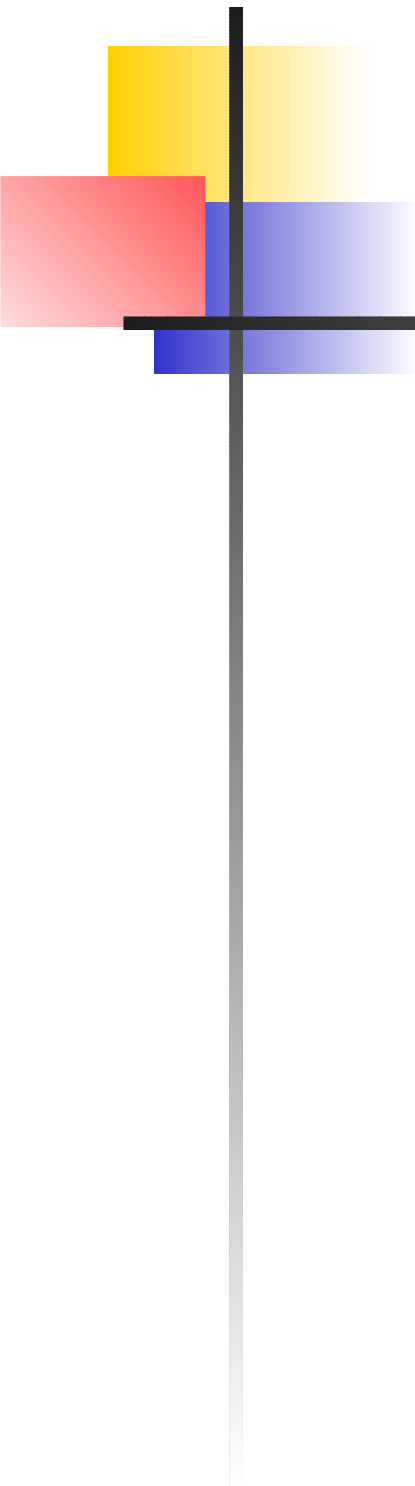




Generalized Gamma Frailty Model and Inference

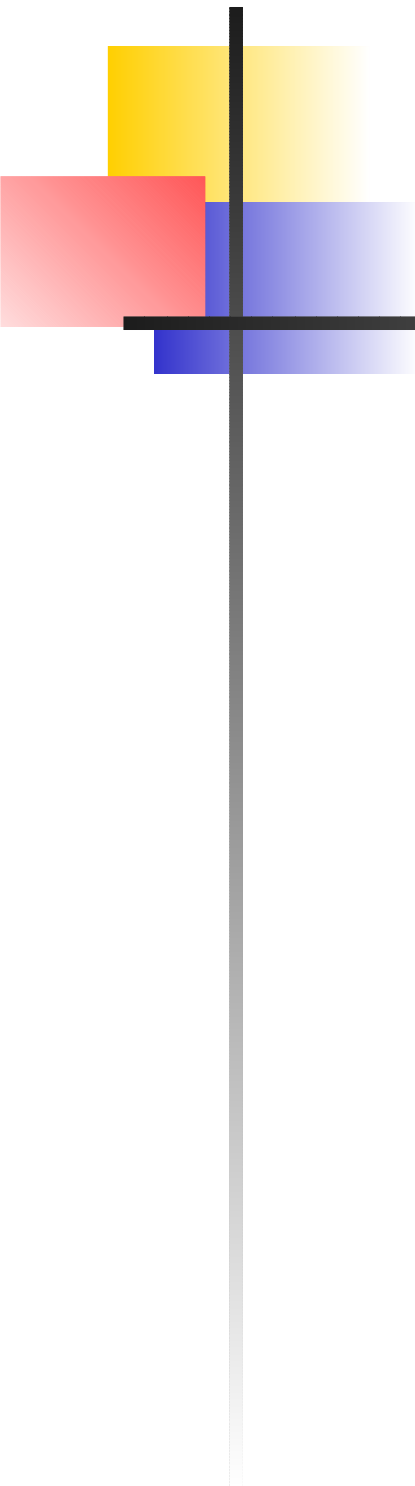
Prof. N. Balakrishnan

Dept. of Mathematics & Statistics

McMaster University

Hamilton, Canada

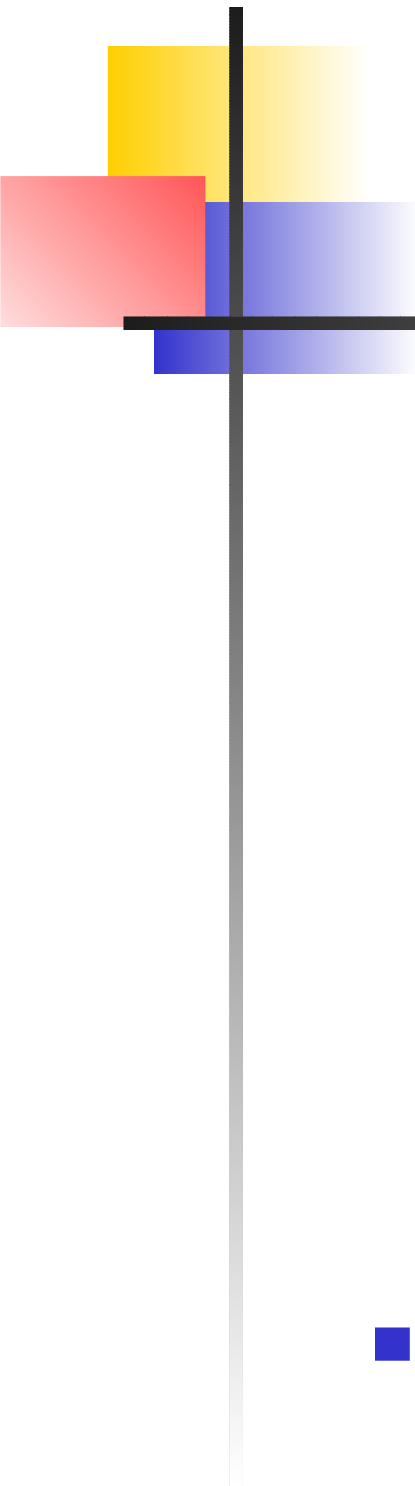
bala@mcmaster.ca



In collaboration with

Prof. Yingwei Peng
Dept. of Mathematics & Statistics
Memorial University of Newfoundland
St. John's
Newfoundland
Canada

ypeng@math.mun.ca

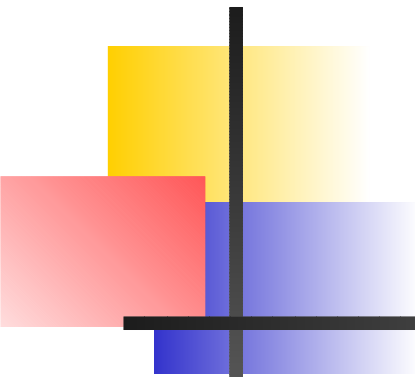


In collaboration with

Prof. Yingwei Peng
Dept. of Mathematics & Statistics
Memorial University of Newfoundland
St. John's
Newfoundland
Canada

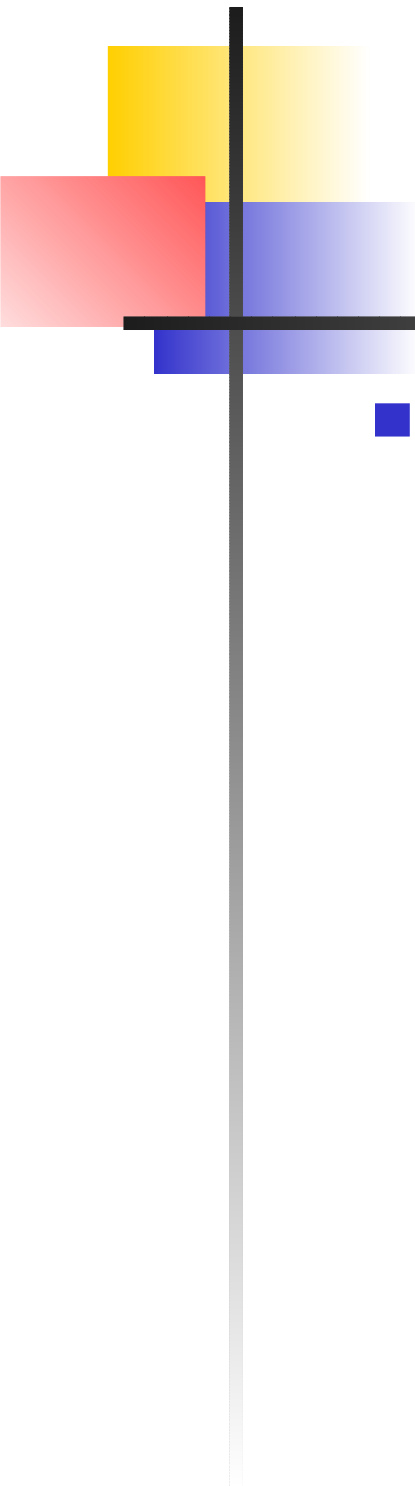
ypeng@math.mun.ca

- *Statistics in Medicine*, **25**, 2797 – 2816 (2006)



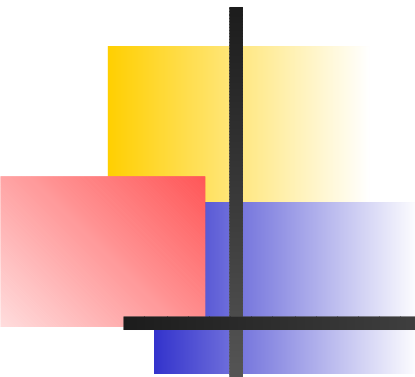
Roadmap

1. Introduction
2. Generalized Gamma Frailty Model
3. Estimation Method
 - 3.1. Monte Carlo Approximation of the Likelihood
 - 3.2. Choices of the Baseline Distribution
4. Simulation Study
5. Goodness-of-Fit
6. Illustrative Examples
7. References



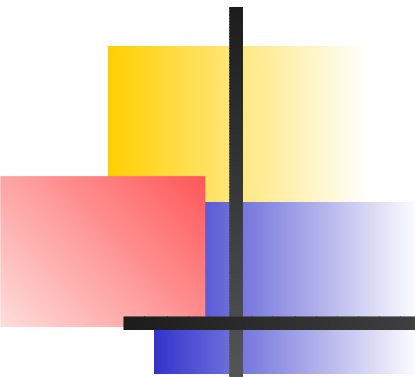
Introduction

- Multivariate or cluster failure time data are commonly encountered in survival analysis, and finding an appropriate method to model the correlation among the observations is a very important issue.



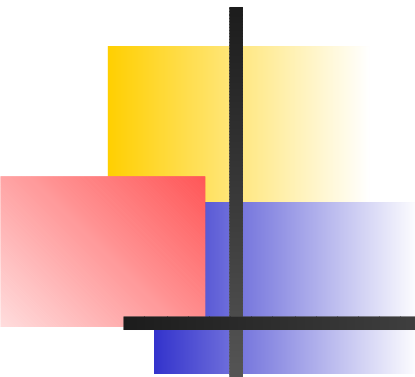
Introduction

- Multivariate or cluster failure time data are commonly encountered in survival analysis, and finding an appropriate method to model the correlation among the observations is a very important issue.
- A convenient choice for modeling the correlation in survival data is the so-called *frailty model*, which was originally proposed by Vaupel *et al.* (1979).



Introduction

- Multivariate or cluster failure time data are commonly encountered in survival analysis, and finding an appropriate method to model the correlation among the observations is a very important issue.
- A convenient choice for modeling the correlation in survival data is the so-called *frailty model*, which was originally proposed by Vaupel *et al.* (1979).
- A frailty model for clustered data is defined as follows.

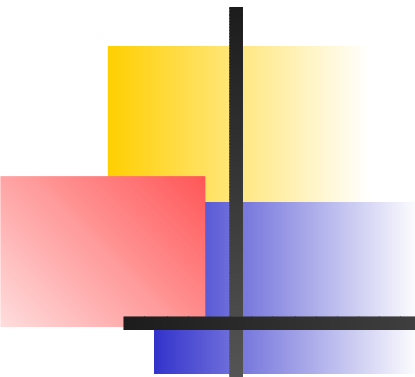


Introduction (cont.)

■ Let

$$(t_{ij}, \delta_{ij}, x_{ij}), i = 1, \dots, n, j = 1, \dots, m_i,$$

be the failure time, censoring indicator, and the covariate of the j -th individual in the i -th cluster, where $\delta_{ij} = 1$ if t_{ij} is uncensored and 0 otherwise.



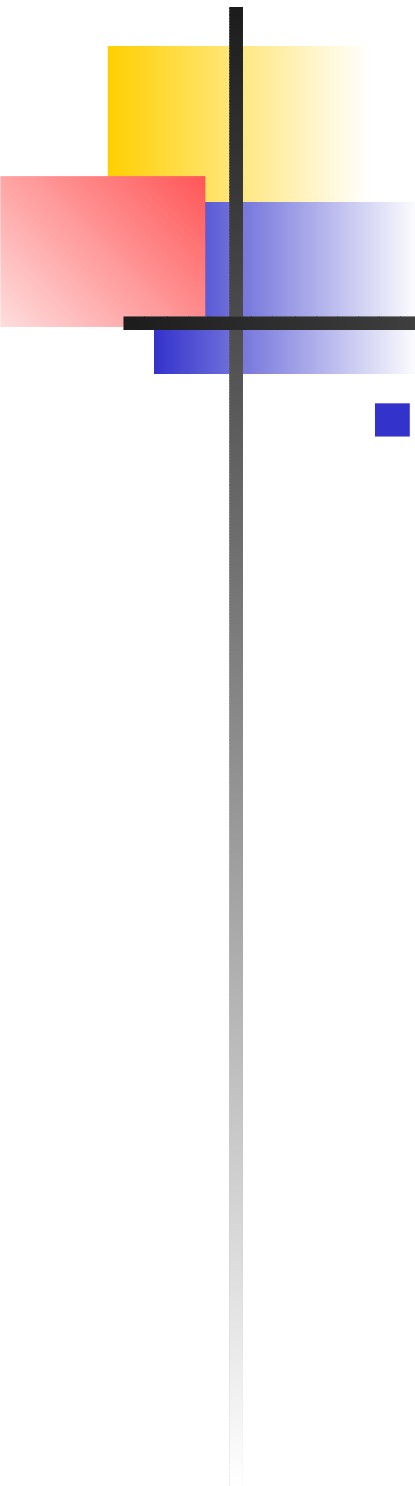
Introduction (cont.)

- Let

$$(t_{ij}, \delta_{ij}, x_{ij}), i = 1, \dots, n, j = 1, \dots, m_i,$$

be the failure time, censoring indicator, and the covariate of the j -th individual in the i -th cluster, where $\delta_{ij} = 1$ if t_{ij} is uncensored and 0 otherwise.

- Let y_i denote the unobserved frailty shared by the individuals in the i -th cluster.

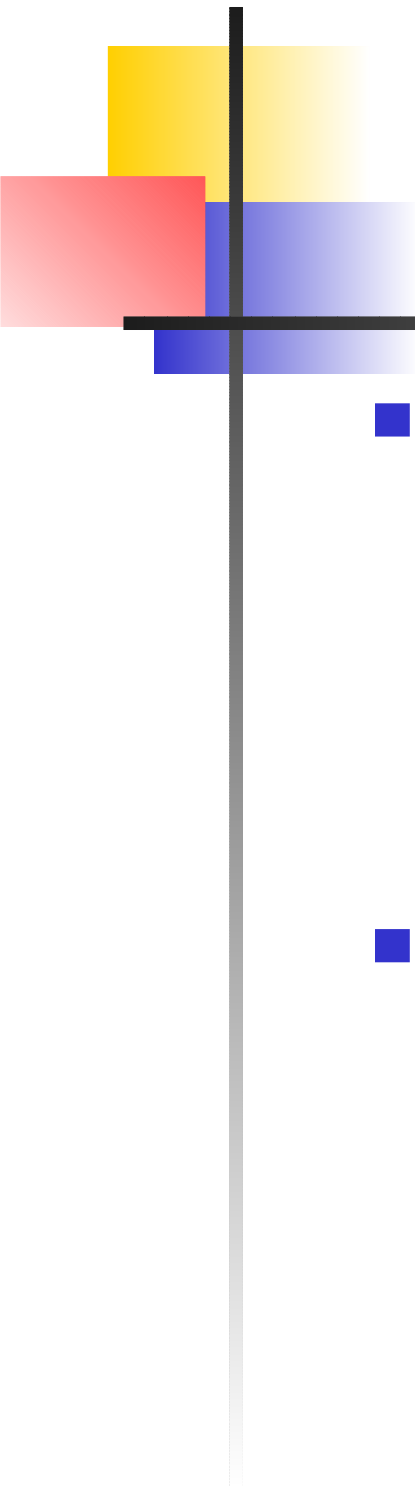


Introduction (cont.)

- Given y_i , the frailty model specifies that t_{ij} are independent with a proportional hazard function

$$h(t_{ij}|y_i) = y_i h_0(t_{ij}) \exp(\beta' x_{ij}),$$

where $h_0(t)$ is an arbitrary baseline hazard function.



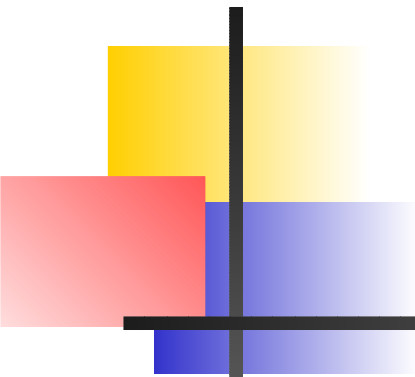
Introduction (cont.)

- Given y_i , the frailty model specifies that t_{ij} are independent with a proportional hazard function

$$h(t_{ij}|y_i) = y_i h_0(t_{ij}) \exp(\beta' x_{ij}),$$

where $h_0(t)$ is an arbitrary baseline hazard function.

- The frailties y_i are usually assumed to be iid with a distribution referred to as the *frailty distribution*.



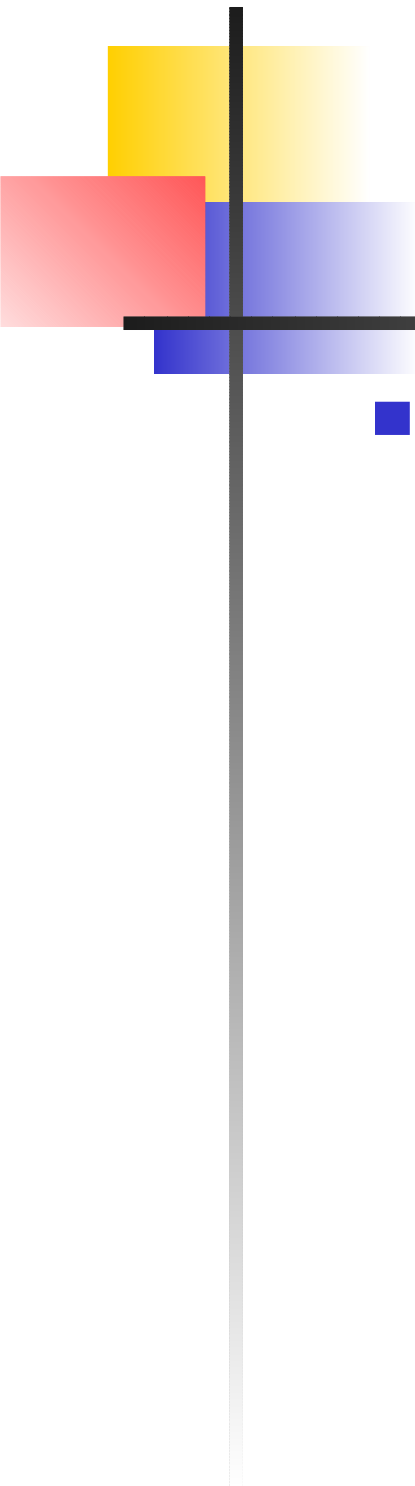
Introduction (cont.)

- Given y_i , the frailty model specifies that t_{ij} are independent with a proportional hazard function

$$h(t_{ij}|y_i) = y_i h_0(t_{ij}) \exp(\beta' x_{ij}),$$

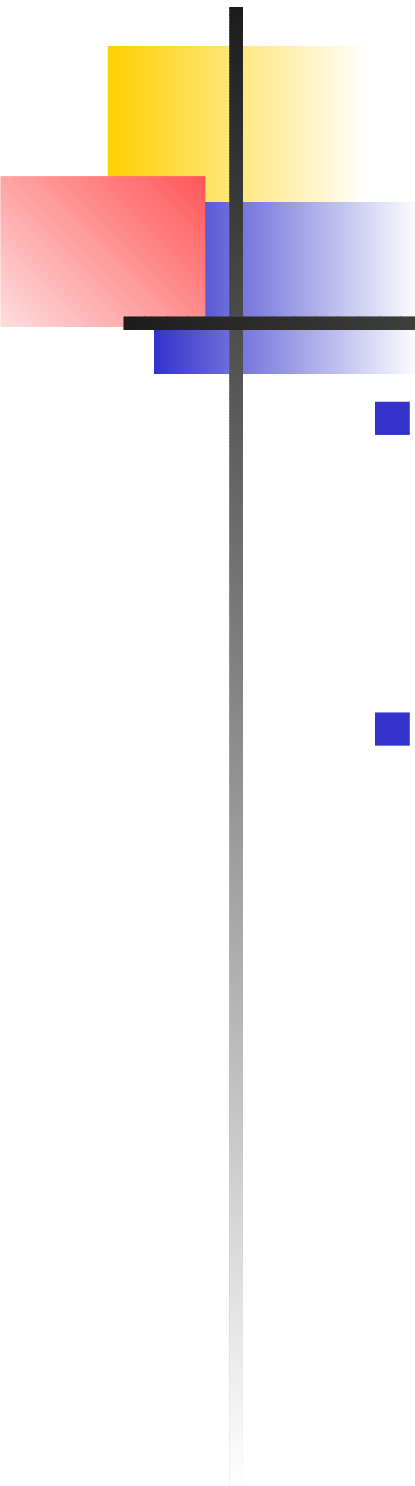
where $h_0(t)$ is an arbitrary baseline hazard function.

- The frailties y_i are usually assumed to be iid with a distribution referred to as the *frailty distribution*.
- A popular choice for the frailty distribution is the gamma distribution since it facilitates EM algorithm [Klein (1992) and Nielsen *et al.* (1992)].



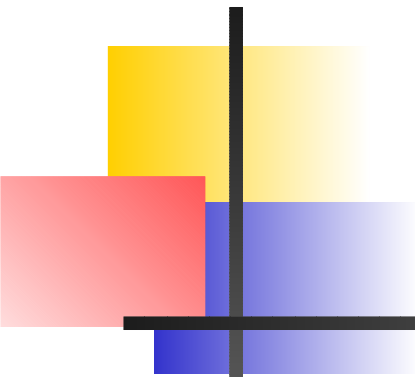
Introduction (cont.)

- Another possibility is the positive stable distribution proposed by Hougaard (1986a,b), which preserves the proportionality of hazards in the marginal distribution.



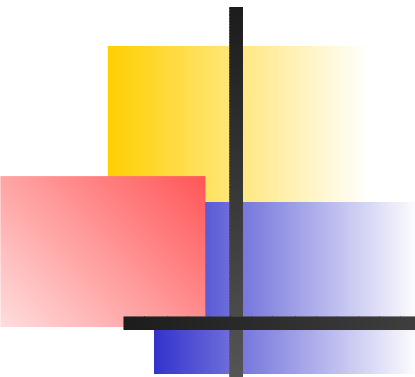
Introduction (cont.)

- Another possibility is the positive stable distribution proposed by Hougaard (1986a,b), which preserves the proportionality of hazards in the marginal distribution.
- Both gamma and positive stable distributions have simple Laplace transforms and hence are convenient to use.



Introduction (cont.)

- Another possibility is the positive stable distribution proposed by Hougaard (1986a,b), which preserves the proportionality of hazards in the marginal distribution.
- Both gamma and positive stable distributions have simple Laplace transforms and hence are convenient to use.
- The convenience, however, does not assure that the fit is good. There are no biological reasons that motivate the two distributions [Hougaard (1995)].

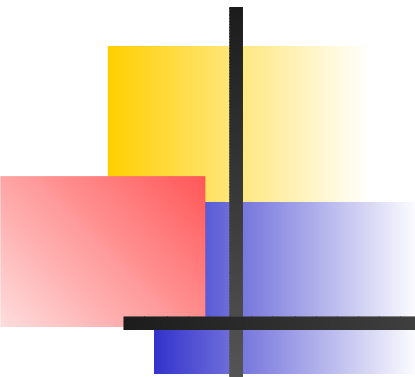


Generalized Gamma Frailty Model

- A random variable $Y > 0$ follows the generalized gamma distribution if the density function is

$$g(y) = \begin{cases} |q|\sigma^{-1}(q^{-2})^{q^{-2}} \lambda(\lambda y)^{q^{-2}(\frac{q}{\sigma})-1} \exp[-q^{-2}(\lambda y)^{\frac{q}{\sigma}}] / \Gamma(q^{-2}), & q \neq 0, \\ (\sqrt{2\pi}\sigma y)^{-1} \exp\{-[\log(\lambda y)]^2 / (2\sigma^2)\}, & q = 0, \end{cases}$$

where $-\infty < q < \infty$ and $\sigma > 0$ are shape parameters and $\lambda > 0$ is a scale parameter.



