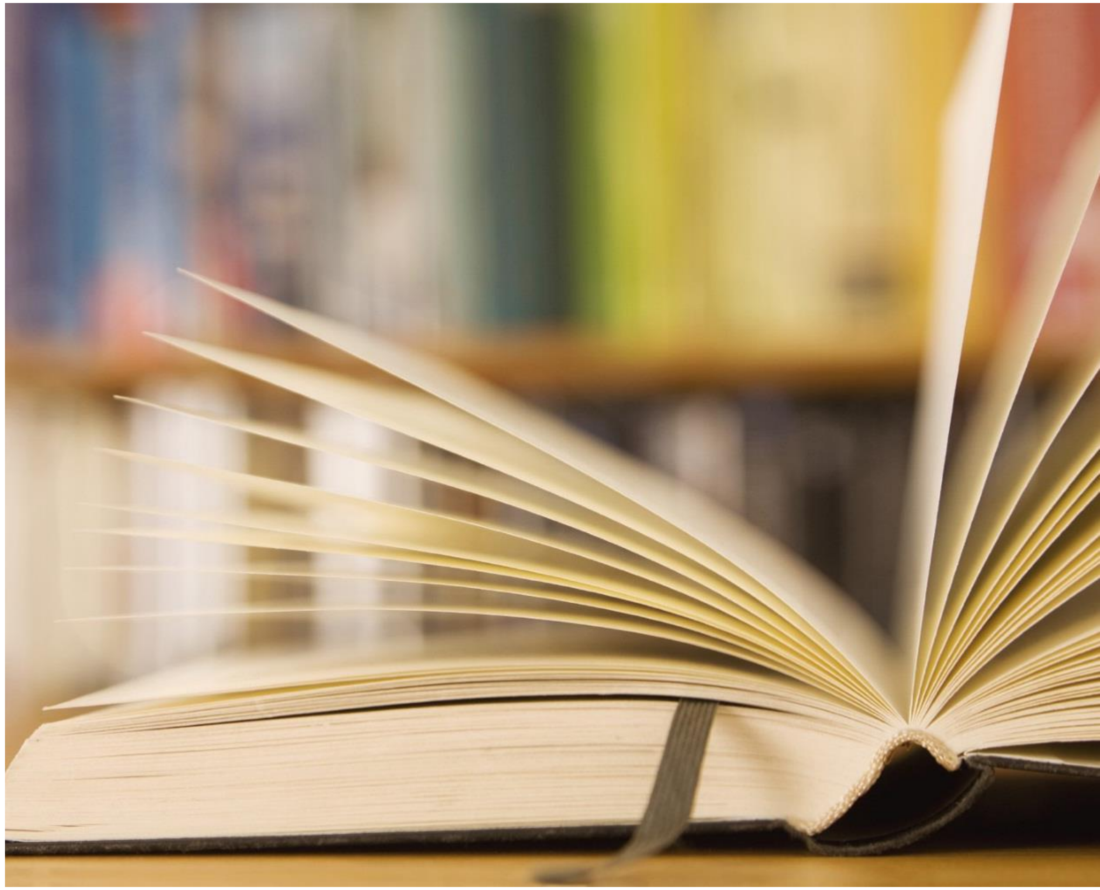
PENALIZED LIKELIHOOD PARAMETER ESTIMATION FOR ADDITIVE HAZARD MODELS WITH INTERVAL CENSORED DATA

Kasun Rathnayake ; A/Prof Jun Ma

Department of Statistics
Faculty of Science and Engineering
Macquarie University

MACQUARIE
University



Outline

- Introduction
  - Survival models
  - Types of censoring
  - Research problem
- Methodology
- Examples
- Conclusion/ Discussion
- References

Introduction – Survival models

Proportional Hazards model

$$h(t) = h_0(t).\exp(X^T\beta)$$

- Proposed by Cox (1972)
- Covariates act multiplicatively on some unknown baseline hazard rate

