardized error distribution or GED with shape parameter ν has density defined as

$$f_{ged}^{std}(y|\nu) = \kappa(\nu) \exp\left\{-\frac{1}{2} \mid \frac{y}{\lambda_\nu} \mid^{\nu}\right\}$$

where

$$\lambda_\nu = \left\{\frac{2^{-2/\nu}\Gamma(\nu^{-1})}{\Gamma(3/\nu)}\right\}^{1/2} \quad \text{and} \quad \kappa(\nu) = \frac{\nu}{\lambda_\nu 2^{1+1/\nu}\Gamma(\nu^{-1})}$$

- The shape parameter ν determines the tail weight with smaller values giving greater tail weight
- When $\nu = 2$, a GED is a normal distribution and when $\nu = 1$, it is a double exponential distribution

AIC and BIC

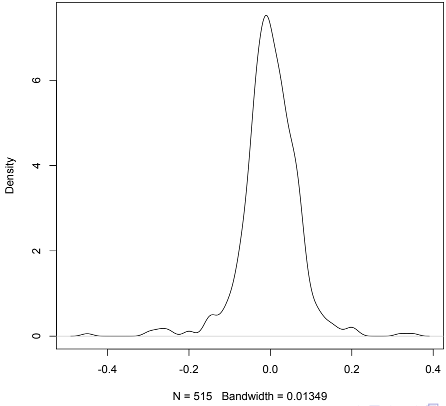
- An important practical problem is choosing between two or more statistical models that might be appropriate for a data set
- The maximized value of the log-likelihood, denoted by $\log(L(\hat{\theta}_{ML}))$ can be used to measure how well a model fits the data or to compare the fits of two model or more models
- The problem is that $\log(L(\hat{\theta}_{ML}))$ can be increased by just adding parameters to the model. The additional parameters do not necessarily mean that the model is better description of the the data generating-mechanism. They may just add complexity.
- Models should be compared both by fit to data and model complexity
- AIC (Akaike's information criterion) and BIC (Bayesian information criterion) are two means for achieving a trade off between fit and complexity. they are defined as

$$\begin{aligned} AIC &= -2\log(L(\hat{\theta}_{ML})) + 2p \\ BIC &= -2\log(L(\hat{\theta}_{ML})) + \log(n)p \end{aligned}$$

  where p is the number of parameters in the model and n is the sample size
- For both criteria, **smaller is better**

- In this example we fit different distributions to the changes in the risk free returns $y_t = r_t - r_{t-1}$ (The data is in Capm in Ecdat).
  `> data(Capm, package = "Ecdat") x = diff(Capm$rf)`
- density estimate of data

Figure: Density estimate for changes in risk free returns



Example

- In R, we fit the standardized t using stdFit (y)
- In R, we fit the skewed t distribution using sstdFit(y)
- In R we fit the generalized error distribution using gedFit (y)
- In R we fit the skewed generalized error distribution using sgedFit(y)

- If we fit the standardized t distribution we get, we can use
  $>$ fitdistr(x,"t")
- the results is
  - $-\log(L) = -693$ (hence AIC$= -2(693) + 2(3) = -1380$)
  - m $= 0.00121$ (mean estimate)
  - sd$= 0.07247$ (scale parameter)
  - df $= 3.33397$ (degrees of freedom estimate)
- The parameters, in order, are the mean, the scale parameter, and the degrees of freedom. The numbers in parentheses are the standard errors.
- Next, we fit the t-distribution by writing a function to return the negative log-ikelihood and using R's optim() function to minimize the log-likelihood.
- We compute standard errors by using solve() to invert the Hessian and then taking the square roots of the diagonal elements of the inverted Hessian.
- We also compute AIC and BIC.