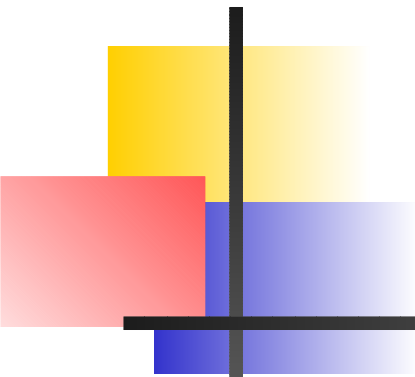
Generalized Gamma Frailty Model

- A random variable $Y > 0$ follows the generalized gamma distribution if the density function is

$$g(y) = \begin{cases} |q|\sigma^{-1}(q^{-2})^{q^{-2}} \lambda(\lambda y)^{q^{-2}(\frac{q}{\sigma})-1} \exp[-q^{-2}(\lambda y)^{\frac{q}{\sigma}}] / \Gamma(q^{-2}), & q \neq 0, \\ (\sqrt{2\pi}\sigma y)^{-1} \exp\{-[\log(\lambda y)]^2 / (2\sigma^2)\}, & q = 0, \end{cases}$$

where $-\infty < q < \infty$ and $\sigma > 0$ are shape parameters and $\lambda > 0$ is a scale parameter.

- It includes many distributions as special cases:



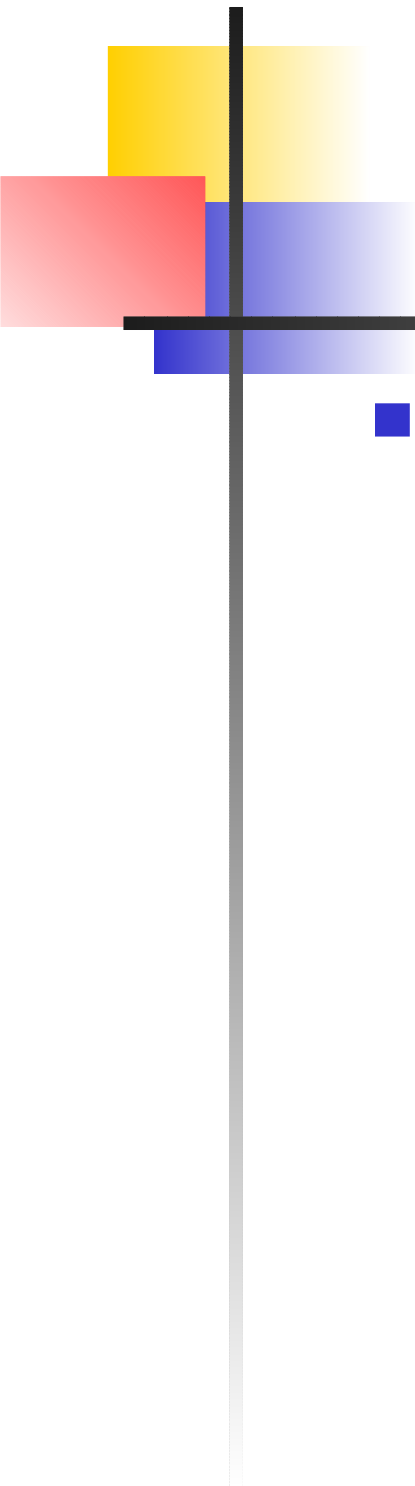
Generalized Gamma Frailty Model

- A random variable $Y > 0$ follows the generalized gamma distribution if the density function is

$$g(y) = \begin{cases} |q|\sigma^{-1}(q^{-2})^{q^{-2}} \lambda(\lambda y)^{q^{-2}(\frac{q}{\sigma})-1} \exp[-q^{-2}(\lambda y)^{\frac{q}{\sigma}}] / \Gamma(q^{-2}), & q \neq 0, \\ (\sqrt{2\pi}\sigma y)^{-1} \exp\{-[\log(\lambda y)]^2 / (2\sigma^2)\}, & q = 0, \end{cases}$$

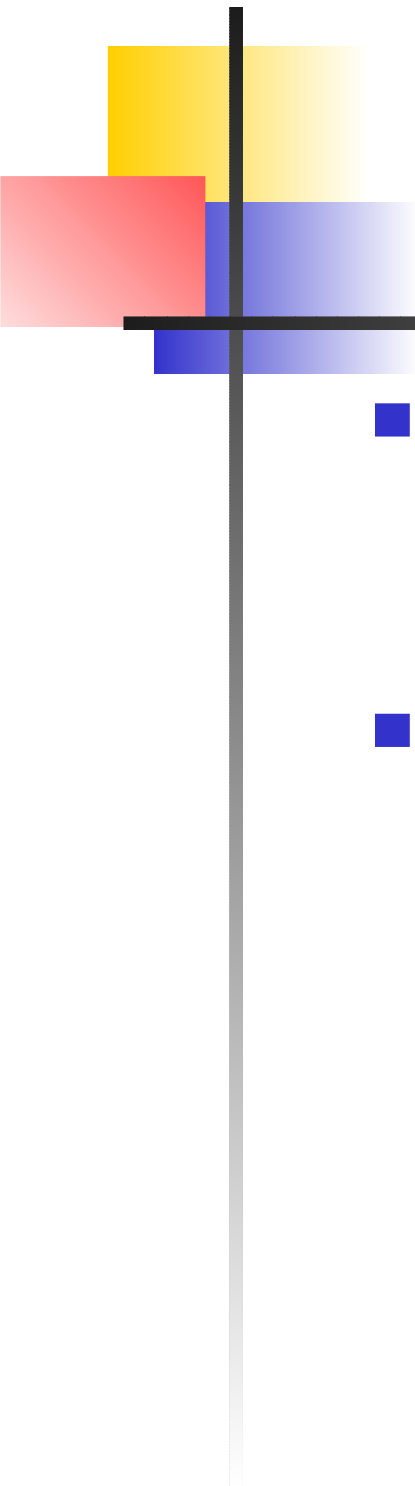
where $-\infty < q < \infty$ and $\sigma > 0$ are shape parameters and $\lambda > 0$ is a scale parameter.

- It includes many distributions as special cases:
 - Weibull distribution when $q = 1$;
 - lognormal distribution when $q = 0$;
 - gamma distribution when $q/\sigma = 1$.



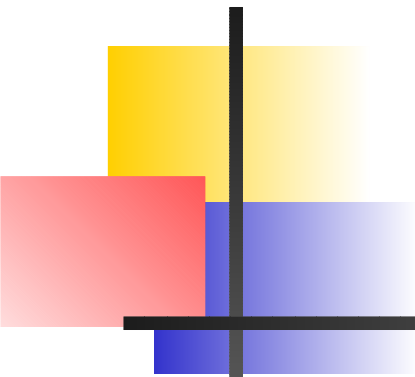
GG Frailty Model (cont.)

- It has considerable flexibility to capture the characteristics in a distribution that might have been missed by using its special cases [Yamaguchi (1992)].



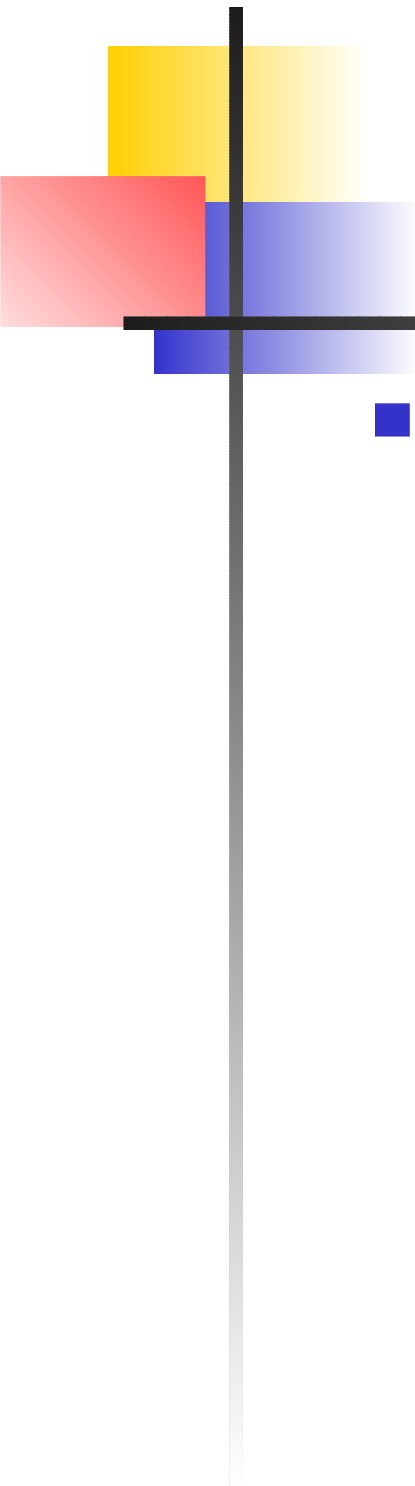
GG Frailty Model (cont.)

- It has considerable flexibility to capture the characteristics in a distribution that might have been missed by using its special cases [Yamaguchi (1992)].
- In the generalized gamma frailty model, the frailty term y_i in the frailty model follows the generalized gamma distribution.



GG Frailty Model (cont.)

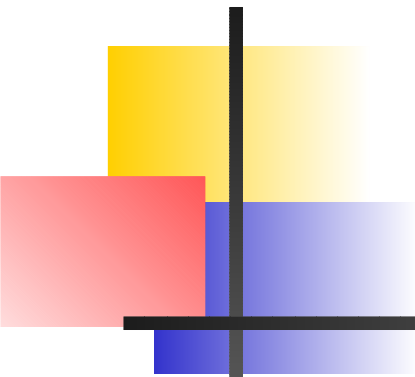
- It has considerable flexibility to capture the characteristics in a distribution that might have been missed by using its special cases [Yamaguchi (1992)].
- In the generalized gamma frailty model, the frailty term y_i in the frailty model follows the generalized gamma distribution.
- To make the parameters in the frailty distribution and the baseline distribution $h_0(t)$ identifiable, one usually sets the mean of the frailty distribution be one.



GG Frailty Model (cont.)

- The mean of the generalized gamma distribution is

$$\frac{\Gamma(q^{-2} + \frac{\sigma}{q})}{\Gamma(q^{-2})(q^{-2})^{\frac{\sigma}{q}} \lambda}, \quad q > -1/\sigma.$$



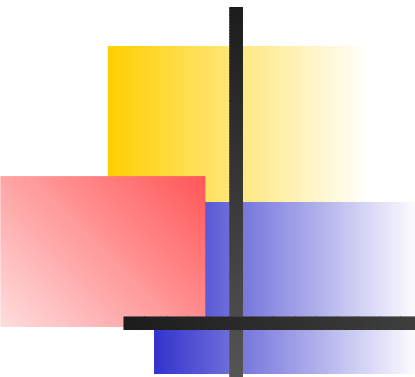
GG Frailty Model (cont.)

- The mean of the generalized gamma distribution is

$$\frac{\Gamma(q^{-2} + \frac{\sigma}{q})}{\Gamma(q^{-2})(q^{-2})^{\frac{\sigma}{q}} \lambda}, \quad q > -1/\sigma.$$

- If the mean is set to one, then

$$\lambda = \frac{\Gamma(q^{-2} + \frac{\sigma}{q})}{\Gamma(q^{-2})(q^{-2})^{\frac{\sigma}{q}}}.$$



GG Frailty Model (cont.)

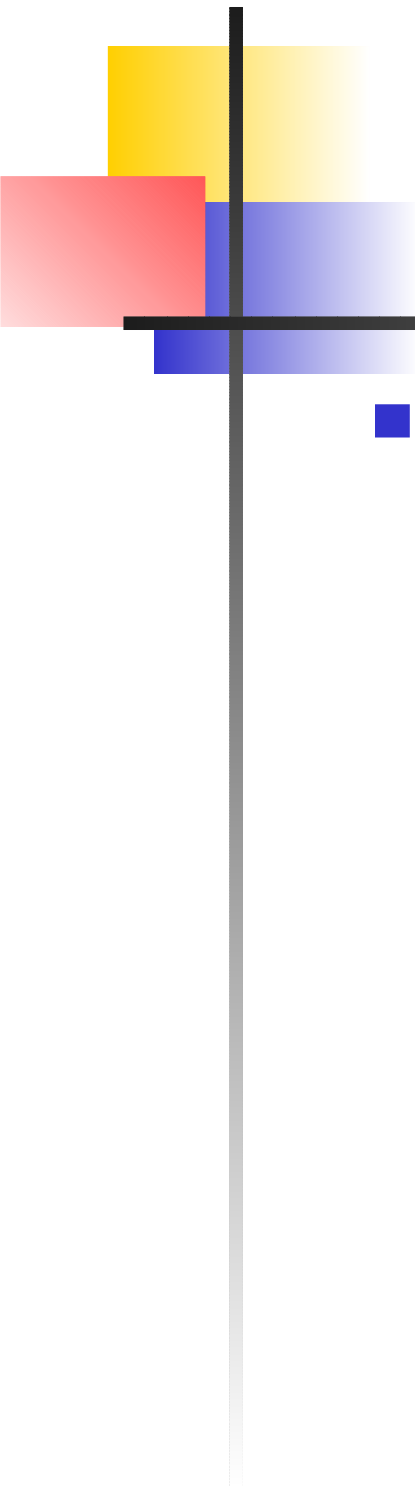
- The mean of the generalized gamma distribution is

$$\frac{\Gamma(q^{-2} + \frac{\sigma}{q})}{\Gamma(q^{-2})(q^{-2})^{\frac{\sigma}{q}} \lambda}, \quad q > -1/\sigma.$$

- If the mean is set to one, then

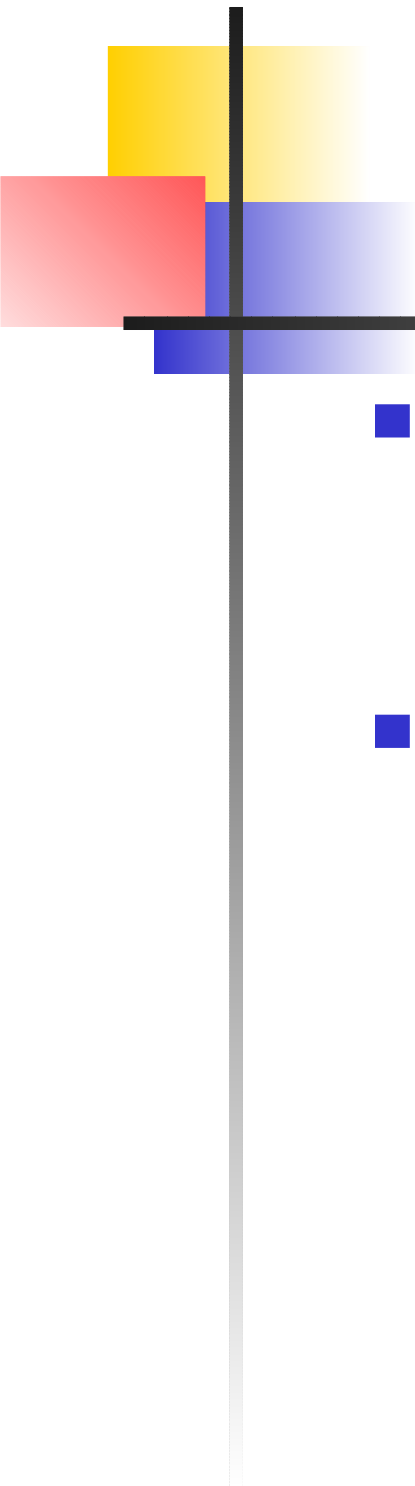
$$\lambda = \frac{\Gamma(q^{-2} + \frac{\sigma}{q})}{\Gamma(q^{-2})(q^{-2})^{\frac{\sigma}{q}}}.$$

- Gamma frailty [Klein (1992)] when $q/\sigma = 1$;
- Lognormal frailty model [McGilchrist and Aisbett (1991), McGilchrist (1993)] when $q = 0$.



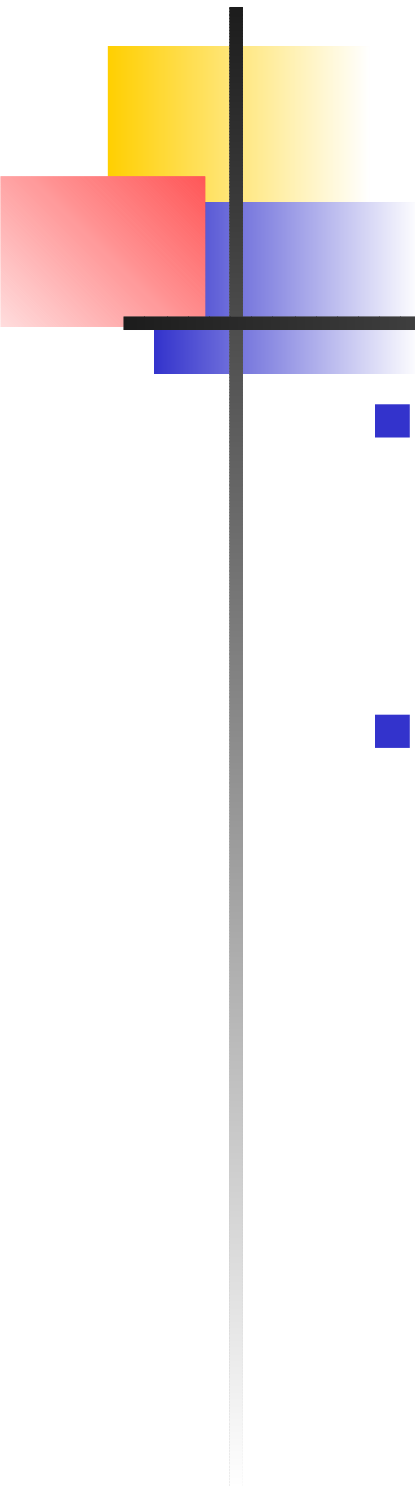
GG Frailty Model (cont.)

- If the baseline distribution $h_0(t)$ is specified up to a few parameters denoted by γ , the unknown parameters in this frailty model are q , σ , β and γ .



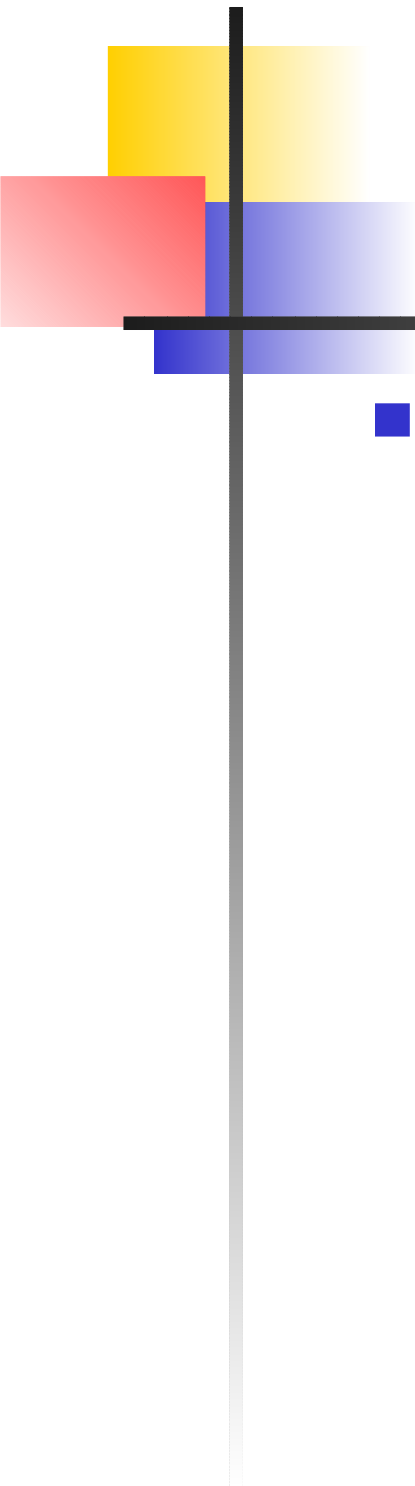
GG Frailty Model (cont.)

- If the baseline distribution $h_0(t)$ is specified up to a few parameters denoted by γ , the unknown parameters in this frailty model are q , σ , β and γ .
- The tails of a frailty distribution can determine the type of dependence among the correlated observations:



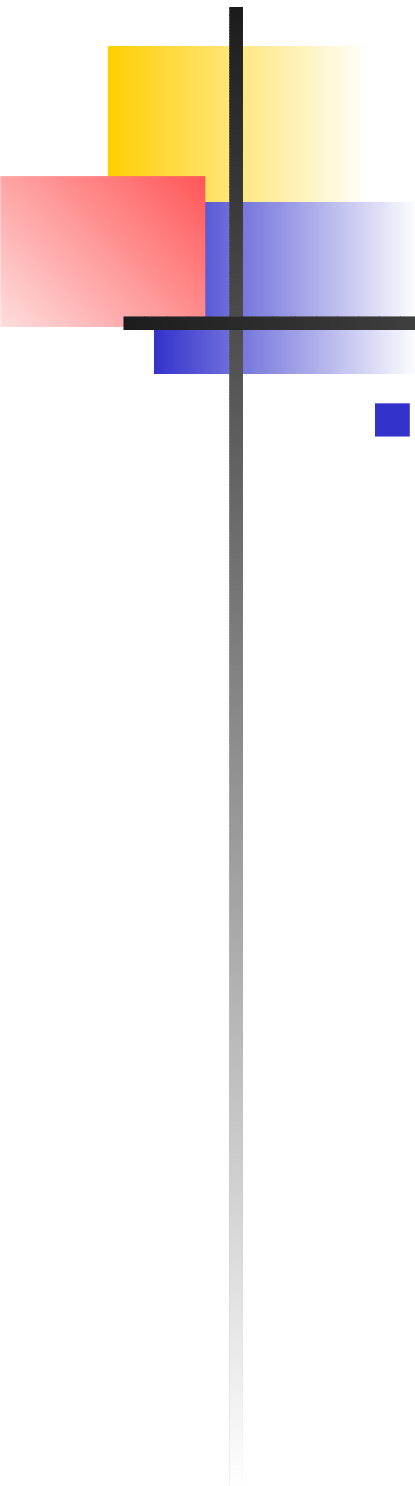
GG Frailty Model (cont.)

- If the baseline distribution $h_0(t)$ is specified up to a few parameters denoted by γ , the unknown parameters in this frailty model are q , σ , β and γ .
- The tails of a frailty distribution can determine the type of dependence among the correlated observations:
 - a large right tail corresponds to strong early dependence;
 - a large left tail corresponds to strong late dependence.



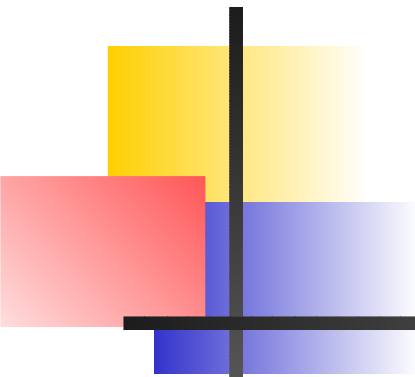
GG Frailty Model (cont.)

- The generalized gamma distribution is capable of handling both early and late dependences since:



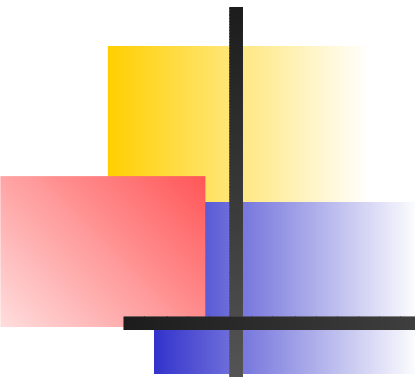
GG Frailty Model (cont.)

- The generalized gamma distribution is capable of handling both early and late dependences since:
 - it has a large left tail when $q > 0$, and is therefore suitable for strong late dependence;
 - a large right tail when $q < 0$, and is therefore suitable for strong early dependence.



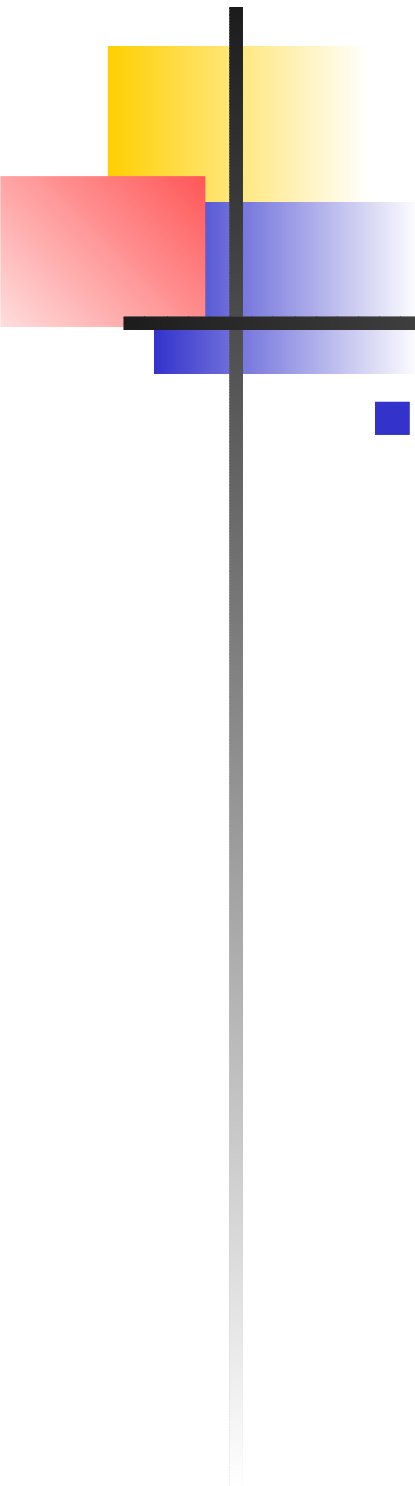
GG Frailty Model (cont.)