Additive Hazards model

$$h(t) = h_0(t) + (X(t))^T\beta(t)$$

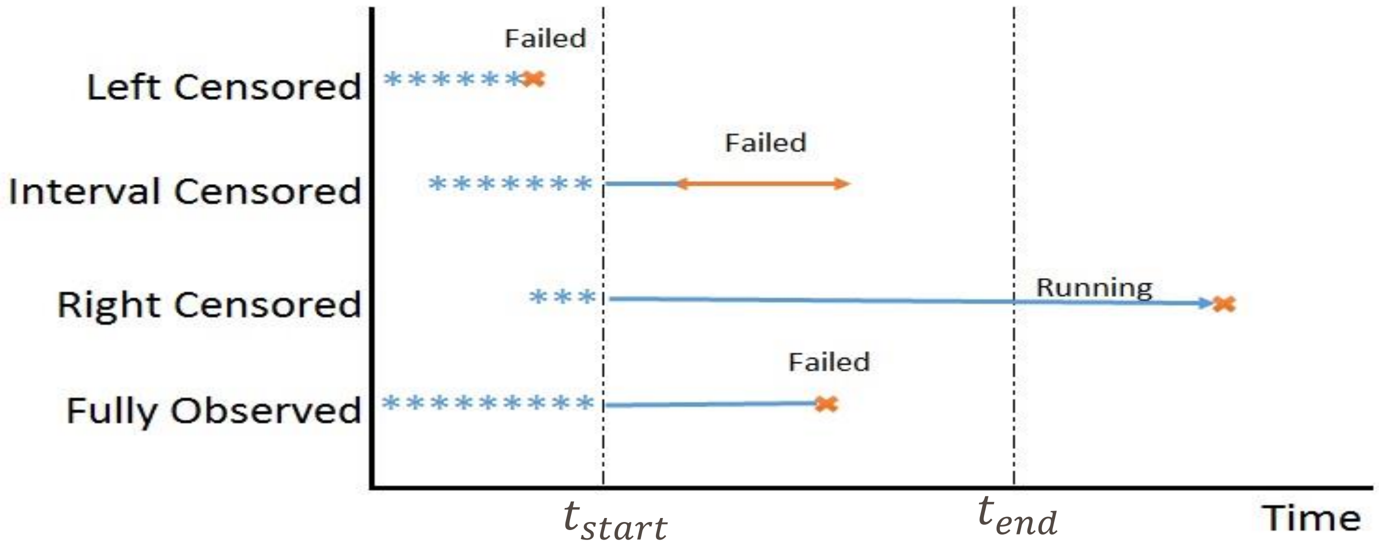
- Proposed by Aalen (1989)
- Covariates act additively on some unknown baseline hazard rate
- $\beta(t)$ is depend on time t

Additive Hazards model – Special case

$$h(t) = h_0(t) + (X(t))^T\beta$$

- Proposed by Lin & Ying (1994)
- Coefficient β is constant
- Lin and Ying considered this model with the baseline $h_0(t)$ be any nonnegative function

Introduction – Types of censoring



T_i- failure time
C_i- censoring time

- Left censoring; $\boldsymbol{t_i = \max(T_i, C_i)}$ - event of interest has already occured before the study starts
- Interval censoring; $\boldsymbol{t_i} \in (\boldsymbol{L_i, R_i})$ - exact event time is unknown, but interval bounding is known
- Right censoring; $\boldsymbol{t_i = \min(T_i, C_i)}$ - study ends before the event has occured
- Fully observed; $\boldsymbol{t_i = T_i}$ - events occured within study period

Introduction – Research problem

- Interested on parameter estimation of the additive hazard model:

$$h_i(t, X_i) = h_0(t) + X_i\beta$$

- For this model the two non-negativity constraints as follows :
  - Constraint 1 : $h_0(t) \geq 0$
  - Constraint 2 : $h_0(t) + X_i\beta \geq 0$

Introduction – Research problem

- Main focus: Estimating the regression coefficients, $\boldsymbol{\beta}$ and the baseline hazard ,$\boldsymbol{h_0(t)}$ **alternately** by considering the two constraints
  - Aalen (1980)
    - Used least square approach to estimate cumulative estimates
    - Without considering constraints
  - Ghosh (2001) & Zeng et al.(2006)
    - Used maximum likelihood approach to estimate β and $H_0(t)$/ $S_0(t)$ $\{= \exp(H_0(t)\}$
    - Imposed constraints on $H_0(t)$ and $H(t)$
  - Farrington (1996)
    - Used GLM approach to estimate β and $h_0(t)$
    - Non-negativity of $h_0(t)$ cannot be guaranteed

MACQUARIE University

Methodology

- Used Maximum Penalized Likelihood **(MPL)** approach
  - To estimate $h_0(t)$ by considering the full likelihood
  - To smooth baseline hazard, a penalty function, $\boldsymbol{J}(\boldsymbol{h_0})$ can be added on $\boldsymbol{h_0}(\boldsymbol{t})$
- Then, maximum penalized likelihood objective function with respect to $\boldsymbol{h_0}(\boldsymbol{t})$ and $\boldsymbol{\beta}$,

$$\left(\widehat{h_0(t)}\,,\hat{\beta}\right) = argmax\,\{\,\boldsymbol{\Phi}\,(h_0(t),\beta) = l\,(h_0(t),\beta) - \lambda.\boldsymbol{J}\big(h_0(t)\big)\}$$

here $\boldsymbol{\lambda}$ is the smoothing parameter

Methodology continued…

- General form of Additive hazard model for observation $\boldsymbol{i}$:

$$\boldsymbol{h_i(t_i, X_i) = h_0(t_i) + X_i\beta}$$

- Constraint 1 : $\boldsymbol{h_0(t_i) \geq 0}$
  - $h_0(t_i) = \sum_{u=1}^{m} \theta_u \psi_u(t_i)$
  - Select a basis function such that $\psi_u(t_i) \geq 0$ for all u
  - $i.e$ need to restrict $\boldsymbol{\theta_u \geq 0}$ to enforce $h_0(t_i) \geq 0$
- Constraint 2 : $\boldsymbol{h_0(t_i) + X_i\beta \geq 0}$
  - $X_i\beta \geq -h_0(t_i)$
  - Possible to re-write above constraint in a simplified form as follows

$$\boldsymbol{X_i\beta \geq -\theta_u \ ; \ t_i \in B_u}$$

Methodology continued…

- Used Augmented Lagrangian (AL) method to treat constraints
- Let $\eta_i = X_i\beta$ to transfer the part of the constraints to $\boldsymbol{\eta_i}$
- Augmented Lagrangian with respect to $\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}$ and $\boldsymbol{\gamma}$,

$$\left(\hat{\theta}\ , \widehat{\beta},\ \hat{\eta}, \hat{\gamma}\right) = argmax_{\theta,\beta,\eta,\gamma}\ \{\mathcal{L}_\alpha(\theta, \beta, \eta, \gamma)\}$$

where $\mathcal{L}_\alpha(\theta, \beta, \eta, \gamma) = \{l\ (\theta, \beta) - \lambda.J(\theta)\}\ -\ \sum_{i=1}^{n} \gamma_i (X_i^T \beta\ - \eta_i)\ -\ \frac{\alpha}{2} \sum_{i=1}^{n} (X_i^T \beta\ - \eta_i)^2$

- These four parameters are updated alternately in each iteration.

Methodology – Alternately updates

- Iteration $(\boldsymbol{k}+\mathbf{1})$ consists of following steps:
- **STEP 1**:With $\boldsymbol{\eta}^{(\boldsymbol{k})}, \boldsymbol{\beta}^{(\boldsymbol{k})}$and $\boldsymbol{\gamma}^{(\boldsymbol{k})}$ obtained $\boldsymbol{\theta}^{(\boldsymbol{k}+\mathbf{1})}$;
  - by running one iteration of the Multiplicative Iterative (MI) algorithm (Ma et al. (2014)
  - followed by a line search
  - this guarantees that each updated $\boldsymbol{\theta}$ value respects the non-negativity constraint

$$\theta^{(k+1)} = argmax_{\theta}\, \mathcal{L}_{\alpha}\left(\theta, \beta^{(k)}, \eta^{(k)}, \gamma^{(k)}\right) \quad ; \boldsymbol{\theta} \geq \mathbf{0} \text{ for all } \boldsymbol{u}$$

Methodology – Alternately updates

- **STEP 2**: With $\boldsymbol{\theta}^{(\boldsymbol{k}+\mathbf{1})}$, $\boldsymbol{\beta}^{(\boldsymbol{k})}$, $\boldsymbol{\gamma}^{(\boldsymbol{k})}$ computed $\boldsymbol{\eta}^{(\boldsymbol{k}+\mathbf{1})}$;
  - by running one iteration of the MI algorithm
  - followed by a line search
  - this step is standard and ensures that $\mathcal{L}_\alpha\left(\theta^{(k+1)}, \beta^{(k)}, \eta, \gamma^{(k)}\right)$ increases as a function of $\boldsymbol{\eta}$

$$\eta^{(k+1)} = argmax_\eta\, \mathcal{L}_\alpha\left(\theta^{(k+1)}, \beta^{(k)}, \eta, \gamma^{(k)}\right) \quad ;\ \boldsymbol{\eta_i} \geq -\boldsymbol{\theta_u}\ ;\ \boldsymbol{t_i} \in \boldsymbol{B_u}$$

Methodology – Alternately updates

- **STEP 3**: With $\boldsymbol{\theta}^{(\boldsymbol{k}+\mathbf{1})}$, $\boldsymbol{\eta}^{(\boldsymbol{k}+\mathbf{1})}$ and $\boldsymbol{\gamma}^{(\boldsymbol{k})}$ computed $\boldsymbol{\beta}^{(\boldsymbol{k}+\mathbf{1})}$;
  - by running one iteration of the Newton algorithm
  - use Armijo's rule to perform line search

$$\beta^{(k+1)} = argmax_{\beta}\, \mathcal{L}_{\alpha}\left(\theta^{(k+1)}, \beta, \eta^{(k+1)}, \gamma^{(k)}\right)$$

$$\beta^{(k+1)} = \beta^{(k)} - \omega^{(k)}. \left[\frac{\partial^2 \mathcal{L}_{\alpha}}{\partial \beta_j \partial \beta_k}\right]^{-1} \left[\frac{\partial \mathcal{L}_{\alpha}}{\partial \beta_j}\right]$$