Example

```
library(fGarch)
 n = length(x)
start = c(mean(x), sd(x), 5)
loglik_t = function(beta) sum( - dt((x - beta[1]) / beta[2],  beta[3],
log = TRUE) + log(beta[2]) )
fit_t = optim(start, loglik_t, hessian = T, method = "L-BFGS-B",
lower = c(-1, 0.001, 1))
AIC_t = 2 * fit_t$value + 2 * 3
BIC_t = 2 * fit_t$value + log(n) * 3
sd_t = sqrt(diag(solve(fit_t$hessian)))
fit_t$par
sd_t
AIC_t
BIC_t
```

The results are below. The estimates and the standard errors agree with those produced by fitdistr(), except for small numerical errors.

```
> fit_t$par
[1] 0.00122 0.04586 3.33655 > sd_t
[1] 0.00245 0.00246 0.49982 > AIC_t
[1] -1380.4
> BIC_t
[1] -1367.6
```

The standardized t-distribution can be ?t by changing dt() to dstd().Then the parameters are the mean, standard deviation, and degrees of freedom.

```
loglik_std = function(beta) sum(- dstd(x, mean = beta[1],
sd = beta[2], nu = beta[3], log = TRUE)) fit_std = optim(start,
 loglik_std, hessian = T, method = "L-BFGS-B", lower = c(-0.1, 0.01, 2.1))
  AIC_std = 2*fit_std$value +2*3
BIC_std = 2*fit_std$value + log(n) * 3
sd_std = sqrt(diag(solve(fit_std$hessian)))
fit_std$par
sd_std
AIC_std
BIC_std
```

AIC and BIC are unchanged, as expected since we are ?tting the same model as before and only changing the parameterization.

```
>fit_std$par
[1] 0.0012144 0.0725088 3.3316132
> sd_std
[1] 0.0024538 0.0065504 0.4986456
> AIC_std
[1] -1380.4
> BIC_std
[1] -1367.6
```

Fitting an F-S skewed t-distribution to changes in risk-free re-turns

```
loglik_sstd = function(beta) sum(- dsstd(x, mean = beta[1],
sd = beta[2], nu = beta[3], xi = beta[4], log = TRUE)) start = c(mean(x),
sd(x), 5, 1) fit_sstd = optim(start, loglik_sstd, hessian = T,
method = "L-BFGS-B", lower = c(-0.1, 0.01, 2.1, -2))
AIC_sstd = 2*fit_sstd$value + 2 * 4
BIC_sstd = 2*fit_sstd$value + log(n) * 4
sd_sstd = sqrt(diag(solve(fit_sstd$hessian)))
fit_sstd$par
sd_sstd
AIC_sstd
BIC_sstd
```

Example

```
> fit_sstd$par
[1] 0.0011811 0.0724833 3.3342759 0.9988491
> sd_sstd
[1] 0.0029956 0.0065790 0.5057846 0.0643003
> AIC_sstd
[1] -1378.4
> BIC_sstd
[1] -1361.4
```

Example

- For this data set we get AIC and BIC for several models and the results are

Distribution	# of parameters	AIC+1300	BIC + 1300
t	3	-80.4	-67.7
skewed t	4	-78.4	-61.4
ged	3	-75.6	-50.9
skewed ged	4	-61.6	-44.6

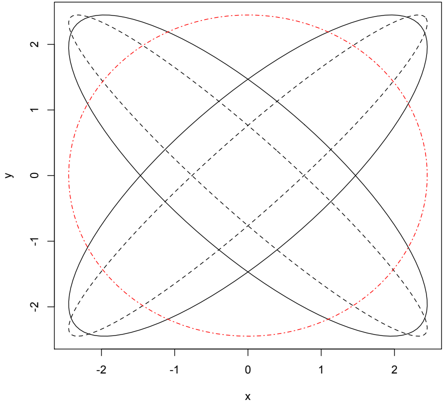
- The t-model is the best by either criteria.

- Suppose $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_d)^T$ has a multivariate normal distribution with density is given by

$$\Phi_d(\mathbf{y}, \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y}-\boldsymbol{\mu})}$$