- The generalized gamma distribution is capable of handling both early and late dependences since:
 - it has a large left tail when $q > 0$, and is therefore suitable for strong late dependence;
 - a large right tail when $q < 0$, and is therefore suitable for strong early dependence.
- Gamma and Weibull frailty models are suitable only strong late dependence.



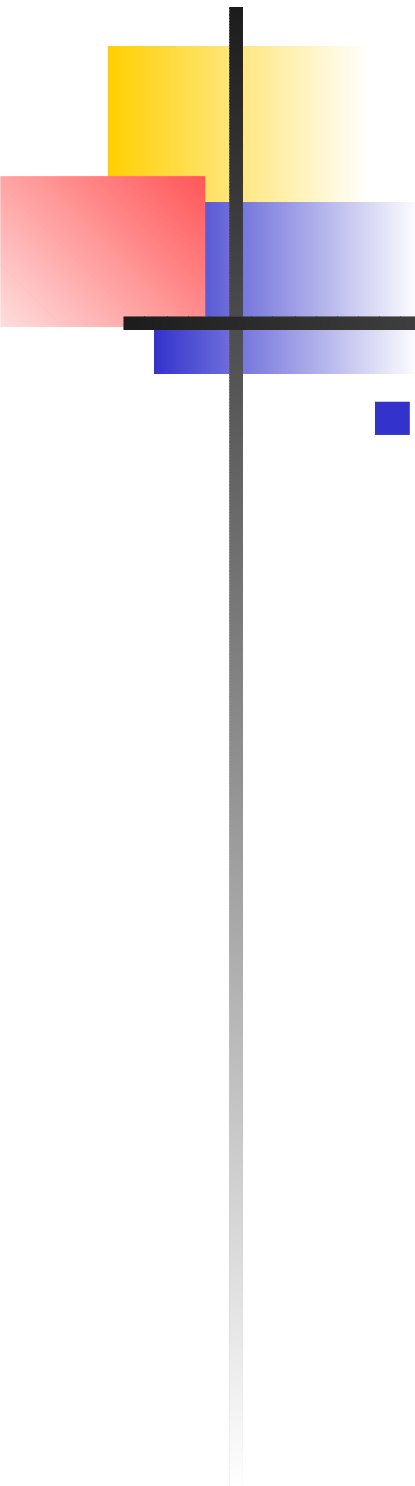
GG Frailty Model (cont.)

- The generalized gamma distribution is capable of handling both early and late dependences since:
 - it has a large left tail when $q > 0$, and is therefore suitable for strong late dependence;
 - a large right tail when $q < 0$, and is therefore suitable for strong early dependence.
- Gamma and Weibull frailty models are suitable only strong late dependence.
- Positive stable model is suitable only for strong early dependence.



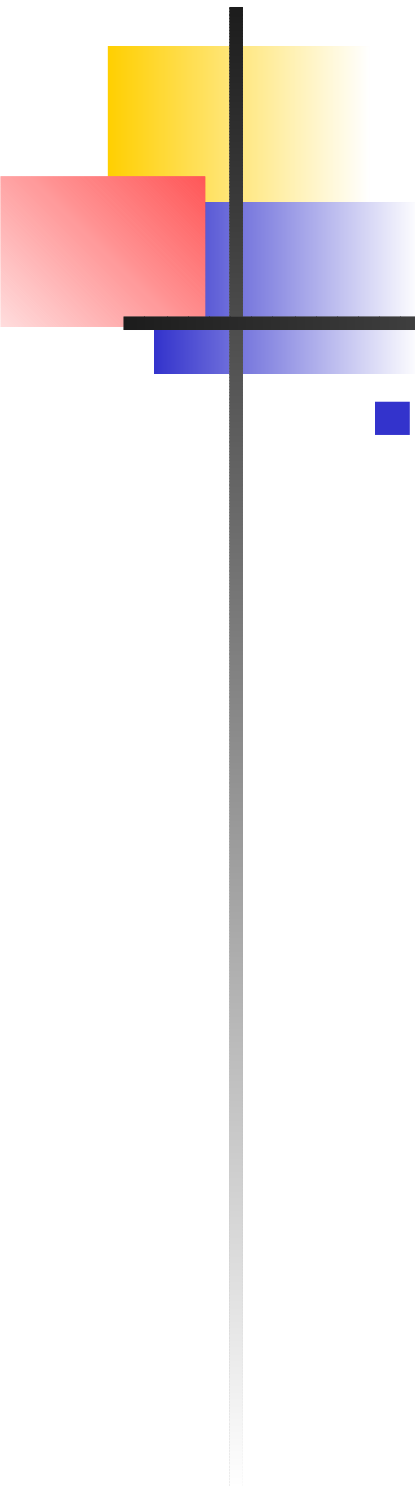
Estimation Method

- There are two main methods proposed in the literature to estimate the parameters of a frailty model.



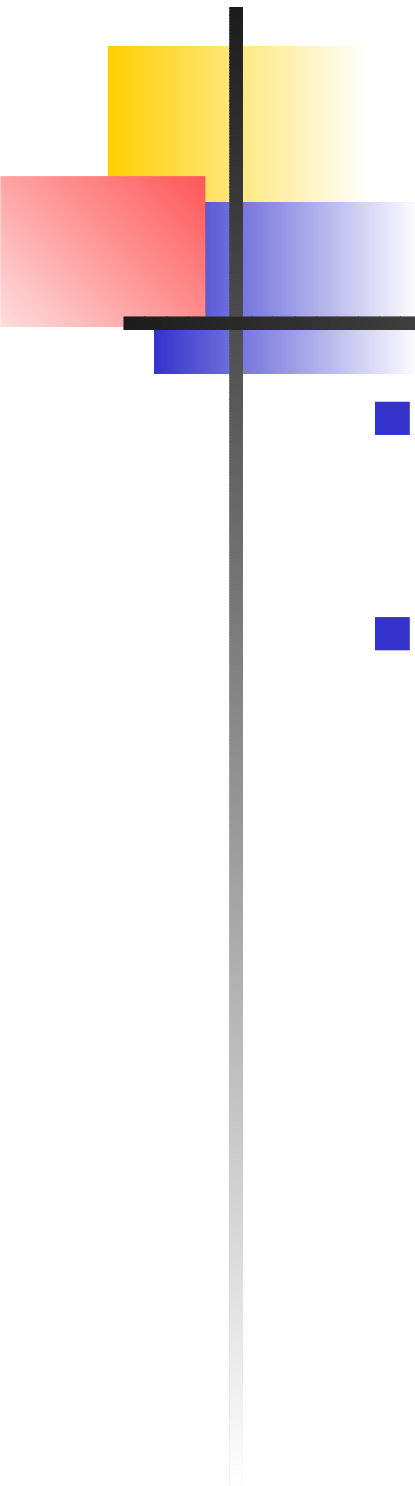
Estimation Method

- There are two main methods proposed in the literature to estimate the parameters of a frailty model.
 - The Maximum Likelihood method, combined with methods for Cox's proportional hazards model, provides MLE of β and a nonparametric estimate of $h_0(t)$;
 - The Best Linear Unbiased Predictor (BLUP) method, also combined with methods for Cox's proportional hazards model to get a nonparametric estimate of $h_0(t)$.



Estimation Method (cont.)

- We use the likelihood method to estimate the parameters of the GG frailty model.



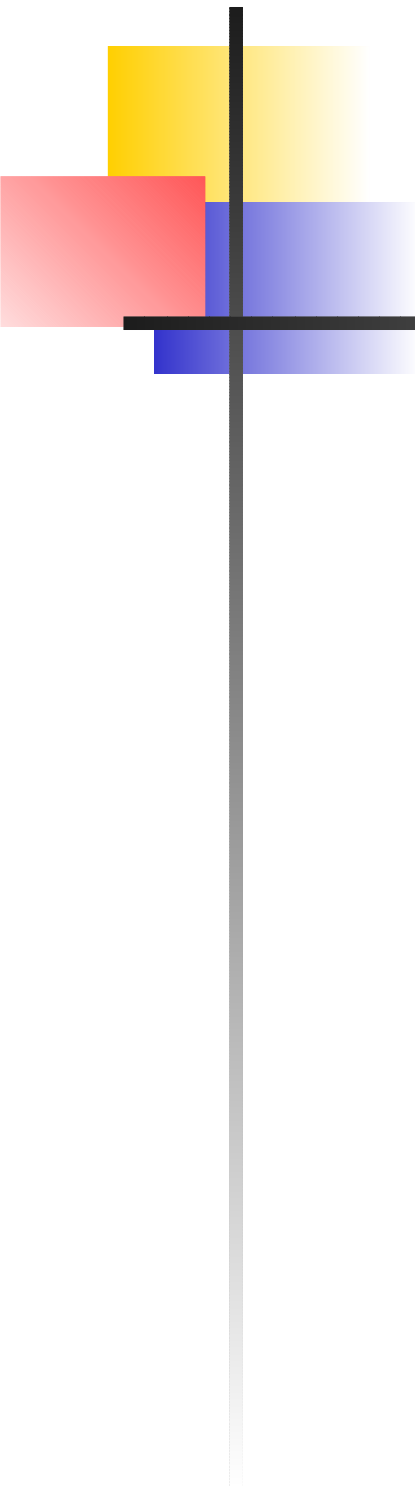
Estimation Method (cont.)

- We use the likelihood method to estimate the parameters of the GG frailty model.
- Given the observed data $(t_{ij}, \delta_{ij}, x_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, m_i$, the unconditional likelihood function is

Estimation Method (cont.)

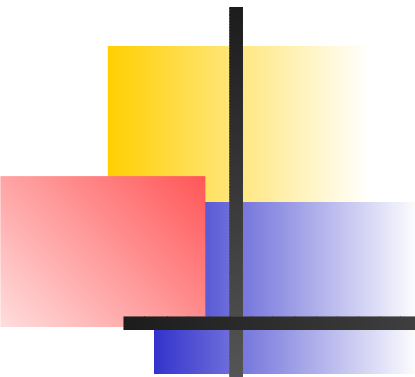
- We use the likelihood method to estimate the parameters of the GG frailty model.
- Given the observed data $(t_{ij}, \delta_{ij}, x_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, m_i$, the unconditional likelihood function is

$$\begin{aligned} L &= \prod_{i=1}^n \int_0^\infty \left(\prod_{j=1}^{m_i} [y_i h_0(t_{ij}; \gamma) e^{\beta' x_{ij}}]^{\delta_{ij}} S_0(t_{ij}; \gamma)^{y_i e^{\beta' x_{ij}}} \right) g(y_i; q, \sigma) dy_i \\ &= \prod_{i=1}^n \left(\prod_{j=1}^{m_i} [h_0(t_{ij}; \gamma) e^{\beta' x_{ij}}]^{\delta_{ij}} \right) \int_0^\infty y_i^{\delta_{i\cdot}} e^{-y_i A_i(\beta, \gamma)} g(y_i; q, \sigma) dy_i, \end{aligned}$$



Estimation Method (cont.)

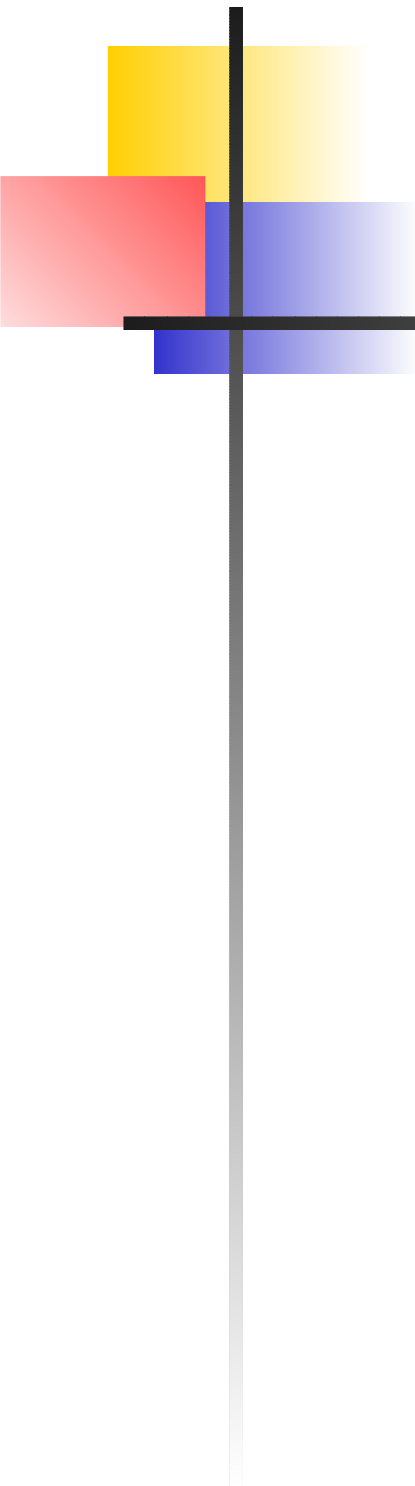
where



Estimation Method (cont.)

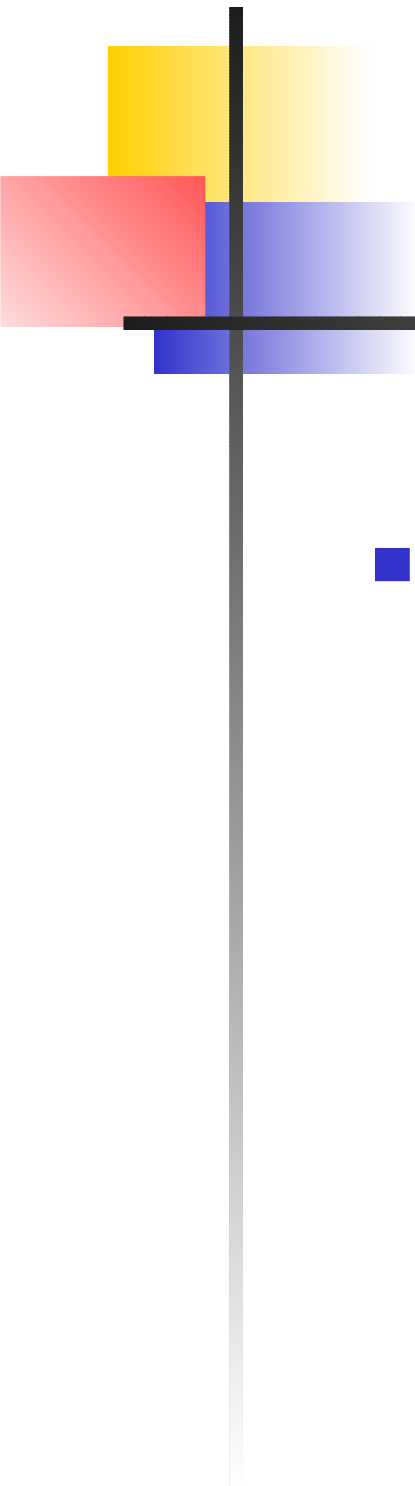
where

- $S_0(t; \gamma) = \exp[-\int_0^t h_0(u; \gamma) du]$ is the baseline survival function,
- $H_0(t; \gamma) = -\log S_0(t; \gamma)$ is the baseline cumulative hazard function,
- $\delta_{i\cdot} = \sum_{j=1}^{m_i} \delta_{ij}$,
- $A_i(\beta, \gamma) = \sum_{j=1}^{m_i} H_0(t_{ij}; \gamma) \exp(\beta' x_{ij})$.



Estimation Method (cont.)

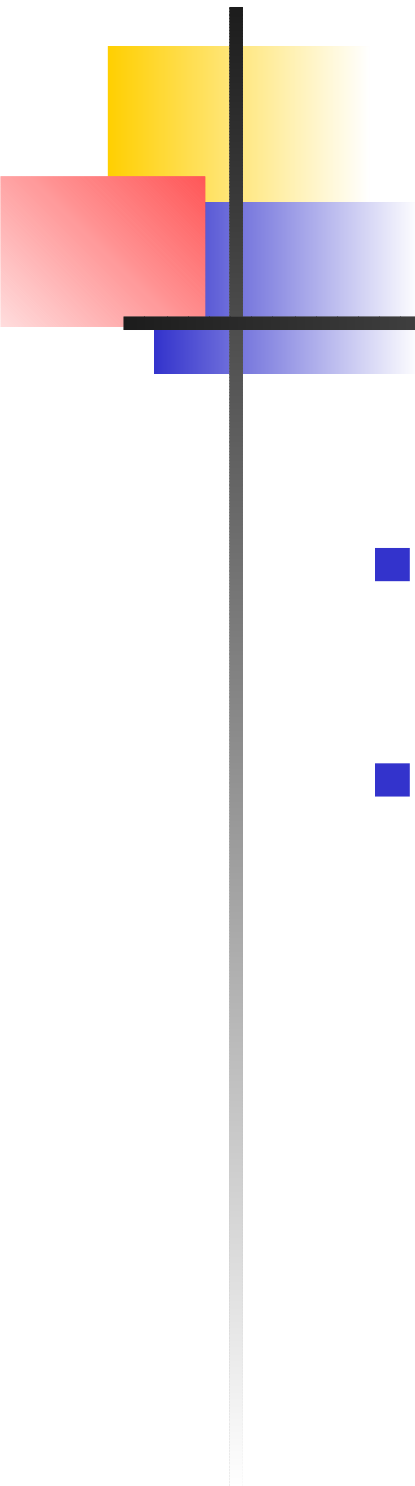
A Monte Carlo Approximation



Estimation Method (cont.)

A Monte Carlo Approximation

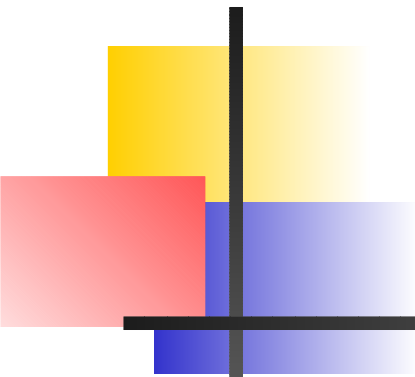
- The integration involved in L is intractable, and we use simulation to approximate the integration.



Estimation Method (cont.)

A Monte Carlo Approximation

- The integration involved in L is intractable, and we use simulation to approximate the integration.
- To simplify the simulation for the approximation, the integration part of the likelihood can be rewritten as

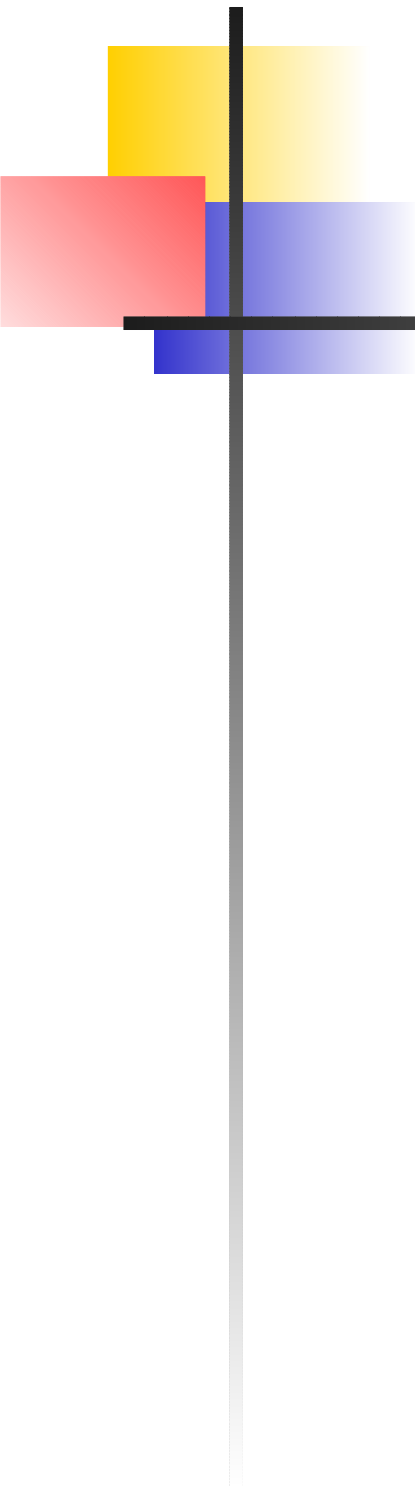


Estimation Method (cont.)

A Monte Carlo Approximation

- The integration involved in L is intractable, and we use simulation to approximate the integration.
- To simplify the simulation for the approximation, the integration part of the likelihood can be rewritten as

$$\begin{aligned} & \int_0^\infty y_i^{\delta_i} \exp(-y_i A_i(\beta, \gamma)) g(y_i; q, \sigma) dy_i \\ &= \int_0^\infty \left(\frac{y_i^{\sigma/q}}{\lambda} \right)^{\delta_i} \exp\left(-\frac{y_i^{\sigma/q}}{\lambda} A_i(\beta, \gamma)\right) f_g(y_i; q^{-2}, q^{-2}) dy_i, \end{aligned}$$



Estimation Method (cont.)

where λ is as before, and

$$f_g(y_i; q^{-2}, q^{-2}) = \frac{(q^{-2})^{q^{-2}} y_i^{q^{-2}-1} \exp[-q^{-2} y_i]}{\Gamma(q^{-2})}$$

is the density function of the gamma distribution with both shape and scale parameters equal to q^{-2} .

Estimation Method (cont.)

where λ is as before, and

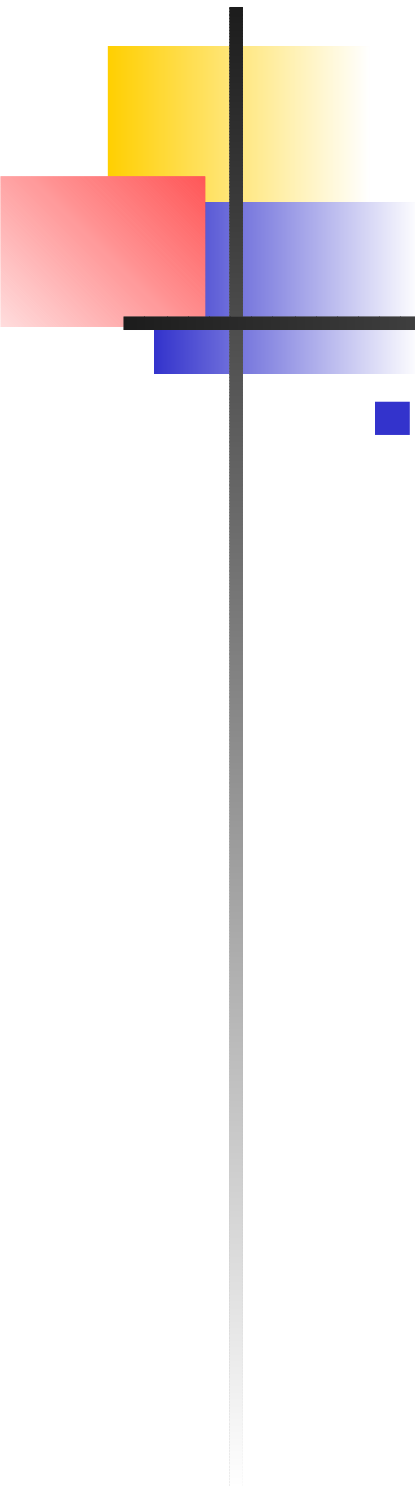
$$f_g(y_i; q^{-2}, q^{-2}) = \frac{(q^{-2})^{q^{-2}} y_i^{q^{-2}-1} \exp[-q^{-2} y_i]}{\Gamma(q^{-2})}$$

is the density function of the gamma distribution with both shape and scale parameters equal to q^{-2} .