- **STEP 4:** With $\boldsymbol{\theta}^{(\boldsymbol{k}+\mathbf{1})}$, $\boldsymbol{\eta}^{(\boldsymbol{k}+\mathbf{1})}$ and $\boldsymbol{\beta}^{(\boldsymbol{k}+\mathbf{1})}$ updated $\boldsymbol{\gamma}^{(\boldsymbol{k}+\mathbf{1})}$ as follows;

$$\gamma^{(k+1)} = \gamma^{(k)} + \alpha. \left(X^T \beta^{(k+1)} - \eta^{(k+1)}\right)$$

Examples

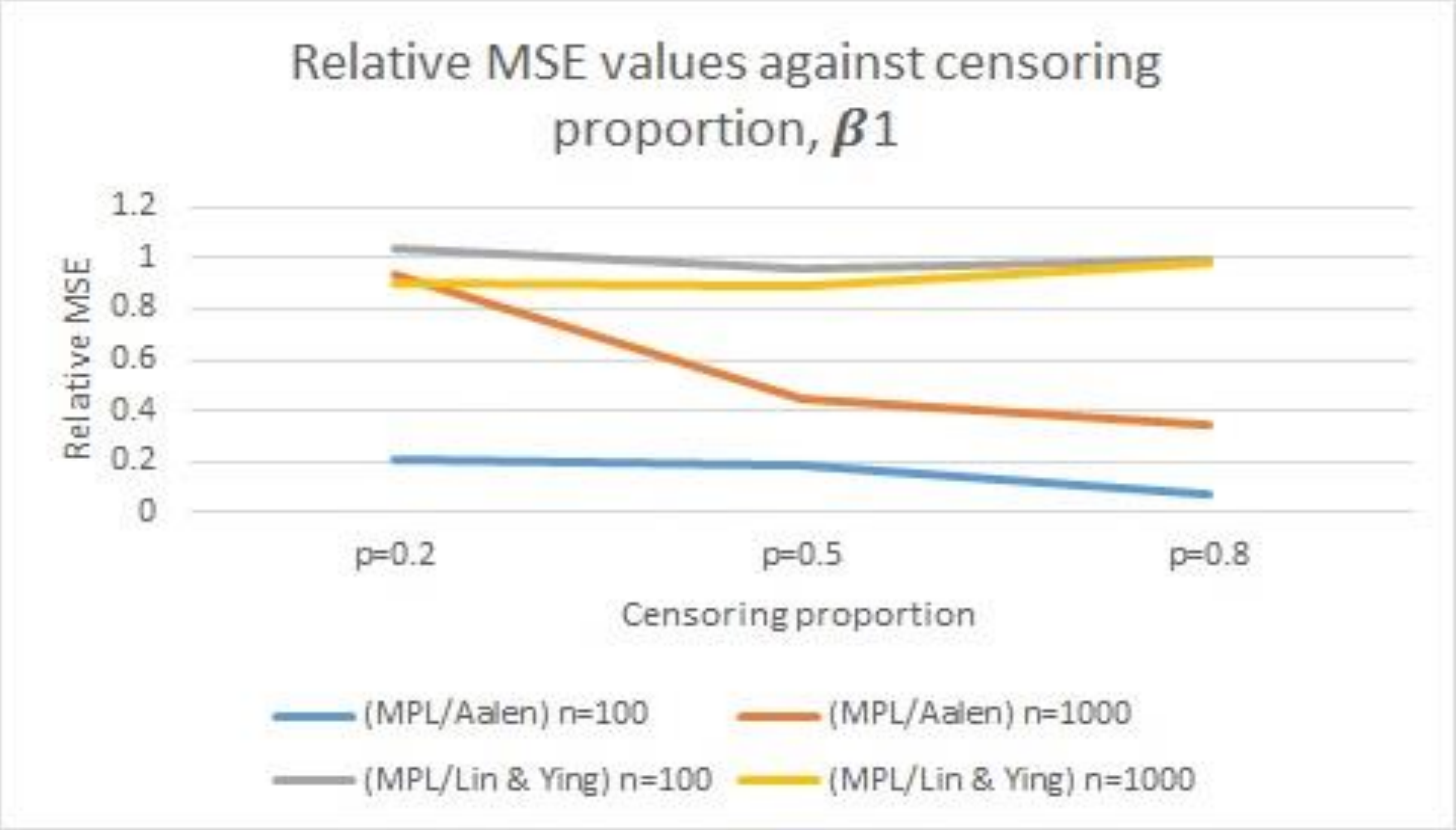
- In the examples, the main aims are to
  - demonstrate this method of estimating $\boldsymbol{\beta}$ and $\boldsymbol{h_0(t)}$ works well
  - study the behavior of the results under different censoring proportions and number of events
  - compare the results with the existing methods
- Indicator function is used as the basis function in $\boldsymbol{h_0(t)}$ estimation
- Arbitrary selected smoothing parameter, $\boldsymbol{\lambda}$ $(= \mathbf{0.05})$ value is used
- Simulate survival times (t) from Weibull distribution with hazard;