  where $\mathbf{y} = (y_1, y_2, \ldots, y_d)^T, \Sigma = Var(\mathbf{Y})$ and $|\Sigma|$ is the determinant of Σ.
- If $\sigma_{ij} = 0$ for all $i \neq j$, then $Y_1, Y_2, \ldots, Y_d$ are independent
- If $\mathbf{c} = (c_1, c_2, \ldots, c_d)^T$ then $\mathbf{c}^T\mathbf{Y} \sim N(\mathbf{c}^T\boldsymbol{\mu}, \mathbf{c}^T\Sigma\mathbf{c})$
- if $\mathbf{c} = (1, 0, \ldots, 0)^T$ then $Y_1 = \mathbf{c}^T\mathbf{Y} \sim N(\mathbf{c}^T\boldsymbol{\mu} = \mu_1, \mathbf{c}^T\Sigma\mathbf{c} = \sigma_1^2)$
- If $\mathbf{Y}$ is the vector of the returns of d assets and $\mathbf{c}^T\mathbf{Y}$ is the return of a portfolio, then the return of the portfolio has $N(\mathbf{c}^T\boldsymbol{\mu}, \mathbf{c}^T\Sigma\mathbf{c})$

Figure: $\rho = -9.5$ and 0.95 (doted ellipsoids), $\rho = -0.8$ and 0.8 (solid ellipsoids) and $\rho = 0$ (red ellipsoid)



- The assumption of multivariate normality facilitates many useful probability calculations.
- If the returns on a set of assets have a multivariate normal distribution, then the return on any portfolio formed from these assets will be normally distributed.
- This is because the return on the portfolio is the weighted average of the returns on the assets.
- Therefore, the normal distribution could be used, for example, to ?nd the probability of a loss of some size of interest, say, 10 % or more, on the portfolio.
- Such calculations have important applications in finding a value-at-risk; see Chap. 19.
- Unfortunately, often individual returns are not normally distributed, which implies that a vector of returns will not have a multivariate normal distribution.
- Next we will look at an important class of heavy-tailed multivariate distributions.

Multivariate t-Distribution

- The univariate t-distribution is a good model for the returns of individual assets.
- Therefore, it is desirable to have a model for vectors of returns such that the univariate marginals are t-distributed.
- The multivariate t-distribution has this property.

Multivariate t-Distribution

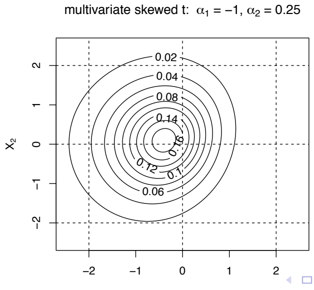
- Let $\mathbf{Z} \sim N(\mathbf{0}, \Lambda)$ and $W \sim \chi^2_\nu$. Assume that Z and W are independent.
- Define

$$\mathbf{Y} = \boldsymbol{\mu} + \sqrt{\frac{\nu}{W}}\mathbf{Z}$$

- $\mathbf{Y}$ has a multivariate t-distrbution ($\mathbf{Y} \sim t_\nu(\boldsymbol{\mu}, \Lambda)$)
- $E(\mathbf{Y}) = \boldsymbol{\mu}$ if $\nu > 1$.
- for $\nu > 2$, $Var(\mathbf{Y}) = \frac{\nu}{\nu-2}\Lambda$
- If $\mathbf{Y} \sim t_\nu(\boldsymbol{\mu}, \Lambda)$ and $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_d)^T$, then $\boldsymbol{\omega}^T\mathbf{Y} \sim t_\nu(\boldsymbol{\omega}^T\boldsymbol{\mu}, \boldsymbol{\omega}^T\Lambda\boldsymbol{\omega})$
- if $\boldsymbol{w} = (1, 0, \ldots, 0)^T$ then $Y_1 = \boldsymbol{w}^T\mathbf{Y} \sim t_\nu(\mu_1, \Lambda_{11})$
- If $\mathbf{Y}$ is the vector of the returns of d assets and $\boldsymbol{w}^T\mathbf{Y}$ is the return of a portfolio, then the return of the portfolio has $t_\nu(\boldsymbol{\omega}^T\boldsymbol{\mu}, \boldsymbol{\omega}^T\Lambda\boldsymbol{\omega})$
- $\Lambda_{ij} = 0$ for all $i \neq j$ does not imply that $Y_1, Y_2, \ldots, Y_d$ are independent