- The required integration is then the expectation of

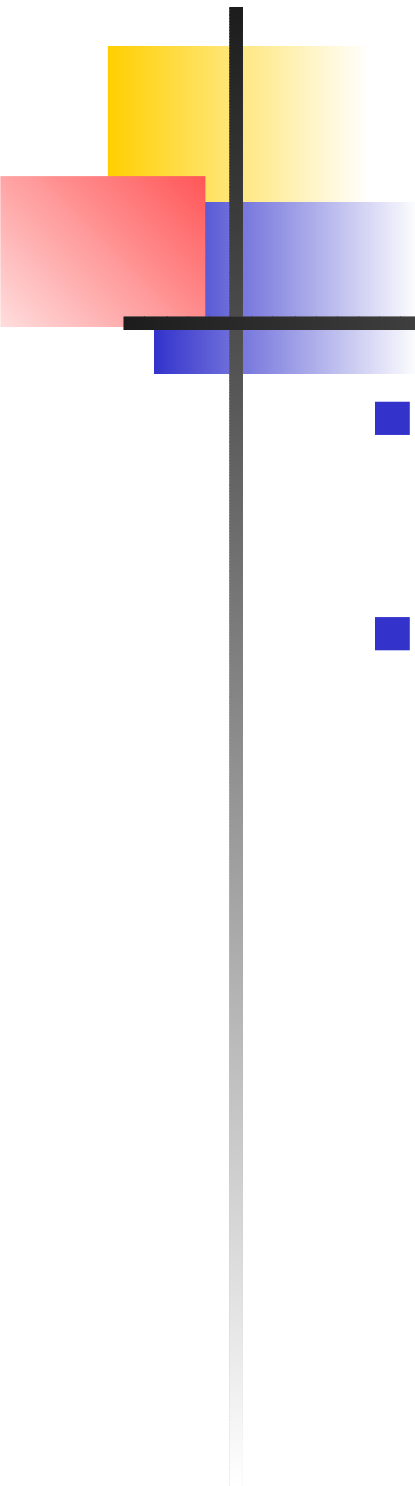
$$\left(\frac{y_i^{\sigma/q}}{\lambda} \right)^{\delta_i} \exp \left(- \frac{y_i^{\sigma/q}}{\lambda} A_i(\beta, \gamma) \right)$$

with respect to $y_i \sim f_g(y_i; q^{-2}, q^{-2})$ above.



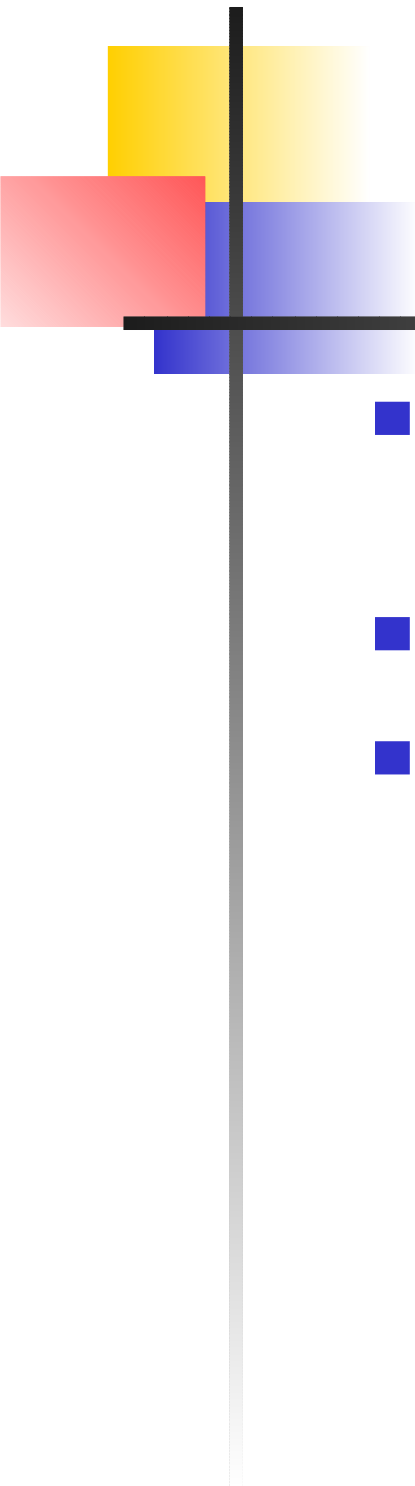
Estimation Method (cont.)

- We can now proceed to approximate the likelihood function via simulation as follows.



Estimation Method (cont.)

- We can now proceed to approximate the likelihood function via simulation as follows.
- Let $\ell(q, \sigma, \beta, \gamma) = \log[\mathbf{L}(q, \sigma, \beta, \gamma)]$ be the log-likelihood.



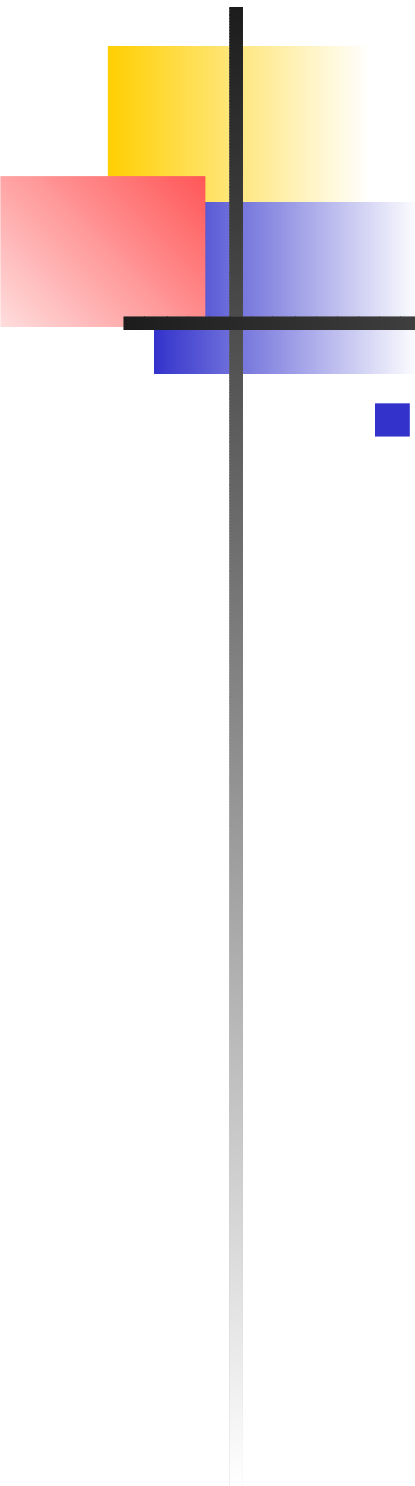
Estimation Method (cont.)

- We can now proceed to approximate the likelihood function via simulation as follows.
- Let $\ell(q, \sigma, \beta, \gamma) = \log[\mathbf{L}(q, \sigma, \beta, \gamma)]$ be the log-likelihood.
- For a fixed q , let $v_{(1)}, \dots, v_{(N)}$ be N independent realizations from gamma distn. $f_g(y_i; q^{-2}, q^{-2})$.

Estimation Method (cont.)

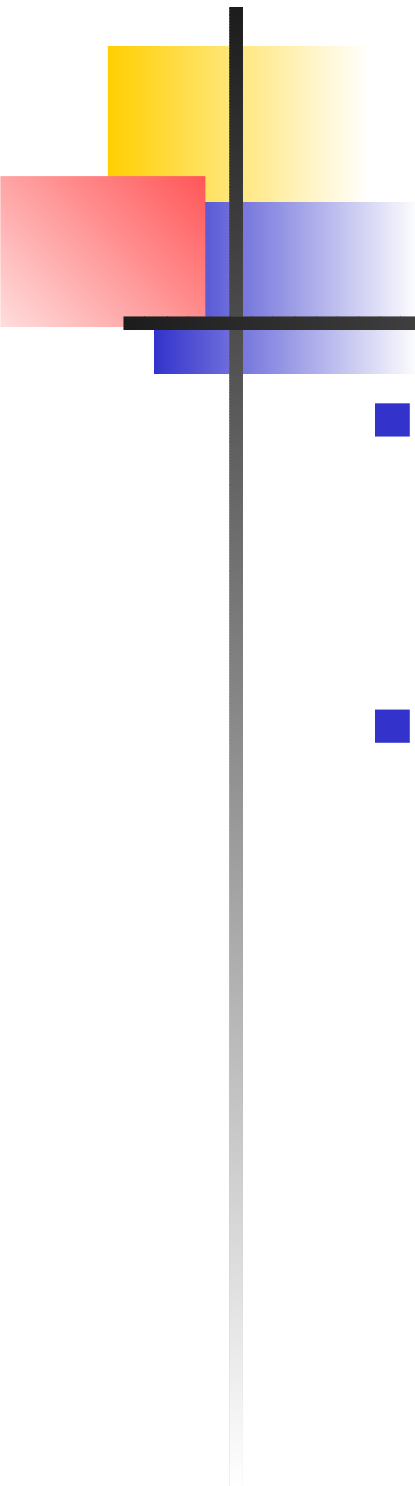
- We can now proceed to approximate the likelihood function via simulation as follows.
- Let $\ell(q, \sigma, \beta, \gamma) = \log[\mathbf{L}(q, \sigma, \beta, \gamma)]$ be the log-likelihood.
- For a fixed q , let $v_{(1)}, \dots, v_{(N)}$ be N independent realizations from gamma distn. $f_g(y_i; q^{-2}, q^{-2})$.
- The log-likelihood function can be approximated by

$$\begin{aligned} \widehat{\ell}(q, \sigma, \beta, \gamma) = & \sum_{i=1}^n \left[\left(\sum_{j=1}^{m_i} \delta_{ij} [\log h_0(t_{ij}; \gamma) + \beta' x_{ij}] \right) \right. \\ & \left. + \log \frac{1}{N} \sum_{k=1}^N \left(\frac{v_{(k)}^{\sigma/q}}{\lambda} \right)^{\delta_i} \exp \left(-\frac{v_{(k)}^{\sigma/q} A_i(\beta, \gamma)}{\lambda} \right) \right]. \end{aligned}$$



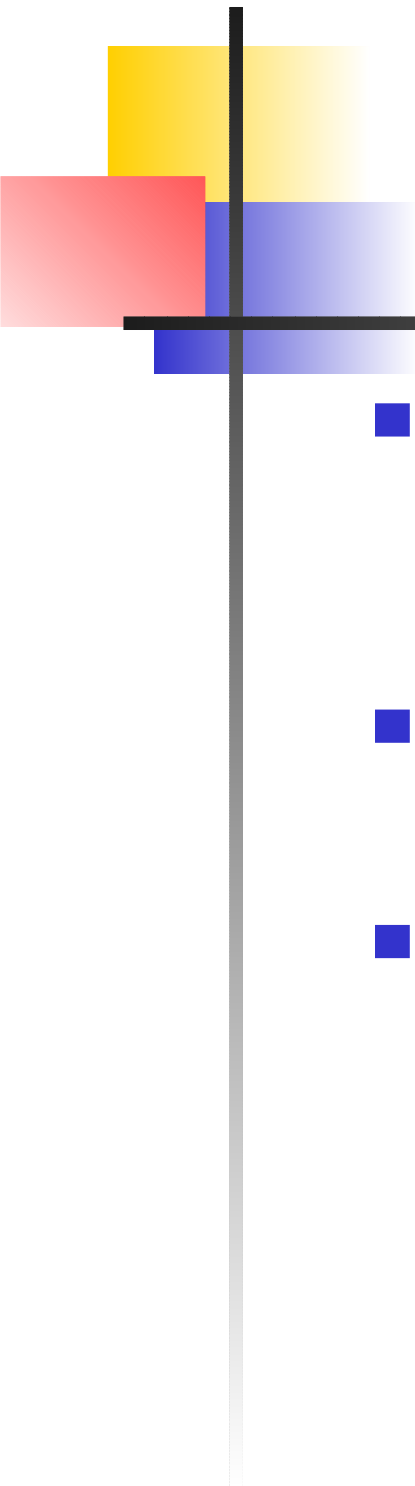
Estimation Method (cont.)

- Similarly, the first and second derivatives (Hessian matrix) of the log-likelihood function with respect to β , σ and γ can be approximated by the simulation.



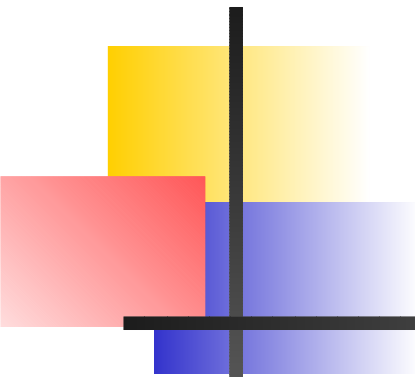
Estimation Method (cont.)

- Similarly, the first and second derivatives (Hessian matrix) of the log-likelihood function with respect to β , σ and γ can be approximated by the simulation.
- Hence for a fixed value of q , the log-likelihood function can be maximized using the Newton-Raphson method.



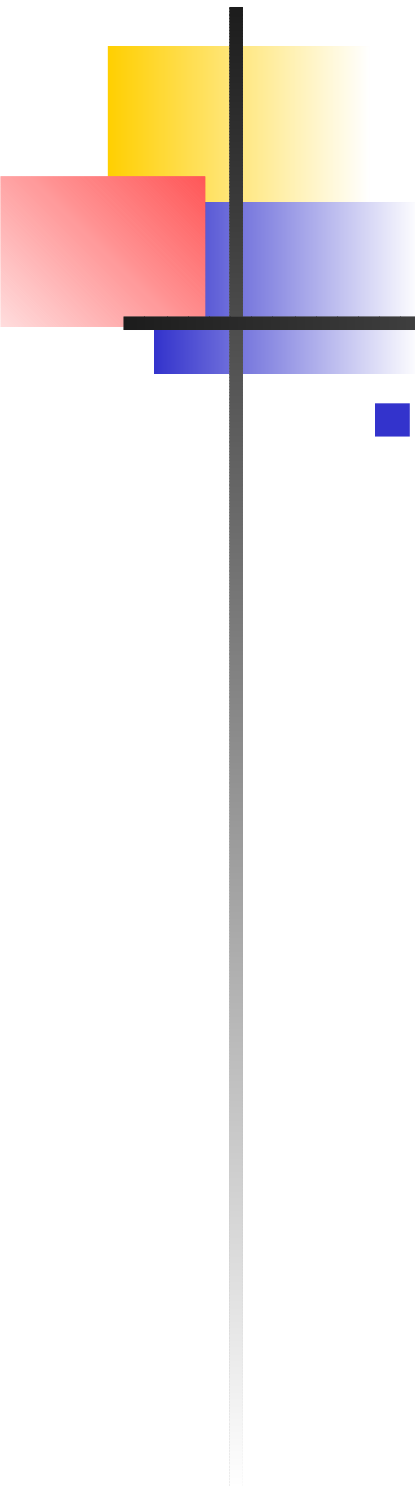
Estimation Method (cont.)

- Similarly, the first and second derivatives (Hessian matrix) of the log-likelihood function with respect to β , σ and γ can be approximated by the simulation.
- Hence for a fixed value of q , the log-likelihood function can be maximized using the Newton-Raphson method.
- The standard errors of the estimates can be approximated from the inverse of the Hessian matrix.



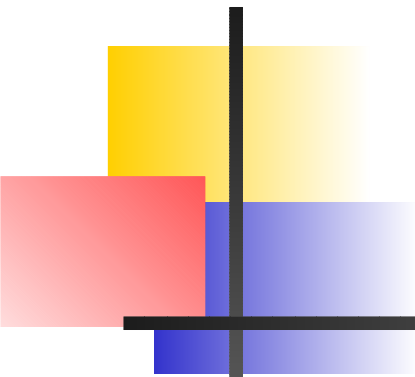
Estimation Method (cont.)

- Similarly, the first and second derivatives (Hessian matrix) of the log-likelihood function with respect to β , σ and γ can be approximated by the simulation.
- Hence for a fixed value of q , the log-likelihood function can be maximized using the Newton-Raphson method.
- The standard errors of the estimates can be approximated from the inverse of the Hessian matrix.
- The same set of $v_{(1)}, \dots, v_{(N)}$ can be used in the maximization regardless of the values of β , σ and γ , in contrast to usual Monte Carlo methods that require a new set of $v_{(1)}, \dots, v_{(N)}$ for each parameter value.



Estimation Method (cont.)

- The parameter q can be estimated by profile likelihood.



Estimation Method (cont.)

- The parameter q can be estimated by profile likelihood.
- When $q = 0$, let $y_i = v_i^\sigma$. Then, the required intergral is

$$\int_0^\infty v_i^{\sigma \delta_i} \exp(-v_i^\sigma A_i(\beta, \gamma)) g_v(v_i) dv_i,$$

where $g_v(v) = e^{-(\log v)^2/2} / (\sqrt{2\pi}v)$ is the standard lognormal pdf.

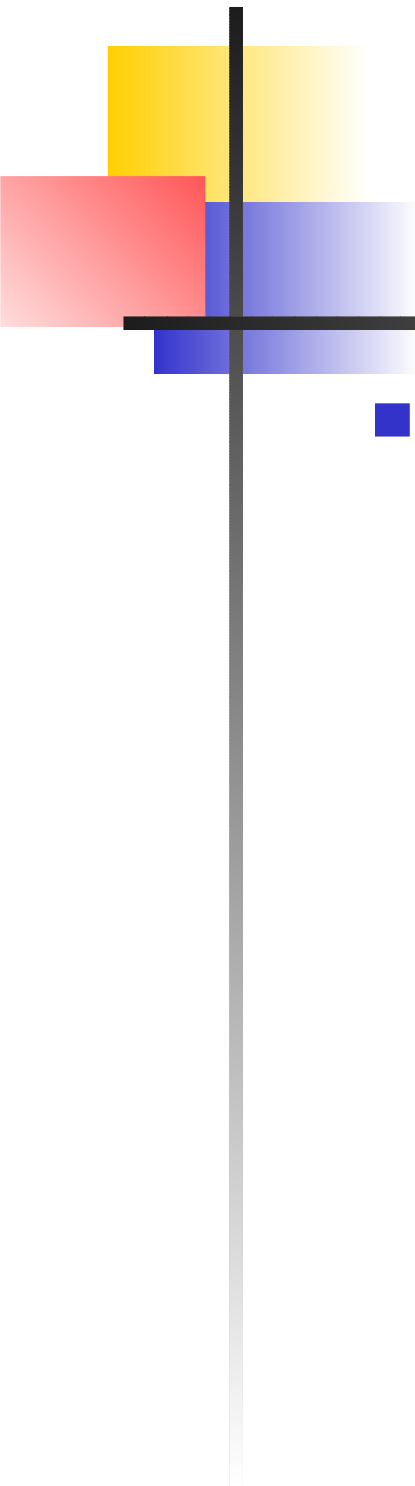
Estimation Method (cont.)

- The parameter q can be estimated by profile likelihood.
- When $q = 0$, let $y_i = v_i^\sigma$. Then, the required intergral is

$$\int_0^\infty v_i^{\sigma \delta_i} \exp(-v_i^\sigma A_i(\beta, \gamma)) g_v(v_i) dv_i,$$

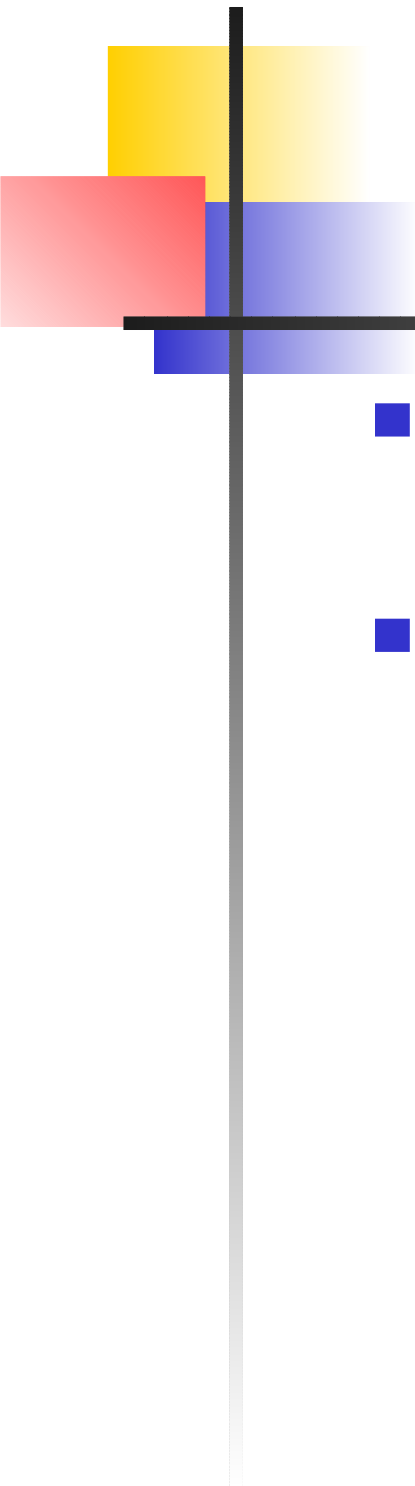
where $g_v(v) = e^{-(\log v)^2/2} / (\sqrt{2\pi}v)$ is the standard lognormal pdf.

- Since the standard lognormal density does not involve any parameter, the Monte Carlo approximation can be done with the same set of N independent realizations from the standard lognormal.



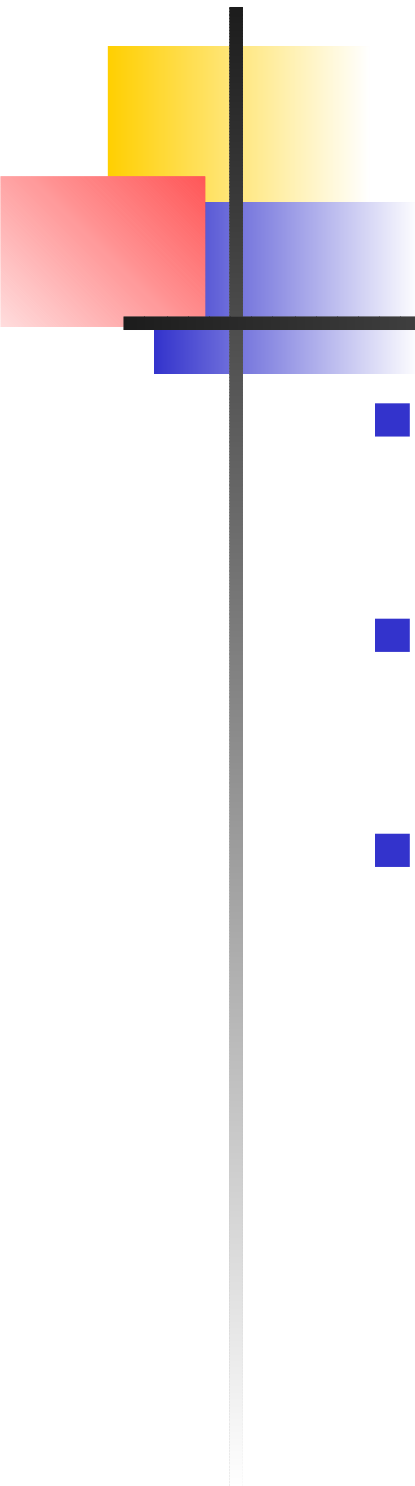
Estimation Method (cont.)

- Monte Carlo simulation approach is conceptually simple, but its convergence is notoriously slow.



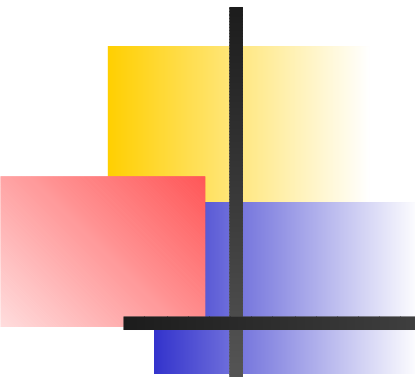
Estimation Method (cont.)

- Monte Carlo simulation approach is conceptually simple, but its convergence is notoriously slow.
- To achieve an accuracy of 0.0001, for example, Monte Carlo simulation size N required is well above 10^5 .



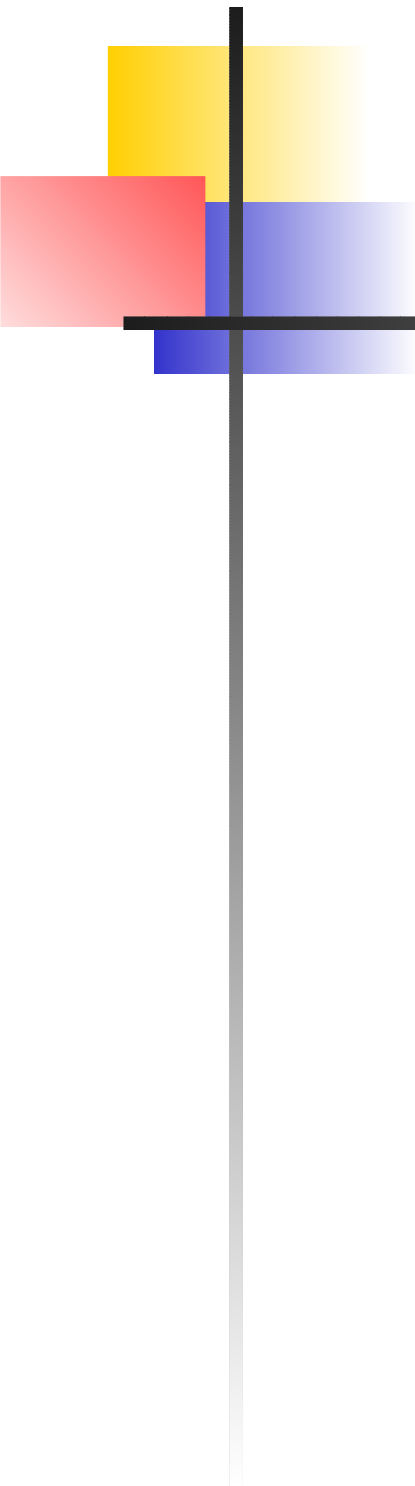
Estimation Method (cont.)