$$\boldsymbol{h_i(t) = 3t^2 + (x_{i1} + 0.6x_{i2} - 0.8x_{i3})}$$

Results – Regression coefficient estimation

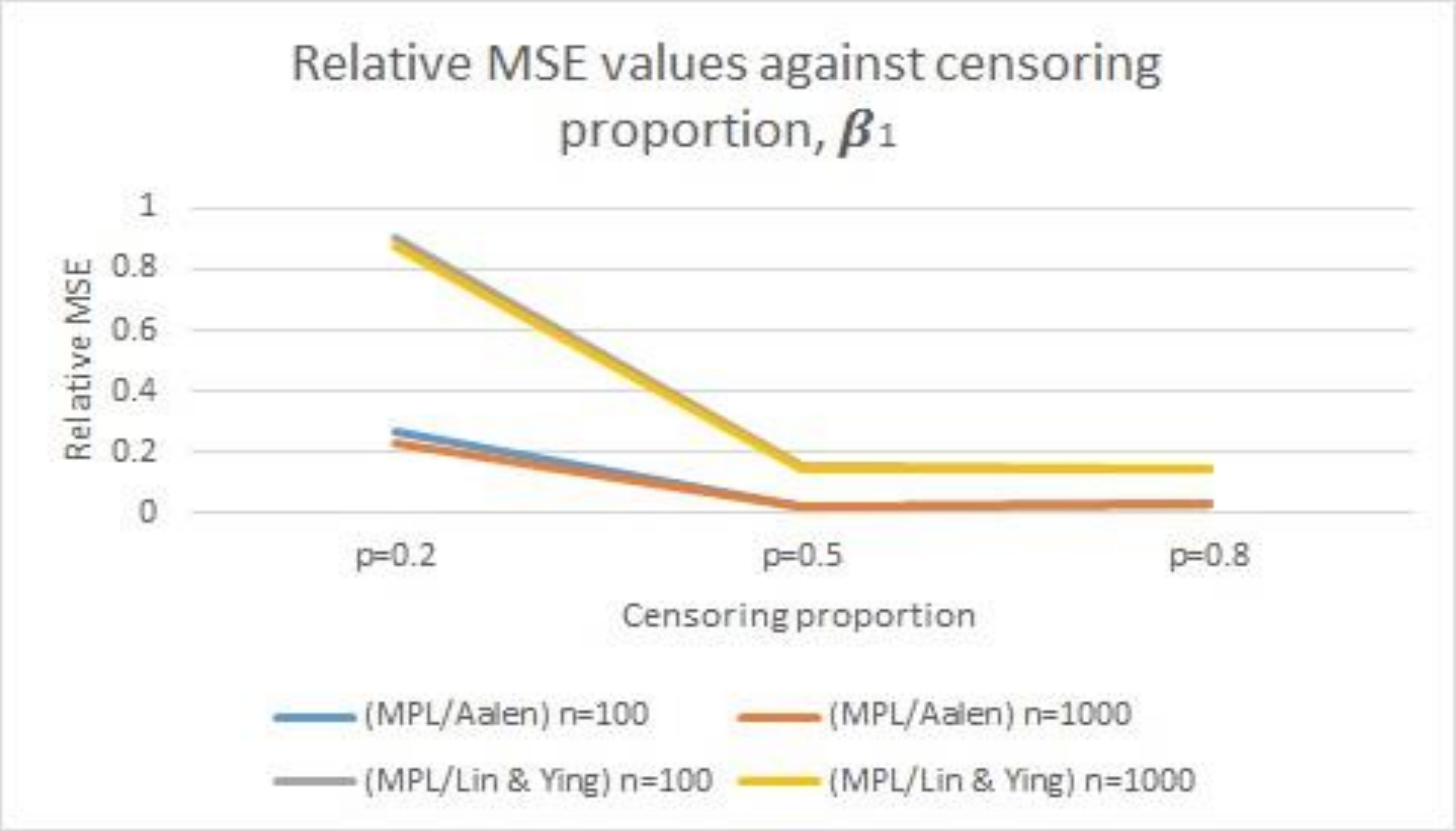
- Considered $n = 100$ & $n = 1000$ with approximate censoring proportions; π_c of 20%, 50% and 80% for each value of n
- Compared results **(MPL)** with two existing estimation procedures;
   i. Aalen's additive hazard models – using "aareg" function of **Survival** R package
   ii. Lin & Ying's additive hazards model – using "ahaz" function of **Ahaz** R package