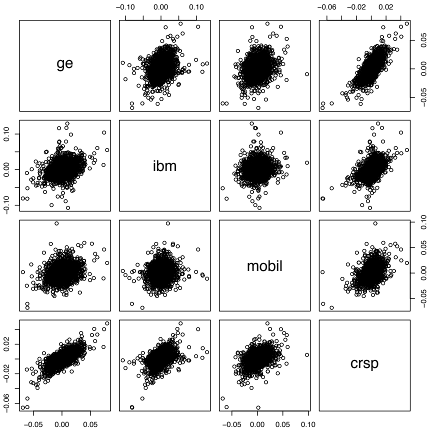
- The skewed t-distribution can also be defined. In addition to the shape parameter ν determining the tail weight, the skewed t-distribution has a vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)^T$ of shape parameters determining the amount of skewness.
- If $\mathbf{Y}$ has skewed t-distrbution, then Y_i is left-skewed, symmetric or right skewed depending on whether $\alpha_i < 0, = 0$ or > 0.
- Example

Figure: Bivariate Skewed t distibutions



Example

- The data we use here is CRSPday data in Rs Ecdat package.
- There are 5 variables representing the returns from 1/3/1969 to 12/31/1998 on three stocks, GE, IBM and Mobil and on CRSP value weighted index (CRSP is the center for research in Security Prices at the University of Chicago).
- the pairwise graph of the dat is



Example

The covariance

	ge	ibm	mobil	crsp
ge	1.88e-04	8.01e-05	5.27e-05	7.61e-05
ibm	8.01e-05	3.06e-04	3.59e-05	6.60e-05
mobil	5.27e-05	3.59e-05	1.67e-04	4.31e-05
crsp	7.61e-05	6.60e-05	4.31e-05	6.02e-05

and the correlation is

	ge	ibm	mobil	crsp
ge	1.000	0.334	0.297	0.715
ibm	0.334	1.000	0.159	0.486
mobil	0.297	0.159	1.000	0.429
crsp	0.715	0.486	0.429	1.000

- Use install.packages(sn) to install the package sn in order to fit a multivariate skewed t
- To fit a multivariate skewed distribution to the data,we use mst.fit(rep(1, n), filename) when n is the number of case in the data and filename is the name of the data set.
- In our case n=2528 and I save the data in a file called data. So I used mst.fit(rep(1, 2528), data).
- To get the estimate of center , Ω, degrees of freedom and the skewness parameters type mst.fit(rep(1,2528),data)$dp and you get in this case
  - $beta

	ge	ibm	mobil	crsp
0.000891	0.000375	0.000656	0.000745	

  - $Omega

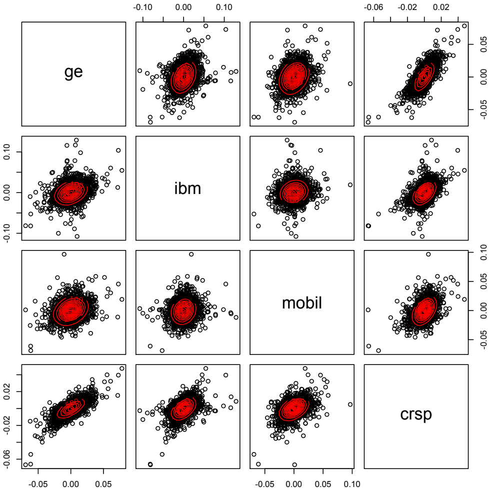
	ge	ibm	mobil	crsp
ge	1.25e-04	4.82e-05	3.43e-05	4.58e-05
ibm	4.82e-05	1.82e-04	2.30e-05	3.83e-05
mobil	3.43e-05	2.30e-05	1.15e-04	2.80e-05
crsp	4.58e-05	3.83e-05	2.80e-05	3.66e-05

  - $alpha

ge	ibm	mobil	crsp
0.003238	0.004947	0.001650	-0.000599

  - $df
    5.92

Figure: Skewed t distibution (from output)



Example

- We can also fit a multivariate normal distribution. For this we need the package mvnmle and to fit the model use mlest( ).
- the results are
  - $muhat

0.001071	0.000700	0.000779	0.000678

  - *sigmahat*

1.88e-04	8.00e-05	5.27e-05	7.60e-05
8.00e-05	3.06e-04	3.59e-05	6.60e-05
5.27e-05	3.59e-05	1.67e-04	4.30e-05
7.60e-05	6.60e-05	4.30e-05	6.02e-05

  - -log(likelihood) $= -81604$