- Monte Carlo simulation approach is conceptually simple, but its convergence is notoriously slow.
- To achieve an accuracy of 0.0001, for example, Monte Carlo simulation size N required is well above 10^5 .
- So, one could instead approximate the required integrals and the derivatives by the Gauss-Kronrod quadrature, which is available in S-Plus and R.



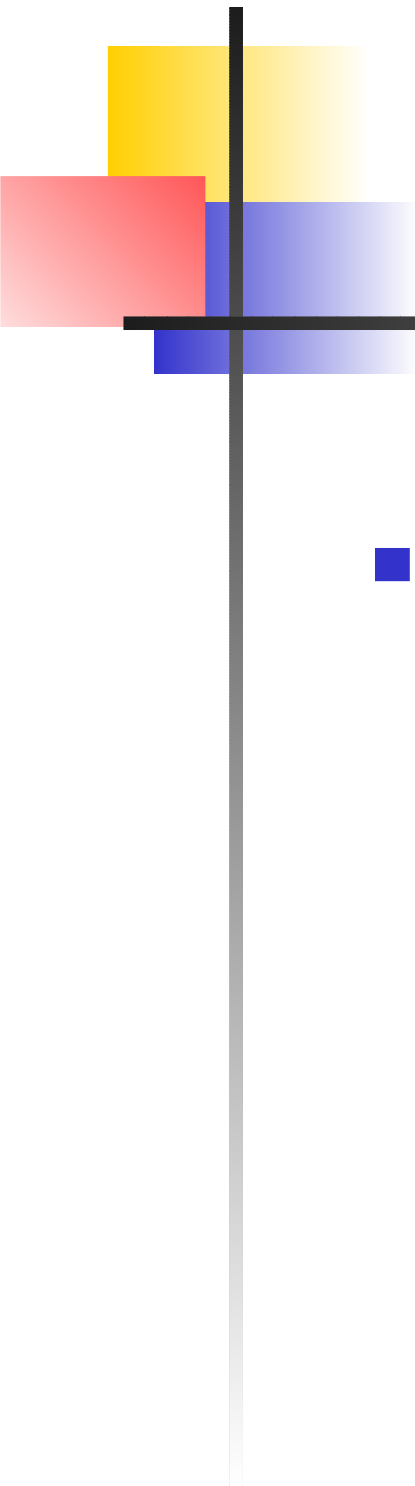
Estimation Method (cont.)

- Monte Carlo simulation approach is conceptually simple, but its convergence is notoriously slow.
- To achieve an accuracy of 0.0001, for example, Monte Carlo simulation size N required is well above 10^5 .
- So, one could instead approximate the required integrals and the derivatives by the Gauss-Kronrod quadrature, which is available in S-Plus and R.
- Standard errors of the estimates of β, σ, γ can be approximated from Hessian matrix for fixed q (at $\hat{q}$); if q is not fixed, they may be underestimated and so bootstrap method may be adopted.



Estimation Method (cont.)

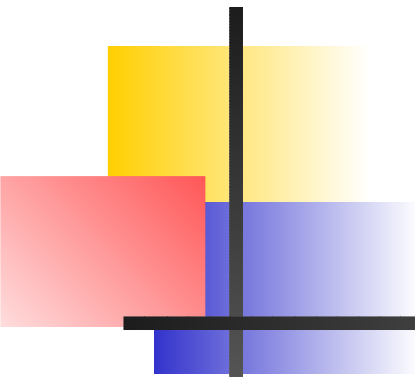
Choices of the baseline distribution



Estimation Method (cont.)

Choices of the baseline distribution

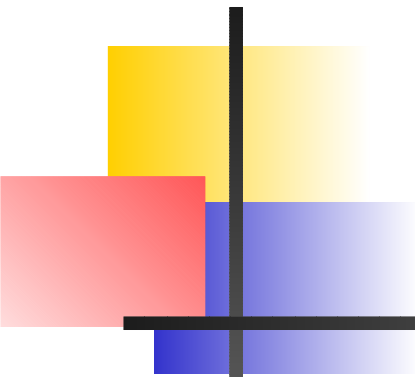
- We have so far assumed that the baseline hazard function $h_0(t; \gamma)$ is specified up to a few unknown parameters γ .



Estimation Method (cont.)

Choices of the baseline distribution

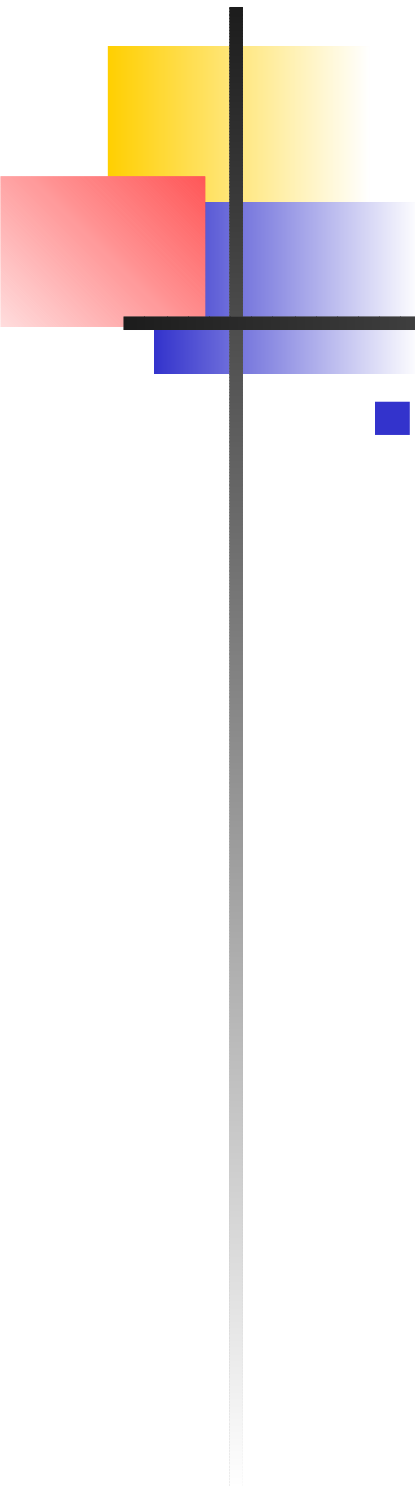
- We have so far assumed that the baseline hazard function $h_0(t; \gamma)$ is specified up to a few unknown parameters γ .
- For example, $h_0(t; \gamma)$ is a hazard function from the exponential or Weibull distribution and γ is the set of parameters in the distribution.



Estimation Method (cont.)

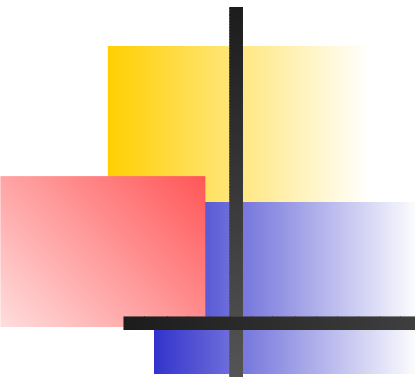
Choices of the baseline distribution

- We have so far assumed that the baseline hazard function $h_0(t; \gamma)$ is specified up to a few unknown parameters γ .
- For example, $h_0(t; \gamma)$ is a hazard function from the exponential or Weibull distribution and γ is the set of parameters in the distribution.
- Even though this method is simple to use, its strong parametric nature is not always desirable and the parametric assumption is usually difficult to verify.



Estimation Method (cont.)

- We could use a piecewise exponential distribution to approximate the baseline distribution if any parametric assumption on the distribution is not desirable.

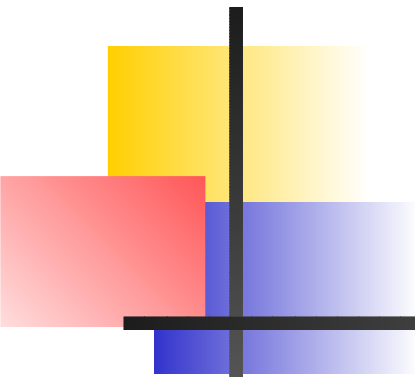


Estimation Method (cont.)

- We could use a piecewise exponential distribution to approximate the baseline distribution if any parametric assumption on the distribution is not desirable.
- We assume the baseline hazard rate as

$$h_0(t|\gamma) = \gamma_j \text{ for } t^{(j)} \leq t < t^{(j+1)}, \quad j = 1, \dots, k,$$

where $t^{(1)} \leq \min(t_i, i = 1, \dots, n) < t^{(2)} < \dots < \max(t_i, i = 1, \dots, n) \leq t^{(k+1)}$ are prespecified cut points that divide observations into k groups.



Estimation Method (cont.)

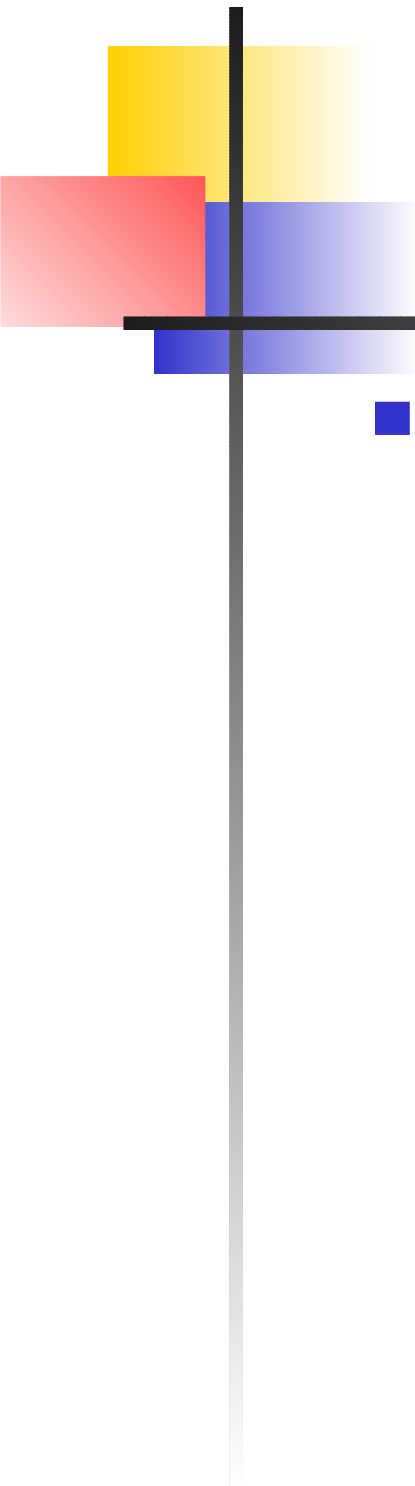
- We could use a piecewise exponential distribution to approximate the baseline distribution if any parametric assumption on the distribution is not desirable.

- We assume the baseline hazard rate as

$$h_0(t|\gamma) = \gamma_j \text{ for } t^{(j)} \leq t < t^{(j+1)}, \quad j = 1, \dots, k,$$

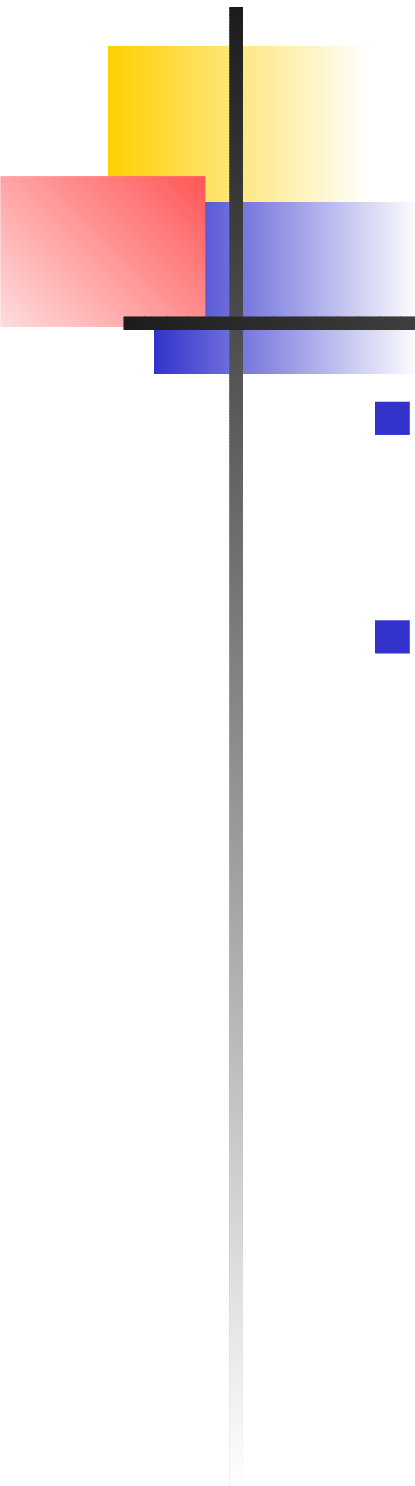
where $t^{(1)} \leq \min(t_i, i = 1, \dots, n) < t^{(2)} < \dots < \max(t_i, i = 1, \dots, n) \leq t^{(k+1)}$ are prespecified cut points that divide observations into k groups.

- The corresponding cumulative hazard function is
$$H(t|\gamma) = \sum_{j=1}^i \gamma_j [\min(t, t^{(j+1)}) - t^{(j)}] \text{ if } t^{(i)} \leq t < t^{(i+1)}.$$



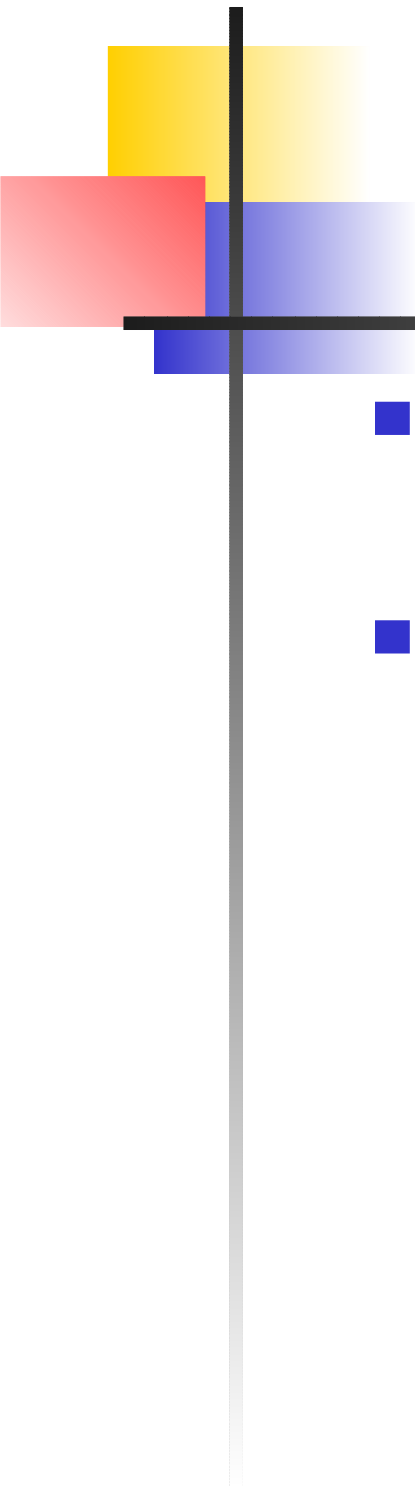
Simulation Study

- We conducted a simulation study to evaluate the performance of the model and the estimation method.



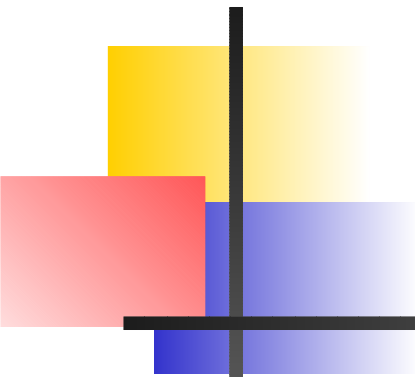
Simulation Study

- We conducted a simulation study to evaluate the performance of the model and the estimation method.
- We considered the following cases:



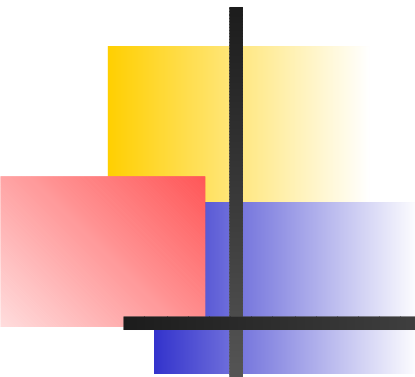
Simulation Study

- We conducted a simulation study to evaluate the performance of the model and the estimation method.
- We considered the following cases:
 - $n = 200$ clusters, $m_i = 2$ failures in each cluster;
 - $n = 100$ clusters, $m_i = 4$ failures in each cluster;
 - $n = 100$ clusters, $m_i = 8$ failures in each cluster.



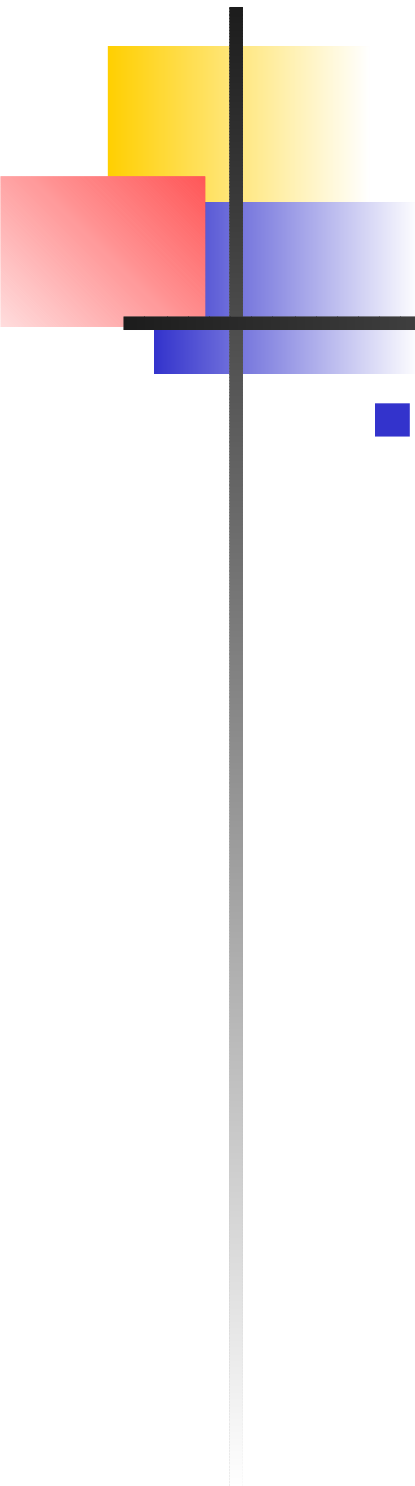
Simulation Study

- We conducted a simulation study to evaluate the performance of the model and the estimation method.
- We considered the following cases:
 - $n = 200$ clusters, $m_i = 2$ failures in each cluster;
 - $n = 100$ clusters, $m_i = 4$ failures in each cluster;
 - $n = 100$ clusters, $m_i = 8$ failures in each cluster.
- Clusters can be regarded as hospitals and each failure as from a patient in the hospital.



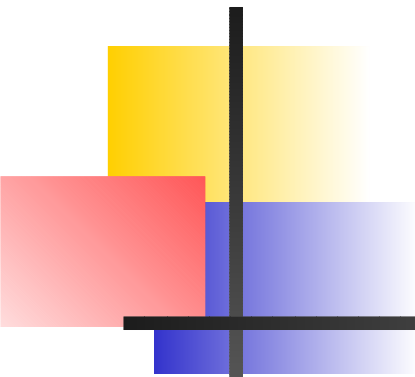
Simulation Study

- We conducted a simulation study to evaluate the performance of the model and the estimation method.
- We considered the following cases:
 - $n = 200$ clusters, $m_i = 2$ failures in each cluster;
 - $n = 100$ clusters, $m_i = 4$ failures in each cluster;
 - $n = 100$ clusters, $m_i = 8$ failures in each cluster.
- Clusters can be regarded as hospitals and each failure as from a patient in the hospital.
- Patients from each hospital are randomly assigned to either a treatment group or to a control group.



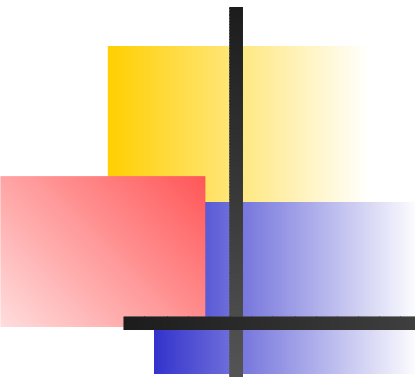
Simulation Study (cont.)

- Failures times from the same hospital (cluster) tend to correlate due to the same environmental or facility effect, which is described by a frailty term following



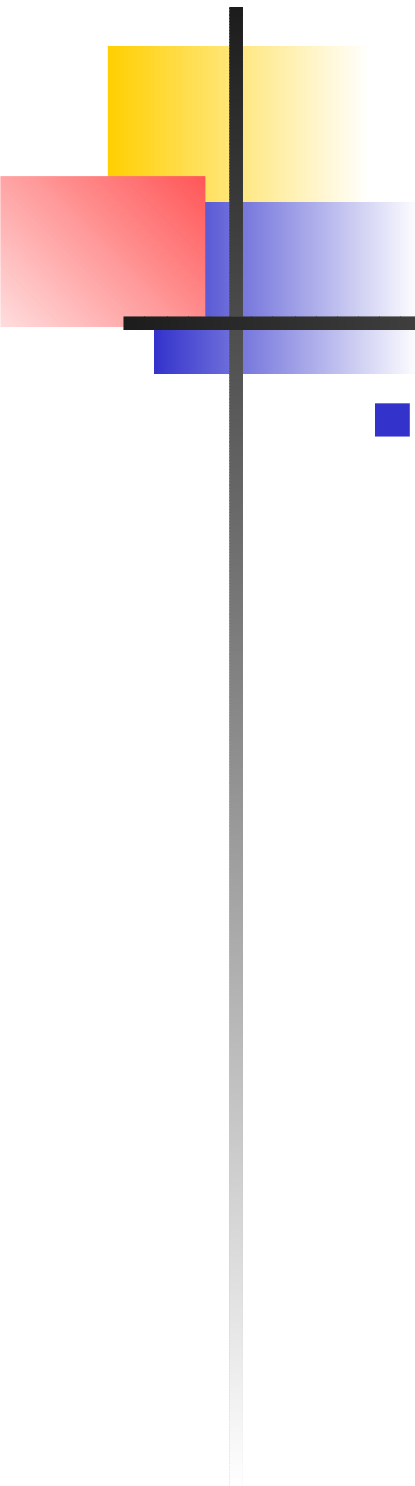
Simulation Study (cont.)

- Failures times from the same hospital (cluster) tend to correlate due to the same environmental or facility effect, which is described by a frailty term following
 - either the lognormal with $\lambda = \sqrt{1.5}$ and $\sigma^2 = \log 1.5$
 - or the gamma distribution with shape parameter as 2 and scale parameter as 0.5
(or $\sigma = q = 1/\sqrt{2}$ and $\lambda = 1$).



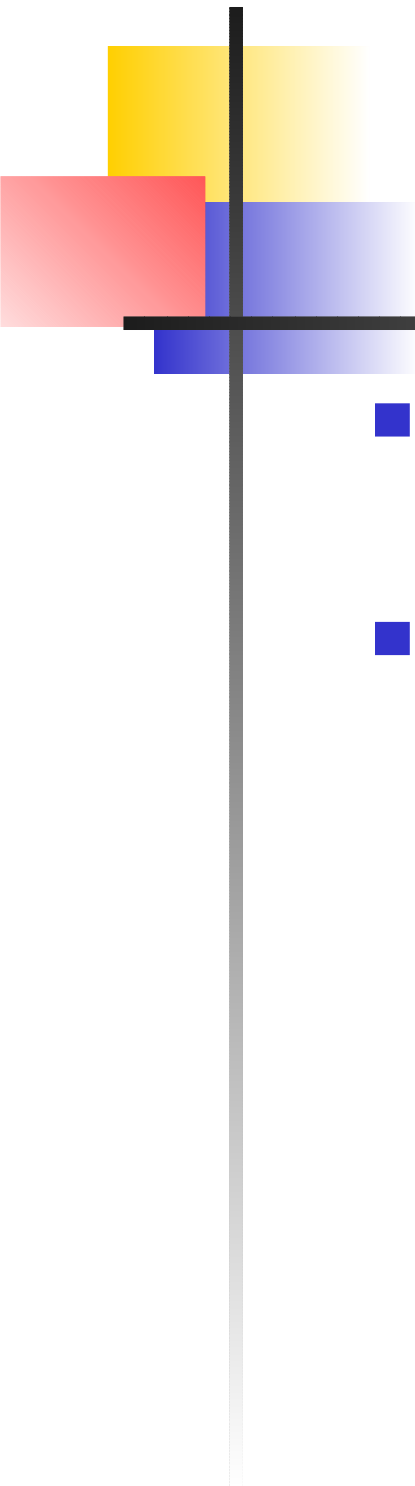
Simulation Study (cont.)

- Failures times from the same hospital (cluster) tend to correlate due to the same environmental or facility effect, which is described by a frailty term following
 - either the lognormal with $\lambda = \sqrt{1.5}$ and $\sigma^2 = \log 1.5$
 - or the gamma distribution with shape parameter as 2 and scale parameter as 0.5
(or $\sigma = q = 1/\sqrt{2}$ and $\lambda = 1$).
- This way, the mean and variance of the frailty distributions are 1 and 0.5, respectively.