Results : study ONE – Right censored data MSE comparison



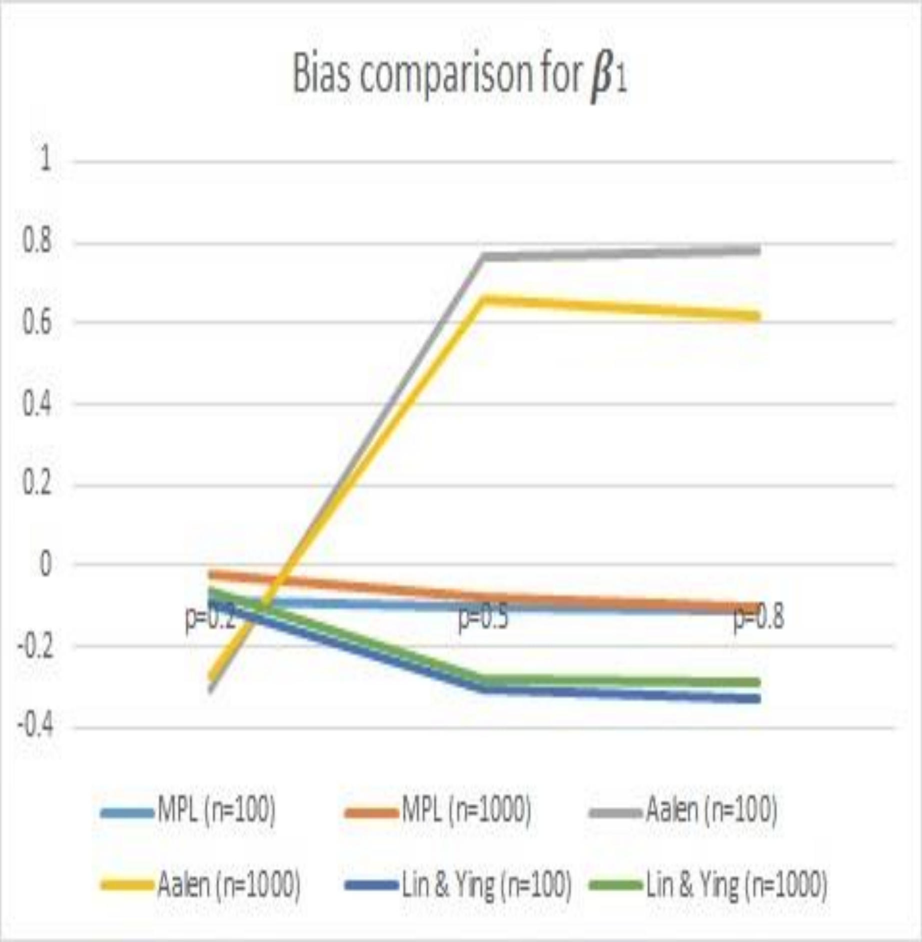
Aalen's method poorly performed with higher censoring proportions
MPL method performs slightly better than Lin & Ying's method

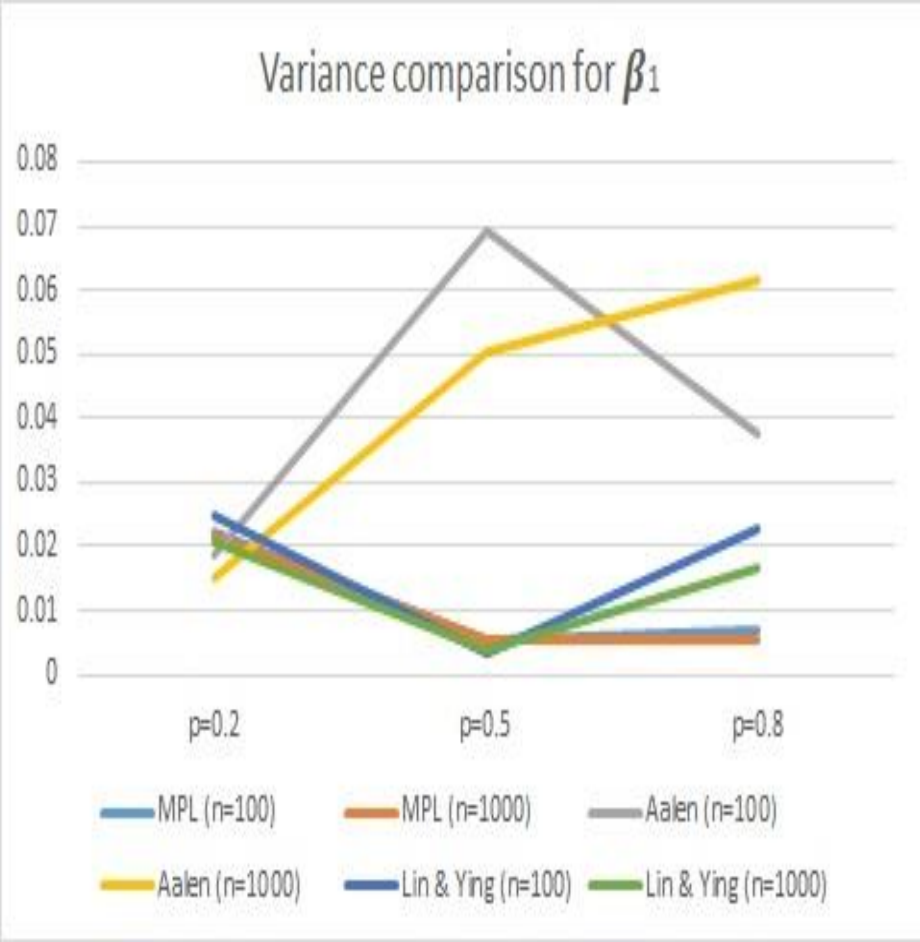
Results : study TWO – Interval censored data MSE comparison