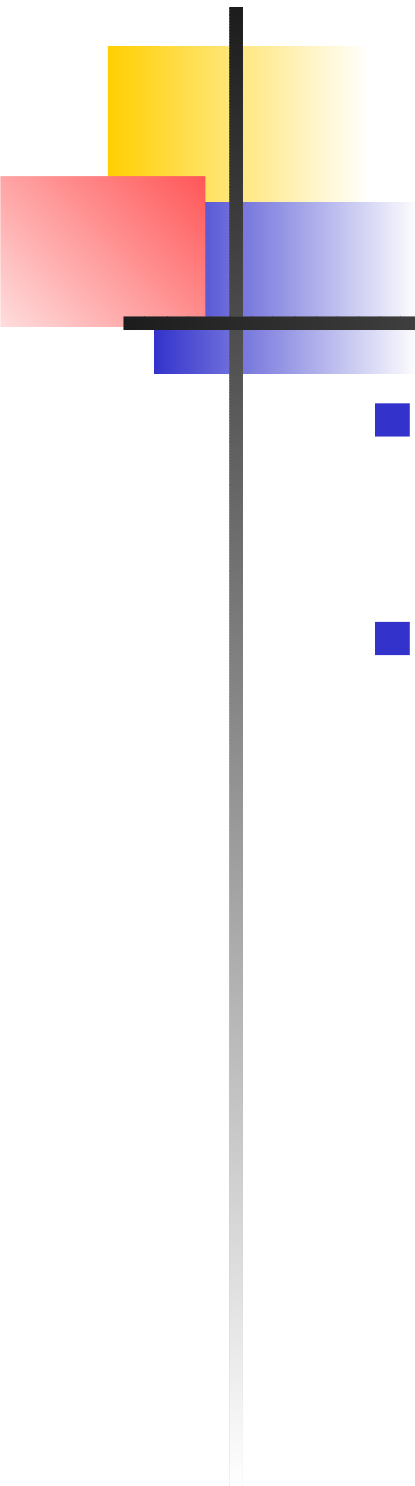
Simulation Study (cont.)

- Given the frailty term, failure times are independent and follow a proportional hazards model.



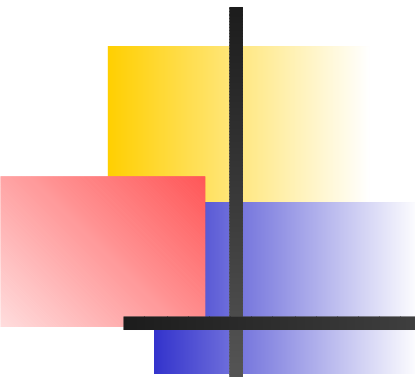
Simulation Study (cont.)

- Given the frailty term, failure times are independent and follow a proportional hazards model.
- Two distributions are considered as baseline:



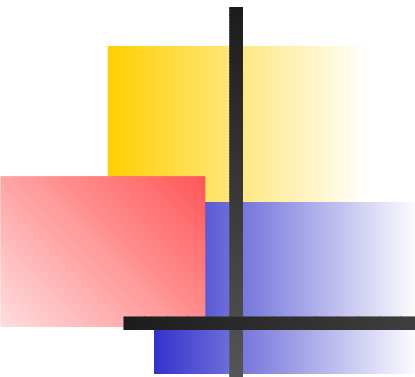
Simulation Study (cont.)

- Given the frailty term, failure times are independent and follow a proportional hazards model.
- Two distributions are considered as baseline:
 - Standard exponential with mean 1;
 - Lognormal with $\lambda = 1$ and $\sigma = 1$.



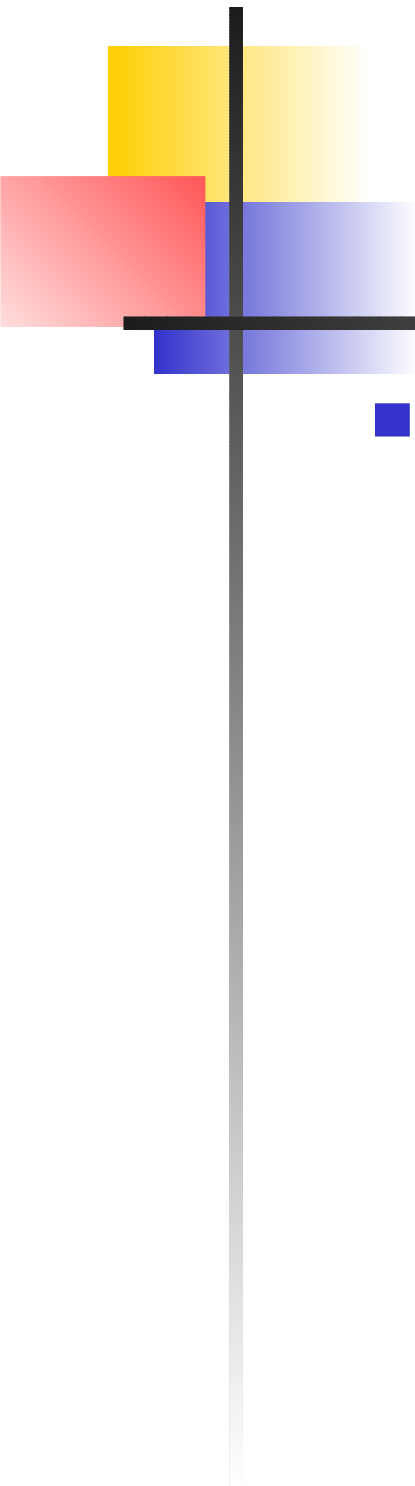
Simulation Study (cont.)

- Given the frailty term, failure times are independent and follow a proportional hazards model.
- Two distributions are considered as baseline:
 - Standard exponential with mean 1;
 - Lognormal with $\lambda = 1$ and $\sigma = 1$.
- We set $\beta = -\log 2$ so that the hazard rate of patients in the treatment group is half of the hazard rate in the control group (the baseline hazard rate).



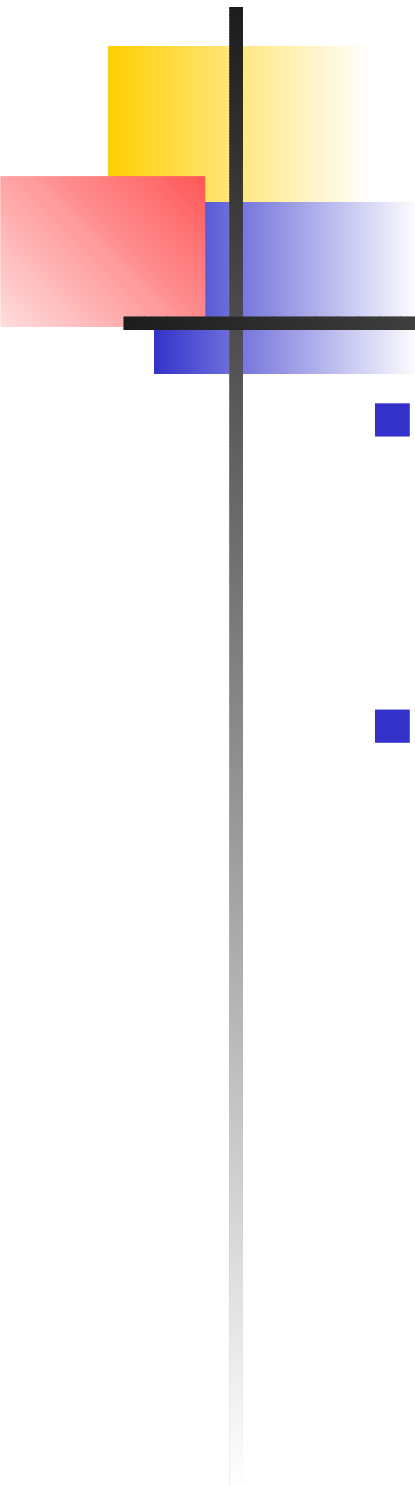
Simulation Study (cont.)

- Given the frailty term, failure times are independent and follow a proportional hazards model.
- Two distributions are considered as baseline:
 - Standard exponential with mean 1;
 - Lognormal with $\lambda = 1$ and $\sigma = 1$.
- We set $\beta = -\log 2$ so that the hazard rate of patients in the treatment group is half of the hazard rate in the control group (the baseline hazard rate).
- Censoring times follow a uniform distribution in $[0, 4.5]$ so that the resulting censoring rates are about 37% for lognormal frailty and 39% for gamma frailty.



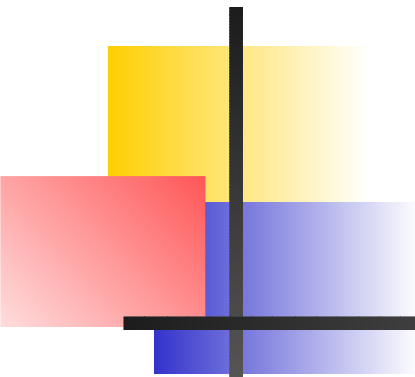
Simulation Study (cont.)

- All the models considered performed almost equally well in estimating the treatment effect, even when the frailty distribution is misspecified.



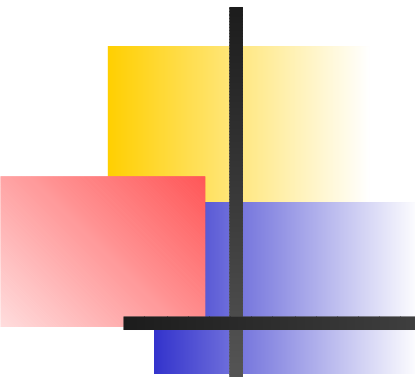
Simulation Study (cont.)

- All the models considered performed almost equally well in estimating the treatment effect, even when the frailty distribution is misspecified.
- This suggests that the frailty distribution may have little impact on the estimation of treatment effect.



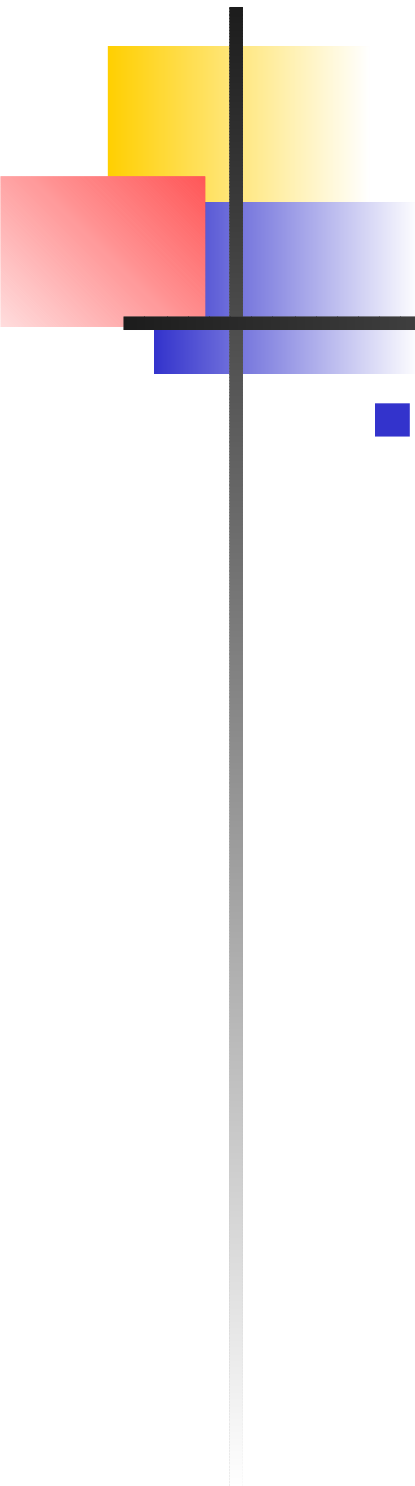
Simulation Study (cont.)

- All the models considered performed almost equally well in estimating the treatment effect, even when the frailty distribution is misspecified.
- This suggests that the frailty distribution may have little impact on the estimation of treatment effect.
- Using the piecewise exponential for baseline hazard results in slightly larger MSEs in the estimates of the treatment effect than under the exponential (perhaps because of more parameters being estimated).



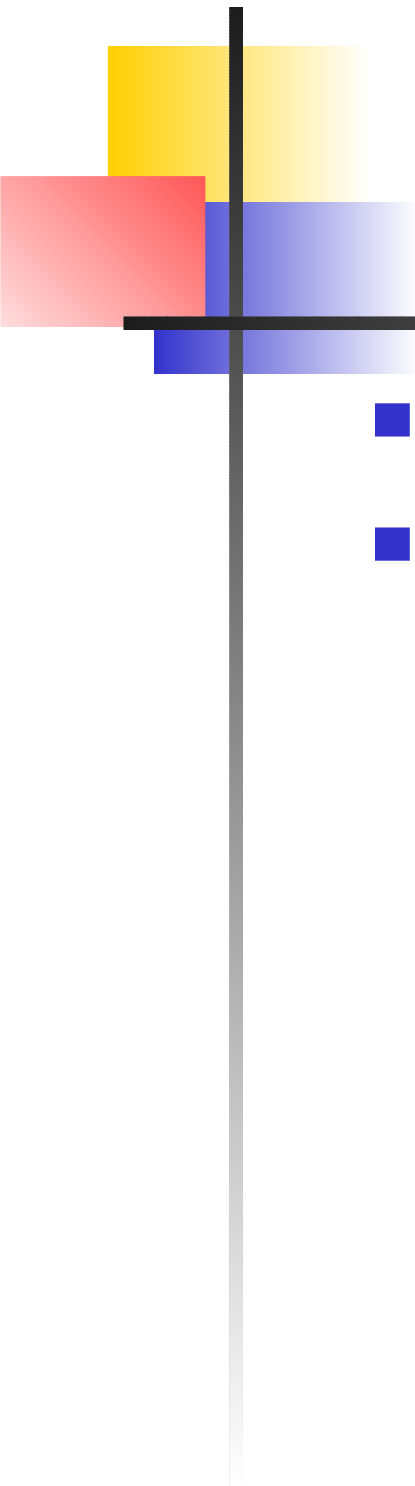
Simulation Study (cont.)

- All the models considered performed almost equally well in estimating the treatment effect, even when the frailty distribution is misspecified.
- This suggests that the frailty distribution may have little impact on the estimation of treatment effect.
- Using the piecewise exponential for baseline hazard results in slightly larger MSEs in the estimates of the treatment effect than under the exponential (perhaps because of more parameters being estimated).
- Differences in the MSEs become negligible when the sample size increases, and the piecewise exponential baseline hazard will be more robust, of course!



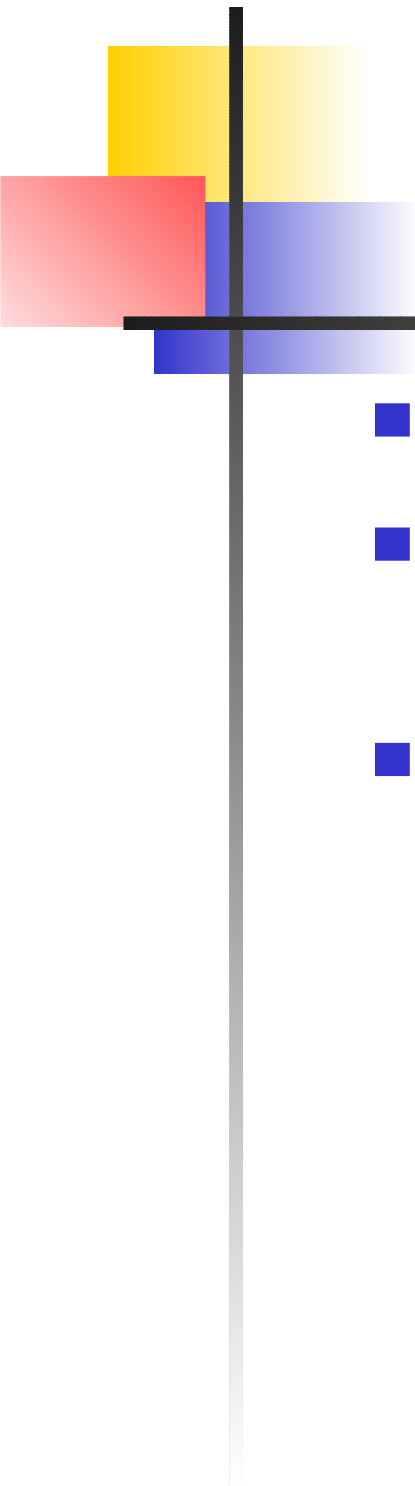
Simulation Study (cont.)

- For frailty variance, noticeable differences are present.



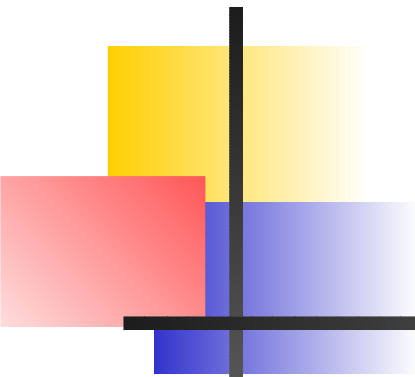
Simulation Study (cont.)

- For frailty variance, noticeable differences are present.
- Gamma frailty had large bias in variance estimate when the true frailty distribution was lognormal.



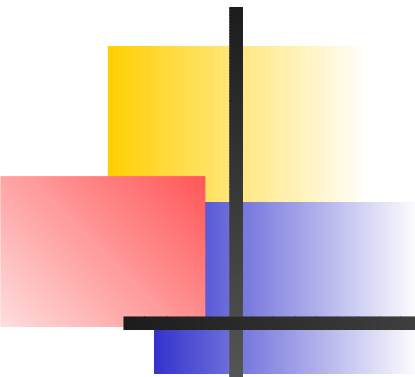
Simulation Study (cont.)

- For frailty variance, noticeable differences are present.
- Gamma frailty had large bias in variance estimate when the true frailty distribution was lognormal.
- Lognormal frailty had large MSE and bias in variance estimate when the true frailty distribution was gamma.



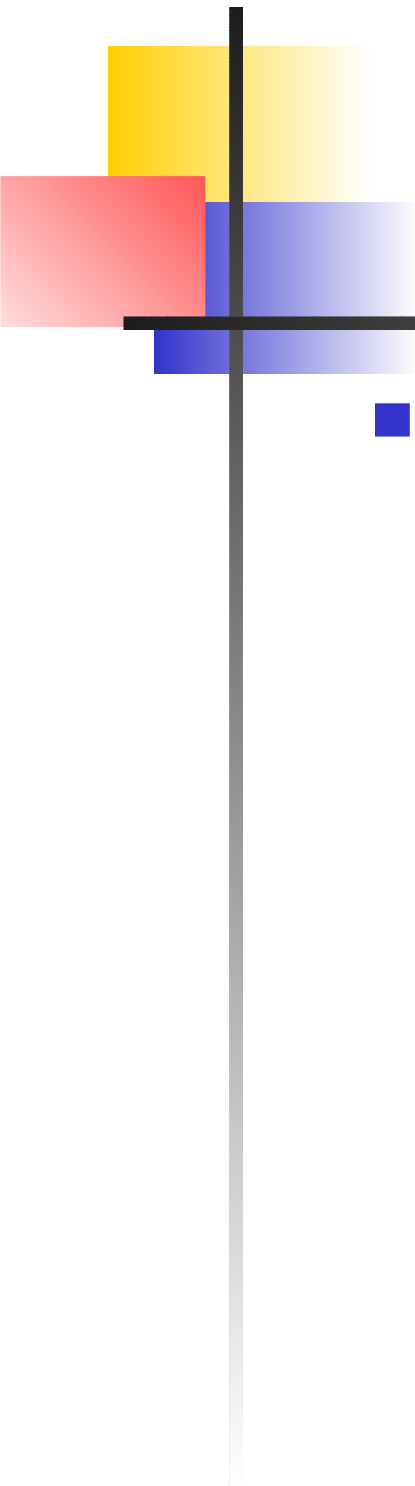
Simulation Study (cont.)

- For frailty variance, noticeable differences are present.
- Gamma frailty had large bias in variance estimate when the true frailty distribution was lognormal.
- Lognormal frailty had large MSE and bias in variance estimate when the true frailty distribution was gamma.
- When the fitted frailty is same as the true one, variance estimates had the smallest bias and MSE, and different baseline distributions appear to have little impact.



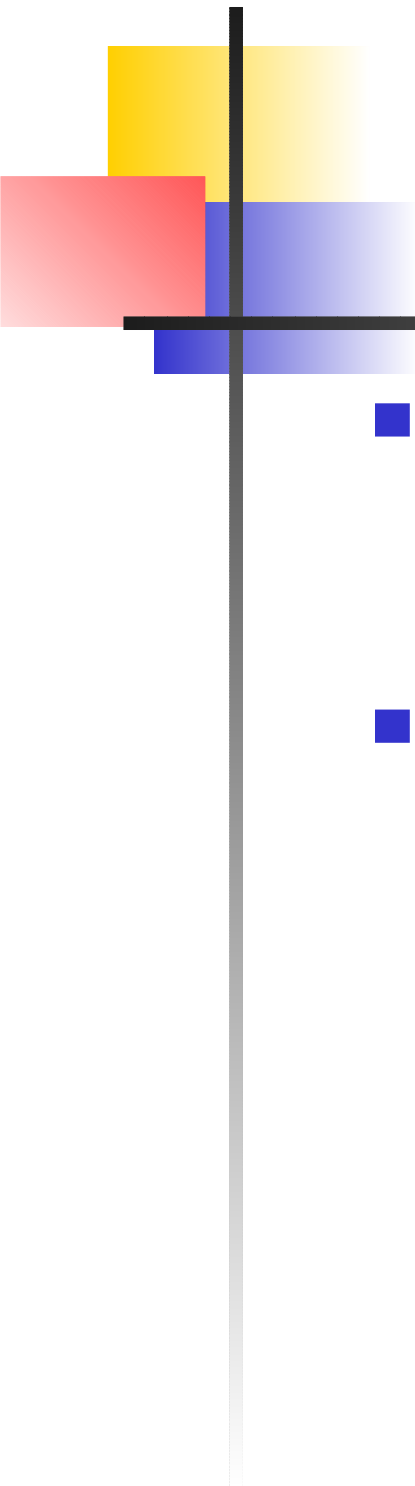
Simulation Study (cont.)

- For frailty variance, noticeable differences are present.
- Gamma frailty had large bias in variance estimate when the true frailty distribution was lognormal.
- Lognormal frailty had large MSE and bias in variance estimate when the true frailty distribution was gamma.
- When the fitted frailty is same as the true one, variance estimates had the smallest bias and MSE, and different baseline distributions appear to have little impact.
- GG frailty model had small MSE and bias in the variance estimation, and comparable to those from the true frailty models.



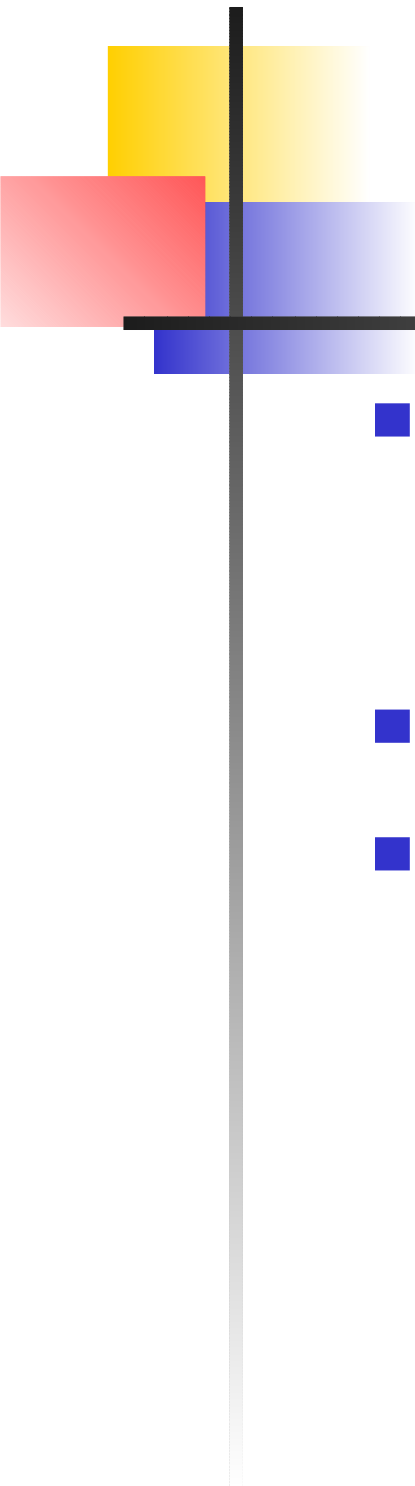
Goodness-of-Fit

- It is of practical interest to explore the use of GG frailty model to select a parsimonious frailty distribution that provides an adequate fit to the data.