Aalen's method poorly performed with interval censored data
Lin & Ying's method not stable with higher censoring proportions

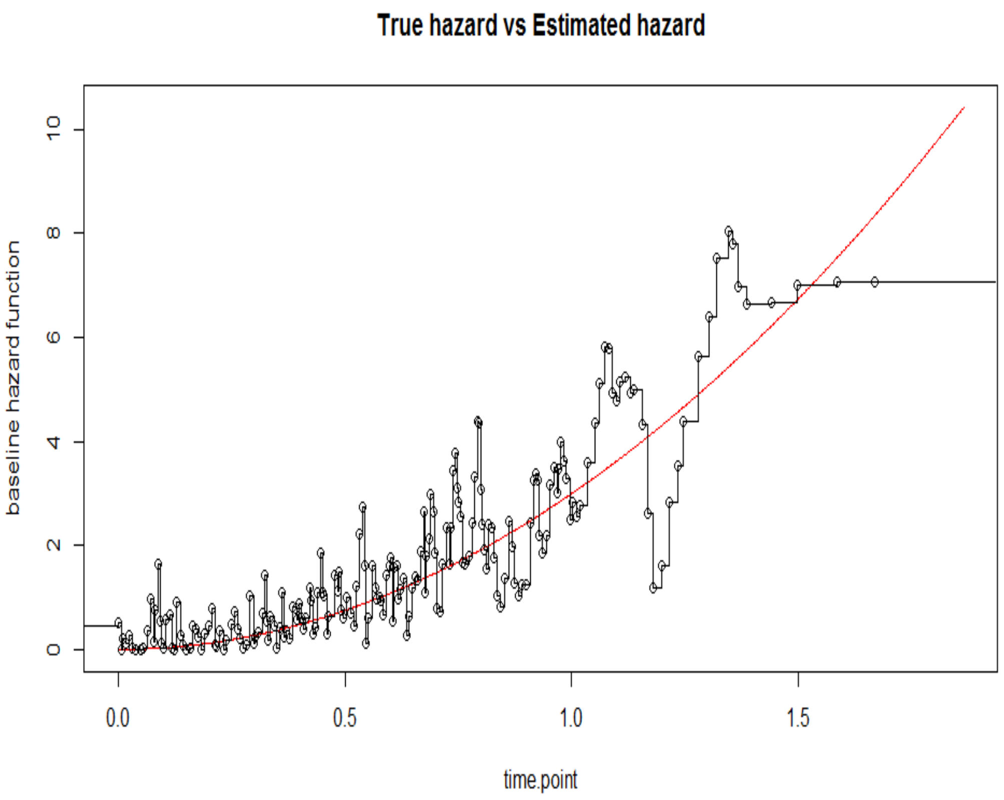
Results : study TWO – Interval censored data
Bias & Variance comparison