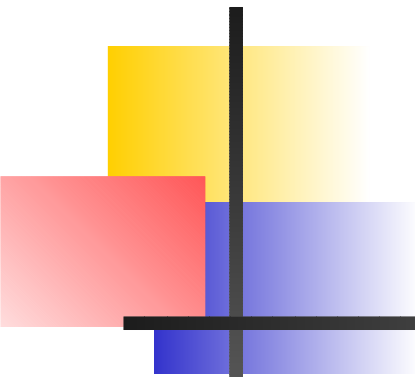
Goodness-of-Fit

- It is of practical interest to explore the use of GG frailty model to select a parsimonious frailty distribution that provides an adequate fit to the data.
- We can use the likelihood ratio statistic for this task.



Goodness-of-Fit

- It is of practical interest to explore the use of GG frailty model to select a parsimonious frailty distribution that provides an adequate fit to the data.
- We can use the likelihood ratio statistic for this task.
- The likelihood ratio (LR) statistic can be used to test



Goodness-of-Fit

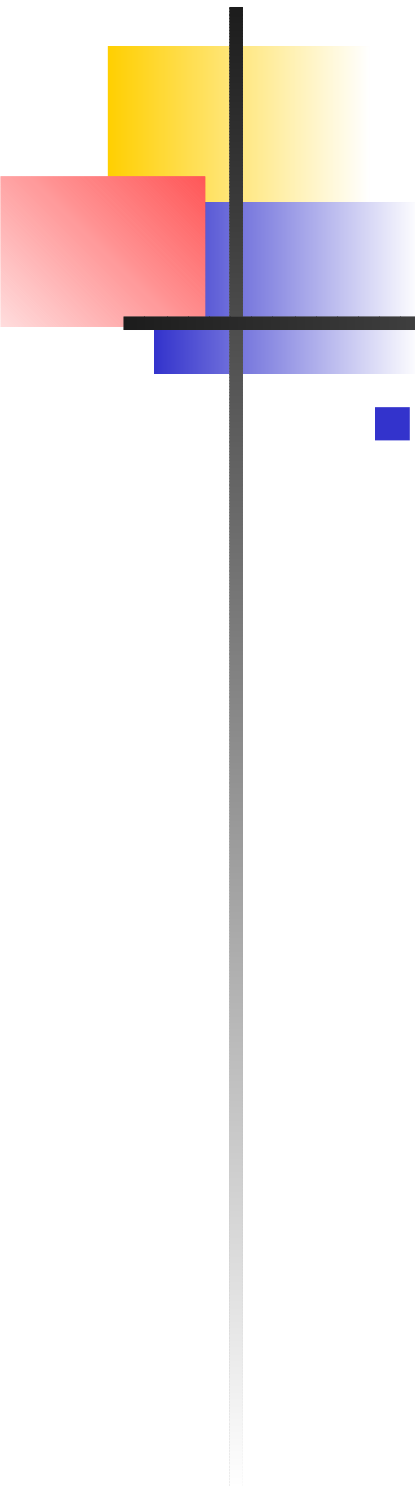
- It is of practical interest to explore the use of GG frailty model to select a parsimonious frailty distribution that provides an adequate fit to the data.
 - We can use the likelihood ratio statistic for this task.
 - The likelihood ratio (LR) statistic can be used to test
 - H_0 : Correlation in survival data can be described by one of gamma, Weibull and lognormal frailty
- versus*
- H_1 : Correlation can be described by a member of GG other than the frailty specified under H_0 .

Goodness-of-Fit (cont.)

Observed levels (in bold) and powers of the LR test at 10% level of significance^a.

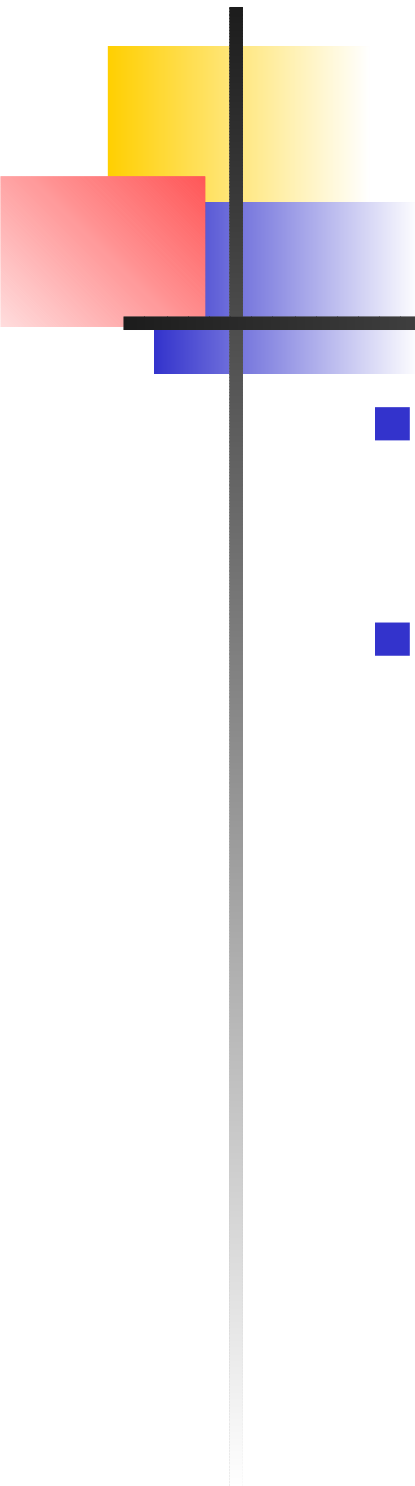
Fitted frailty	<i>True frailty and baseline distributions</i>			
	G & Exp	G & LN	LN & Exp	LN & LN
Gamma	0.114	0.141	0.342	0.122
Weibull	0.143	0.077	0.581	0.357
Lognormal	0.458	0.668	0.088	0.132
Gamma	0.102	0.100	0.341	0.208
Weibull	0.149	0.103	0.586	0.490
Lognormal	0.455	0.522	0.092	0.090

^aFirst block is for *fitting with exponential baseline distribution* and second block is for *fitting with piecewise exponential baseline distribution*.



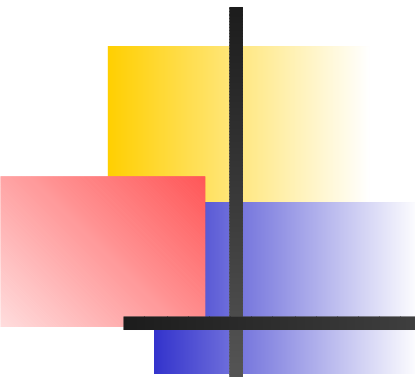
Goodness-of-Fit (cont.)

- As the observed levels are close to 10%, χ^2 distn. provides a good approximation to the null distribution.



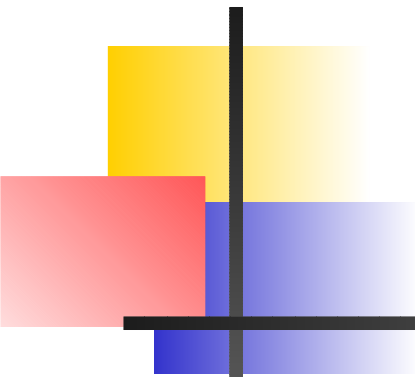
Goodness-of-Fit (cont.)

- As the observed levels are close to 10%, χ^2 distn. provides a good approximation to the null distribution.
- When Gamma is the true frailty, test has good power to reject the Lognormal (with a largest power of 67%).



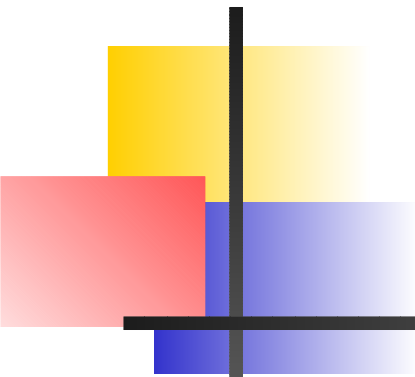
Goodness-of-Fit (cont.)

- As the observed levels are close to 10%, χ^2 distn. provides a good approximation to the null distribution.
- When Gamma is the true frailty, test has good power to reject the Lognormal (with a largest power of 67%).
- When Gamma is the true frailty, test has low power to reject the Weibull. These two models provide similar fits and do not differ in describing correlation in data.



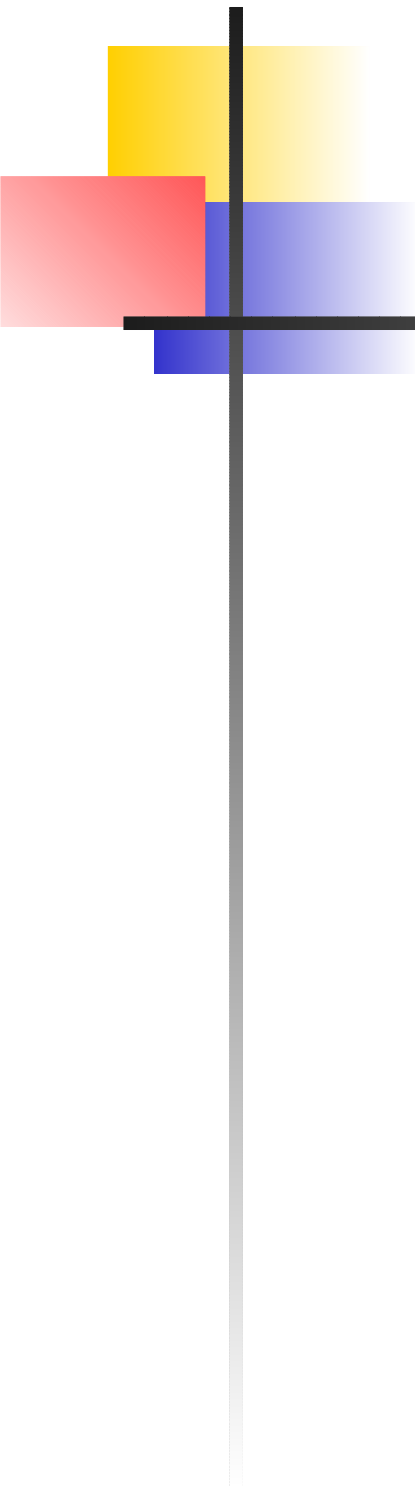
Goodness-of-Fit (cont.)

- As the observed levels are close to 10%, χ^2 distn. provides a good approximation to the null distribution.
- When Gamma is the true frailty, test has good power to reject the Lognormal (with a largest power of 67%).
- When Gamma is the true frailty, test has low power to reject the Weibull. These two models provide similar fits and do not differ in describing correlation in data.
- When Lognormal is the true frailty, test has good power to reject the Weibull (with a largest power of 59%).



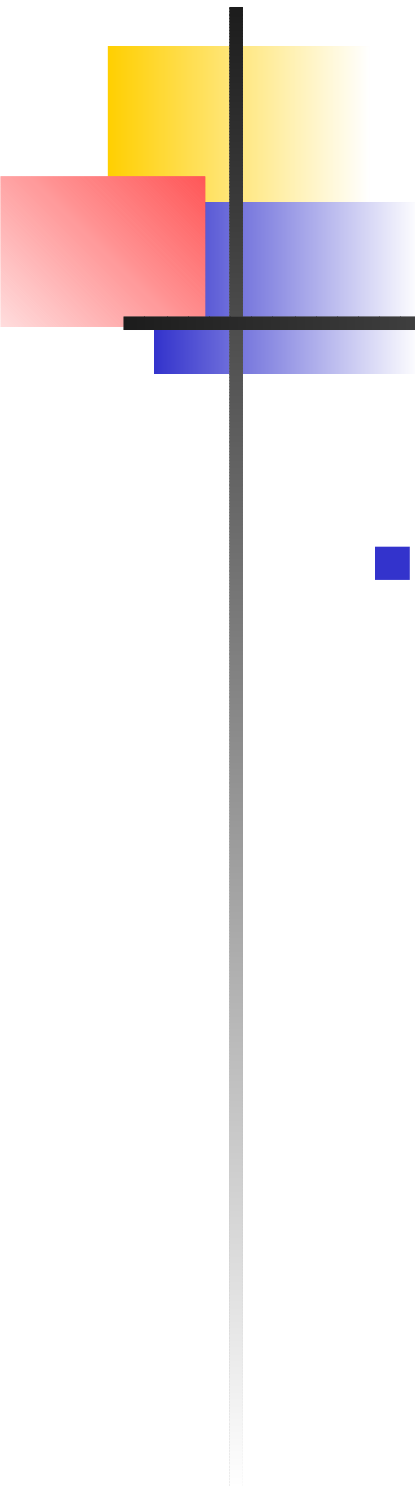
Goodness-of-Fit (cont.)

- As the observed levels are close to 10%, χ^2 distn. provides a good approximation to the null distribution.
- When Gamma is the true frailty, test has good power to reject the Lognormal (with a largest power of 67%).
- When Gamma is the true frailty, test has low power to reject the Weibull. These two models provide similar fits and do not differ in describing correlation in data.
- When Lognormal is the true frailty, test has good power to reject the Weibull (with a largest power of 59%).
- When Lognormal is the true frailty, test has moderate power to reject the Gamma (of about 34%).



Illustrative Examples

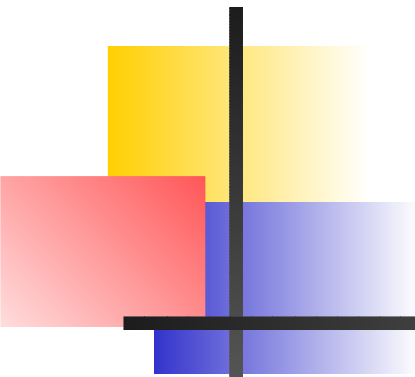
Example 1



Illustrative Examples

Example 1

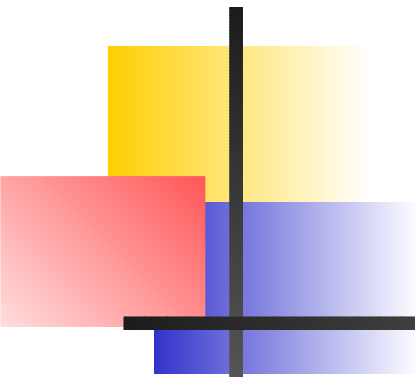
- Danahy *et al.* (1977) studied oral administration of isosorbide dinitrate on 21 heart disease patients.



Illustrative Examples

Example 1

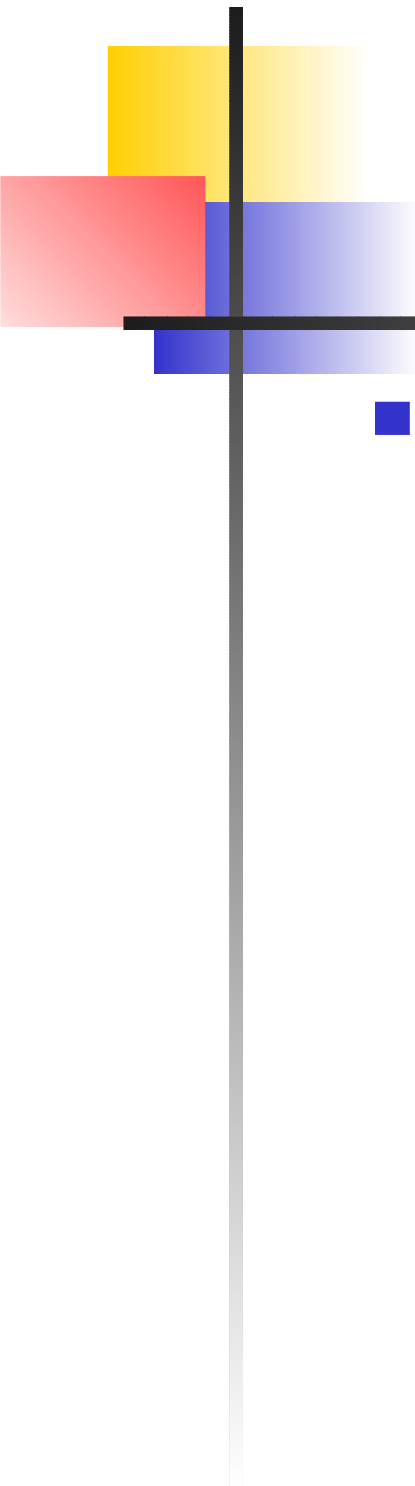
- Danahy *et al.* (1977) studied oral administration of isosorbide dinitrate on 21 heart disease patients.
- Patients were treated with sublingual nitroglycerin (SLT) and with sublingual placebo (SLP), and then with oral isosorbide dinitrate (OI), and oral placebo (OP).



Illustrative Examples

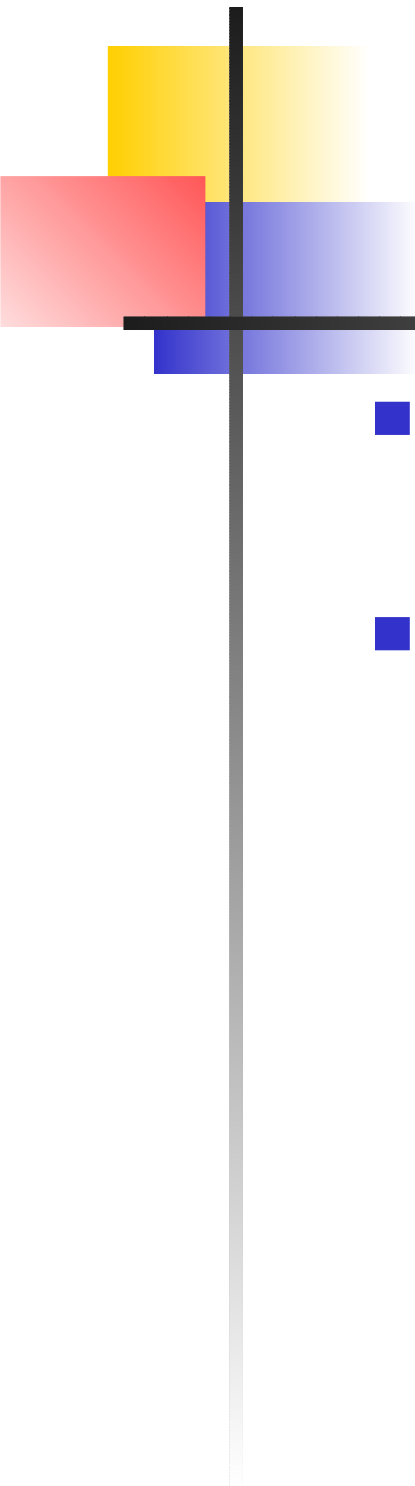
Example 1

- Danahy *et al.* (1977) studied oral administration of isosorbide dinitrate on 21 heart disease patients.
- Patients were treated with sublingual nitroglycerin (SLT) and with sublingual placebo (SLP), and then with oral isosorbide dinitrate (OI), and oral placebo (OP).
- 10 tests of exercise bike pedalling were administered on the patients, one immediately before OI or OP (denoted by OI0 and OP0) and three after OI or OP, at 1, 3 and 5 h (denoted by OI1, OI3, OI5 and OP1, OP3, OP5).



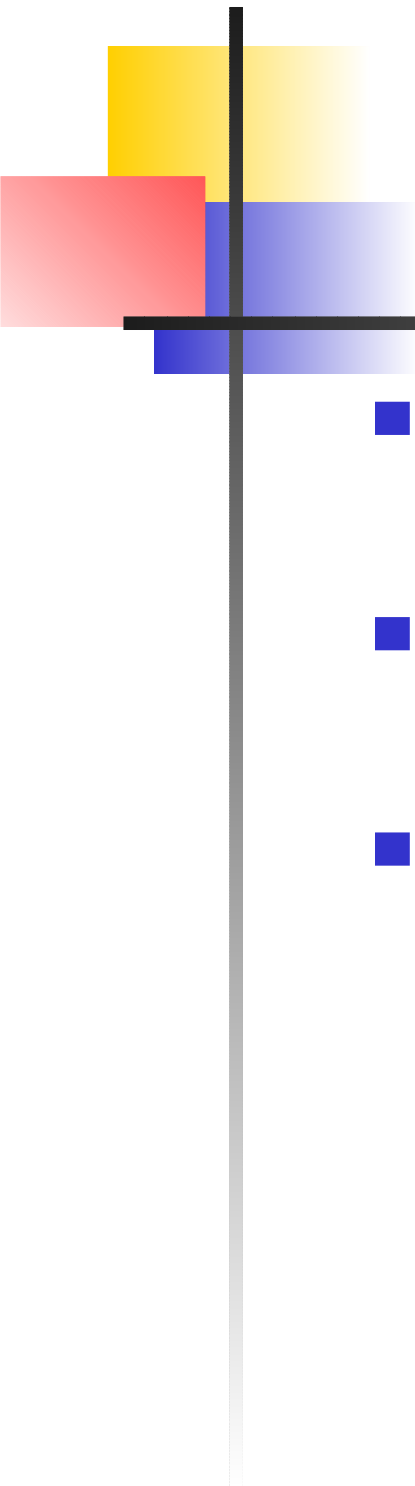
Example 1 (cont.)

- Tests lasted until the patients had angina pectoris and the times to angina pectoris were recorded.



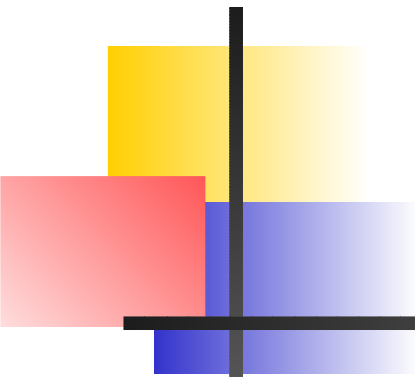
Example 1 (cont.)

- Tests lasted until the patients had angina pectoris and the times to angina pectoris were recorded.
- Some of these times were censored because patients became too exhausted to continue.



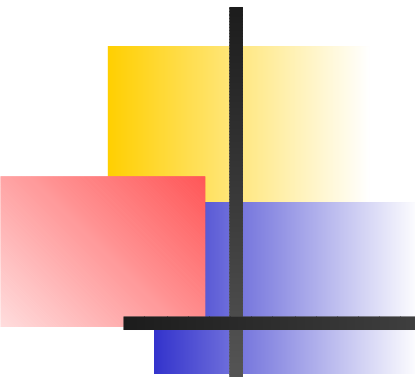
Example 1 (cont.)

- Tests lasted until the patients had angina pectoris and the times to angina pectoris were recorded.
- Some of these times were censored because patients became too exhausted to continue.
- In this case, we have $n = 21$ and $m_i = 10$.



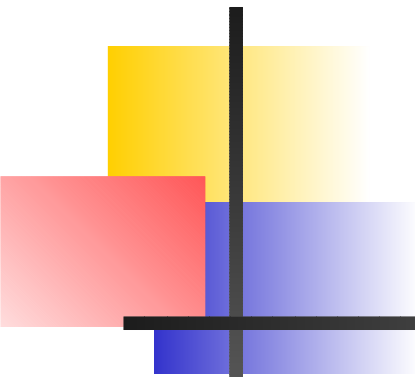
Example 1 (cont.)

- Tests lasted until the patients had angina pectoris and the times to angina pectoris were recorded.
- Some of these times were censored because patients became too exhausted to continue.
- In this case, we have $n = 21$ and $m_i = 10$.
- We analyzed the data with various frailty models by using the baseline distribution to be



Example 1 (cont.)

- Tests lasted until the patients had angina pectoris and the times to angina pectoris were recorded.
- Some of these times were censored because patients became too exhausted to continue.
- In this case, we have $n = 21$ and $m_i = 10$.
- We analyzed the data with various frailty models by using the baseline distribution to be
 - the Weibull distribution
 - the piecewise exponential distribution.



Example 1 (cont.)

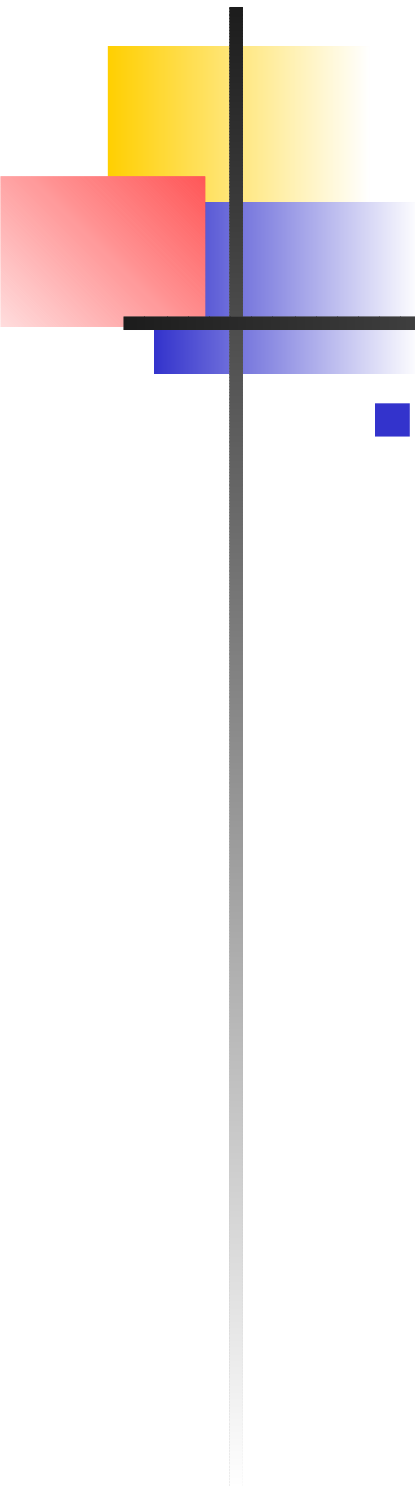
Estimates of β (standard error) from fitted frailty models with the Weibull baseline distribution.

Covariate	Gamma	Weibull	Lognormal	GG
SLT	-1.51(0.34)	-1.53(0.34)	-1.53(0.34)	-1.51(0.34)
OP0	0.70(0.33)	0.71(0.33)	0.72(0.33)	0.67(0.33)
OP1	0.14(0.32)	0.13(0.32)	0.14(0.32)	0.11(0.32)
OP3	0.28(0.33)	0.28(0.33)	0.29(0.33)	0.27(0.32)
OP5	0.64(0.33)	0.65(0.33)	0.66(0.33)	0.66(0.32)
OI0	0.15(0.33)	0.15(0.33)	0.15(0.33)	0.19(0.32)
OI1	-2.65(0.41)	-2.66(0.41)	-2.65(0.41)	-2.54(0.39)
OI3	-1.37(0.36)	-1.38(0.36)	-1.38(0.36)	-1.31(0.35)
OI5	-0.36(0.35)	-0.36(0.35)	-0.35(0.35)	-0.38(0.33)
Frailty Var.	2.49	5.79	49.47	232.27
Max log-likel.	-1125.25	-1123.83	-1122.19	-1120.92

Example 1 (cont.)

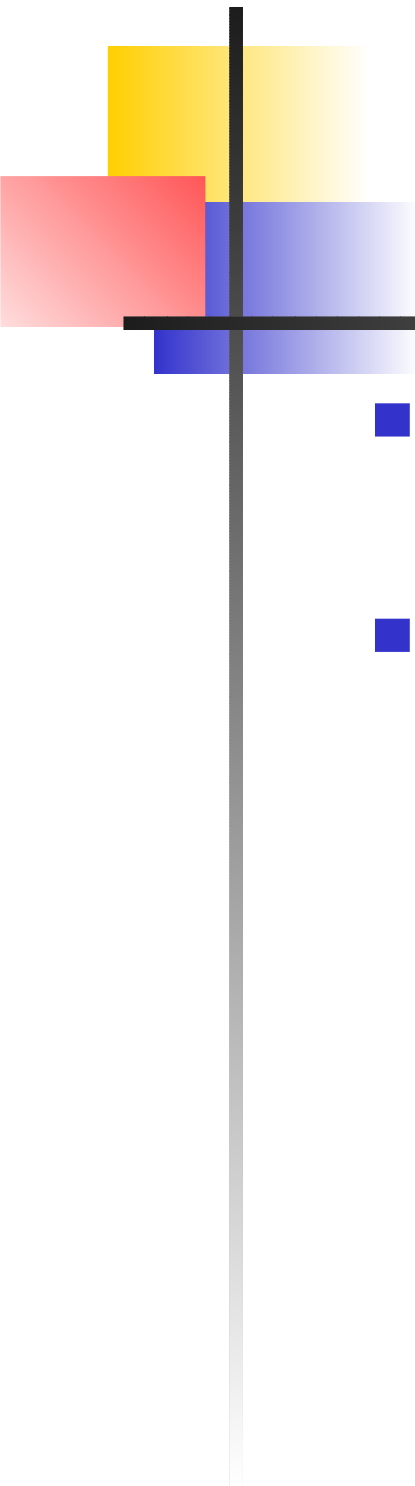
Estimates of β (standard error) from fitted frailty models with the piecewise exponential baseline distribution.

Covariate	Gamma	Weibull	Lognormal	GG
SLT	-1.38(0.34)	-1.36(0.34)	-1.37(0.34)	-1.37(0.34)
OP0	0.59(0.33)	0.58(0.33)	0.58(0.33)	0.56(0.33)
OP1	0.01(0.32)	0.00(0.32)	0.00(0.32)	-0.06(0.31)
OP3	0.22(0.33)	0.20(0.33)	0.20(0.33)	0.17(0.33)
OP5	0.52(0.32)	0.51(0.32)	0.51(0.32)	0.48(0.32)
OI0	0.06(0.32)	0.07(0.33)	0.06(0.32)	0.09(0.32)
OI1	-2.24(0.39)	-2.19(0.39)	-2.17(0.39)	-2.16(0.38)
OI3	-1.32(0.35)	-1.29(0.35)	-1.28(0.35)	-1.28(0.35)
OI5	-0.41(0.34)	-0.40(0.34)	-0.40(0.34)	-0.41(0.41)
Frailty Var.	2.34	4.17	21.46	56.18
Max log-likel.	-1116.67	-1115.34	-1113.11	-1111.39



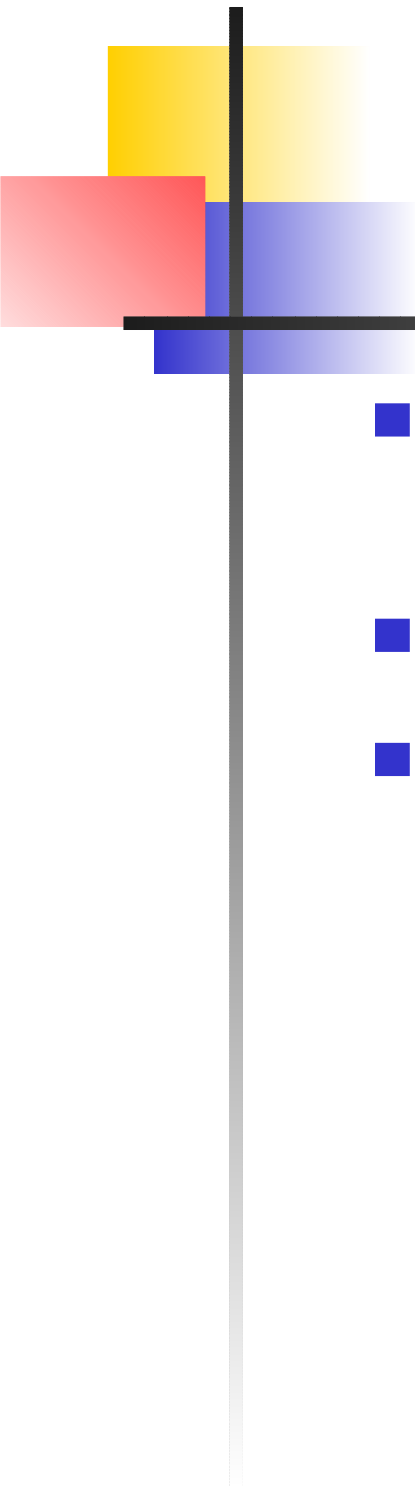
Example 1 (cont.)

- In both cases, the GG frailty model has the largest maximum log-likelihood, and it provides the best fit.



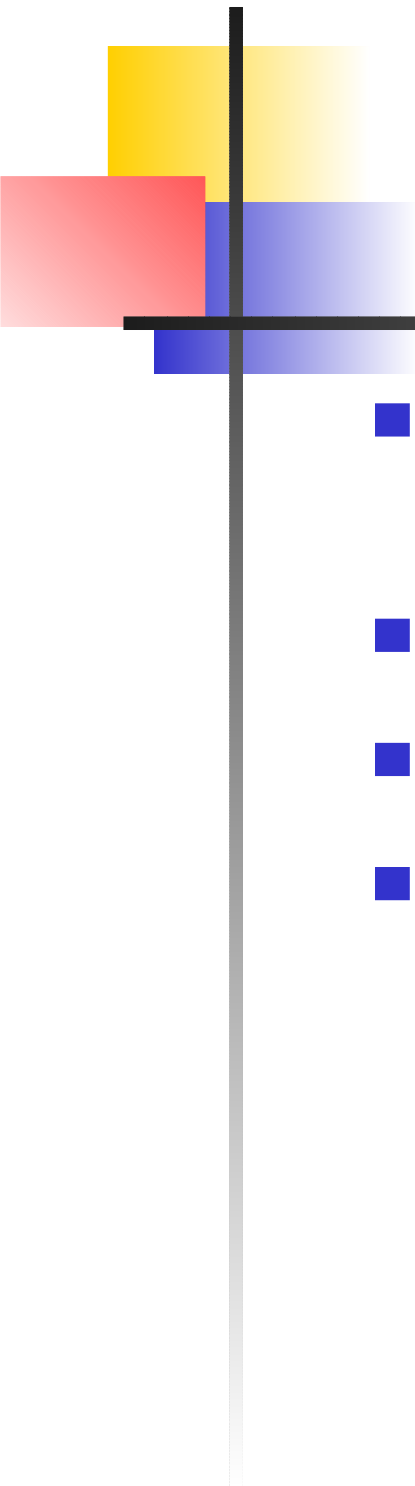
Example 1 (cont.)

- In both cases, the GG frailty model has the largest maximum log-likelihood, and it provides the best fit.
- Gamma model is not acceptable (p -value=0.003).



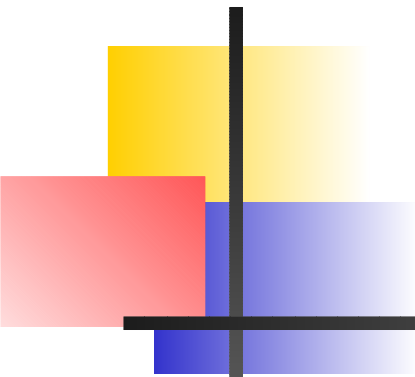
Example 1 (cont.)

- In both cases, the GG frailty model has the largest maximum log-likelihood, and it provides the best fit.
- Gamma model is not acceptable (p -value=0.003).
- Weibull model is not acceptable (p -value=0.02).



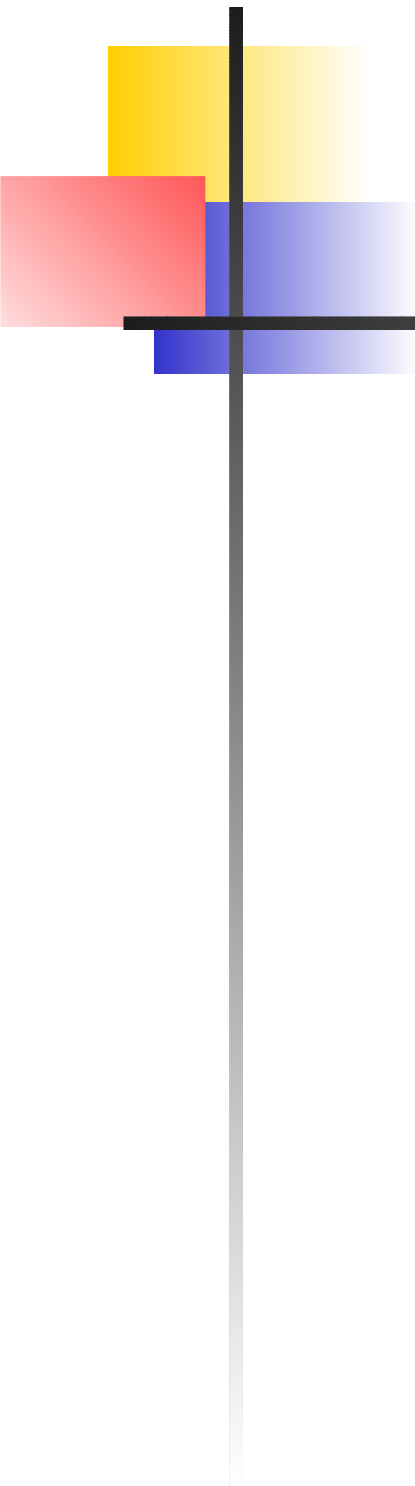
Example 1 (cont.)

- In both cases, the GG frailty model has the largest maximum log-likelihood, and it provides the best fit.
- Gamma model is not acceptable (p -value=0.003).
- Weibull model is not acceptable (p -value=0.02).
- LN model is marginally acceptable (p -value=0.11).



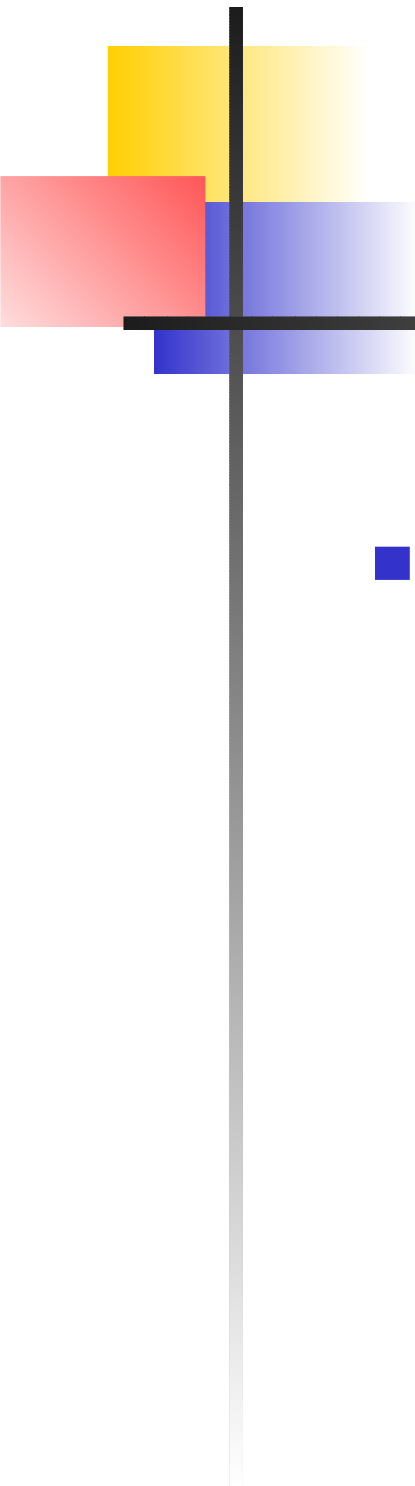
Example 1 (cont.)

- In both cases, the GG frailty model has the largest maximum log-likelihood, and it provides the best fit.
- Gamma model is not acceptable (p -value=0.003).
- Weibull model is not acceptable (p -value=0.02).
- LN model is marginally acceptable (p -value=0.11).
- Estimated frailty variances are increasing from Gamma to GG models. Since the frailty variance signals the strength of correlation among observations within clusters, Gamma and Weibull models seriously underestimate the strength of the correlation.



Illustrative Examples

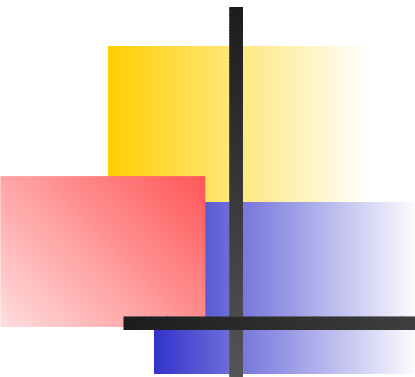
Example 2



Illustrative Examples

Example 2

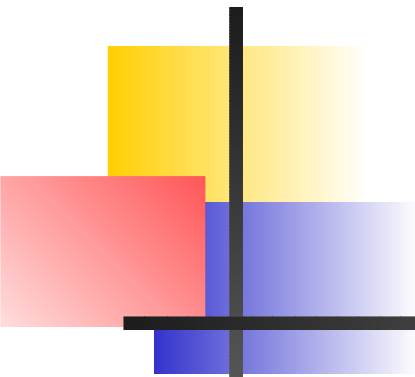
- Data of McGilchrist and Aisbett (1991) contains recurrence times to infection at point of insertion of the catheter for 38 kidney patients using dialysis machine.



Illustrative Examples

Example 2

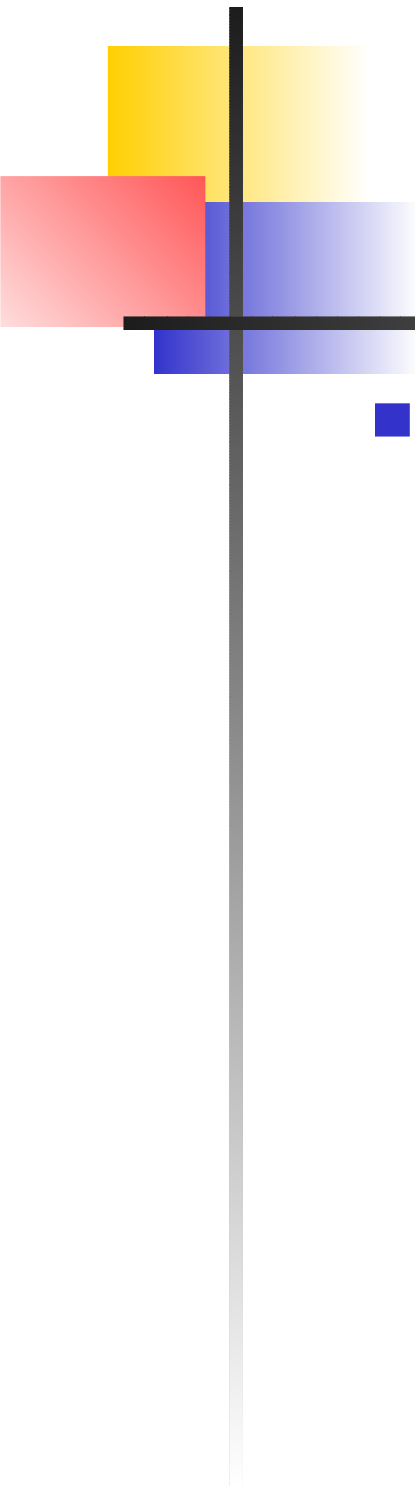
- Data of McGilchrist and Aisbett (1991) contains recurrence times to infection at point of insertion of the catheter for 38 kidney patients using dialysis machine.
- Risk variables are age, sex (1 = male, 2 = female), and disease type (0 = Glomerulo Nephrit, 1 = Acute Nephritis, 2 = Polycystic Kidney Disease, 3 = Other).



Illustrative Examples

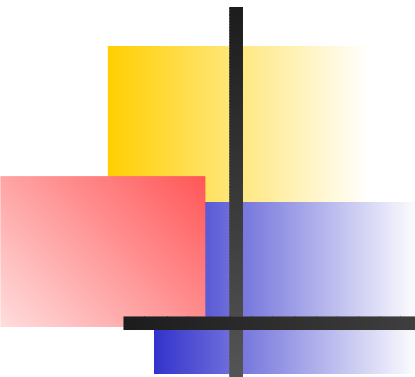
Example 2

- Data of McGilchrist and Aisbett (1991) contains recurrence times to infection at point of insertion of the catheter for 38 kidney patients using dialysis machine.
- Risk variables are age, sex (1 = male, 2 = female), and disease type (0 = Glomerulo Nephrit, 1 = Acute Nephritis, 2 = Polycystic Kidney Disease, 3 = Other).
- After examining the Kaplan-Meier survival curves of the recurrence times and their quartiles, we chose $k = 5$ and $t^{(1)} = 2$, $t^{(2)} = 15$, $t^{(3)} = 40$, $t^{(4)} = 150$, and $t^{(5)} = 350$ as cut points in the piecewise exponential distribution.



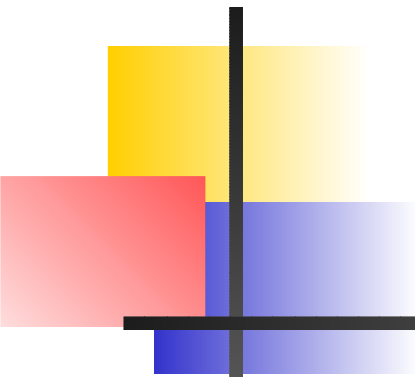
Example 2 (cont.)

- The largest maximized log-likelihood from the GG frailty model indicates that it provides the best fit among the four parametric frailty models.



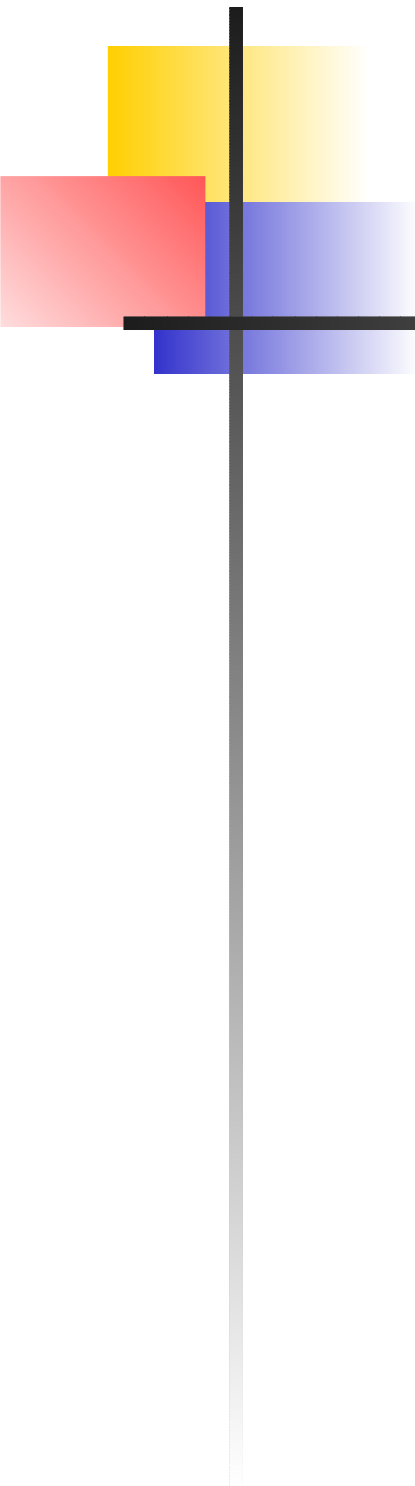
Example 2 (cont.)

- The largest maximized log-likelihood from the GG frailty model indicates that it provides the best fit among the four parametric frailty models.
- The maximized log-likelihood and estimated regression coefficients of the lognormal frailty model, which has one less parameter than the GG frailty model, are very close to those of the GG frailty model.



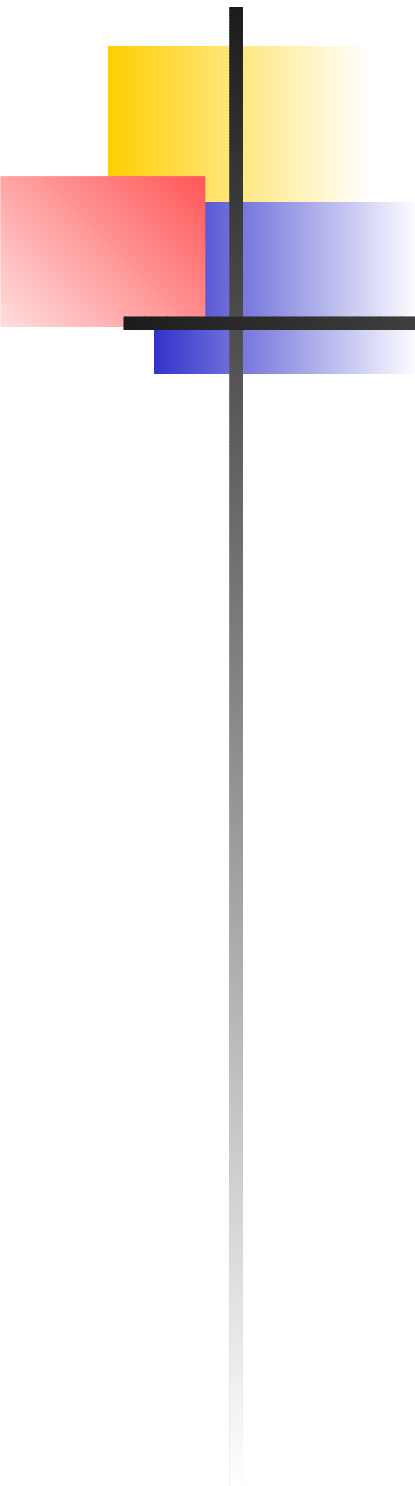
Example 2 (cont.)

- The largest maximized log-likelihood from the GG frailty model indicates that it provides the best fit among the four parametric frailty models.
- The maximized log-likelihood and estimated regression coefficients of the lognormal frailty model, which has one less parameter than the GG frailty model, are very close to those of the GG frailty model.
- Risk variables are similar for the lognormal and GG frailty models except that in the latter the disease type PKD also has a marginal effect, in addition to the significant effect of the sex variable.



Bibliography

- Danahy, D. T., Burwell, D. T., Aronow, W. S. and Prakash, R. (1977). Sustained hemodynamic and antianginal effect of high dose oral isosorbide dinitrate, *Circulation*, **55**, 381–387.
- Hougaard, P. (1986a). A class of multivariate failure time distributions, *Biometrika*, **73**, 671–678.
- Hougaard, P. (1986b). Survival models for heterogenous populations derived from stable distributions, *Biometrika*, **73**, 387–396.
- Hougaard, P. (1995). Frailty models for survival data, *Lifetime Data Analysis*, **1**, 255-273.
- Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on EM algorithm, *Biometrics*, **48**, 795–806.
- McGilchrist, C. A. (1993). REML estimation for survival models with frailty, *Biometrics*, **49**, 221–225.



Bibliography (cont.)

- McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis, *Biometrics*, **47**, 461–466.
- Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sorensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models, *Scandinavian Journal of Statistics*, **19**, 25–43.
- Vaupel, J. W., Manton, K. G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, **16**, 439–454.
- Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of permanent employment in Japan, *Journal of the American Statistical Association*, **87**, 284–292.