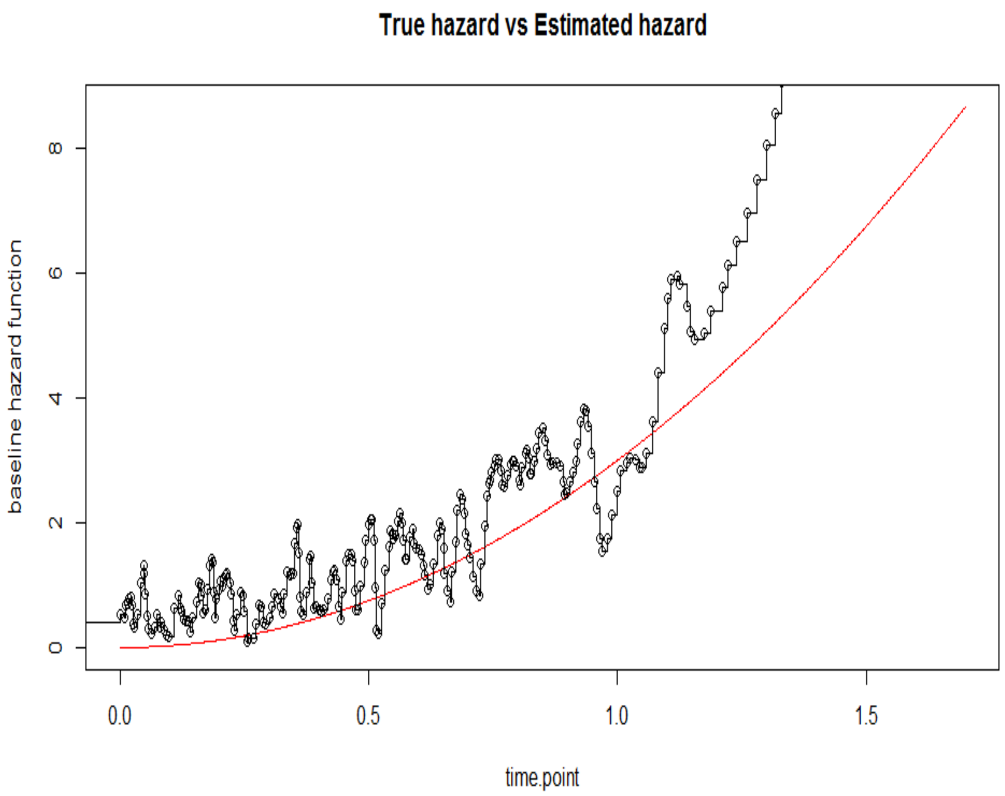


Bias and variance increases with the censoring proportion, but decreases with the sample size

Results - Baseline hazard estimation (Interval censored data)



$n = 1000, \lambda = 0.05$ and $p = 0.2$



$n = 1000, \lambda = 0.05$ and $p = 0.8$

Conclusion

- MPL produces better results than Aalen's method (**Right & Interval cens**)
- MPL produces slightly better results than Lin & Ying's method (**Right cens**) & Lin & Ying's methods produces roughly same results as MPL for lower censoring proportions, but under performed for higher censoring proportions (**Interval cens**)
- MPL produces estimates that are less biased than other methods, leading to a substantial MSE reduction
- Overall, MPL provides a gain in efficiency over Aalen's and Lin & Ying's method
- Baseline hazard estimation performs well even with the random λ value, could be improved by selecting an optimal smoothing parameter

Future work

- Extend the algorithm for different basis functions, including spline function
- Implement a procedure to obtain the optimal smoothing parameter $\boldsymbol{\lambda}$
- Develop asymptotic properties of the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$

References

- Aalen, O.O. (1980), “A model for nonparametric regression analysis of counting processes”, *Lecture notes in Statistics,* 2, Springer-Verlag, 1-25
- Cai, T. and Betensky, R.A. (2003), “Hazard regression for interval censored data with penalized spline”, *Biometrics*, 59,570-579
- J. Ma, S. Heritier and S. Lo, (2014),"On the maximum penalized likelihood approach for proportional hazard models with right censored survival data", *Computational Statistics and Data Analysis*, vol 74, pp 142 - 156, 2014.
- Lin, D. Y. and Ying, Z. (1994), “Semiparametric analysis of the additive risk model”, *Biometrika*, 81:61 – 71.
- Lin, D. Y. and Ying, Z. (2007), “Maximum likelihood estimation in semi-parametric regression models with censored data”, *J. Roy. Statist Soc.B* , 69:507 – 564.
- Luenberger, D. (1984). *Linear and Nonlinear Programming (2nd edition)*. J. Wiley.
- McCullagh, P. and Nelder, J. A. (1989), *Generalized linear models (2nd edition).* Chapman and Hall, London.

THANK YOU…!!!

Literature Review

- Here we review three approaches on fitting the Additive hazards model
- **Least squares approach**
  - Aalen (1980) used formal least squares principle for nonparametric additive hazard model
  - Firstly obtained cumulative versions of the estimates using continuous data
  - Then, the coefficients can be estimated from the slope of the cumulative estimates
  - Leads to well-known Nelson-Aalen estimate for cumulative hazard estimation

*Could not extend this model for interval censored data

Literature Review

- **Maximum Likelihood (ML) approach**
  - Ghosh (2001) and Zeng et al.(2006) developed ML approach for additive hazards model.
  - Ghosh fits the additive hazards model by estimating β and a cumulative baseline hazard function $H_0(.)$
    - Used primal-dual interior point algorithm
    - Algorithm imposes contraints of positivity & monotonic increasing on $H_0(.)$ and the cumulative hazard $H_i(.)$
  - Zeng et al. fit the additive hazards model with interval censored data
    - Using the log likelihood function which is expressed in terms of $S_0(.)$ and β
    - This constraints positivity and monotonic decreasing on $S_0(.)$ by using a logarithm transformation

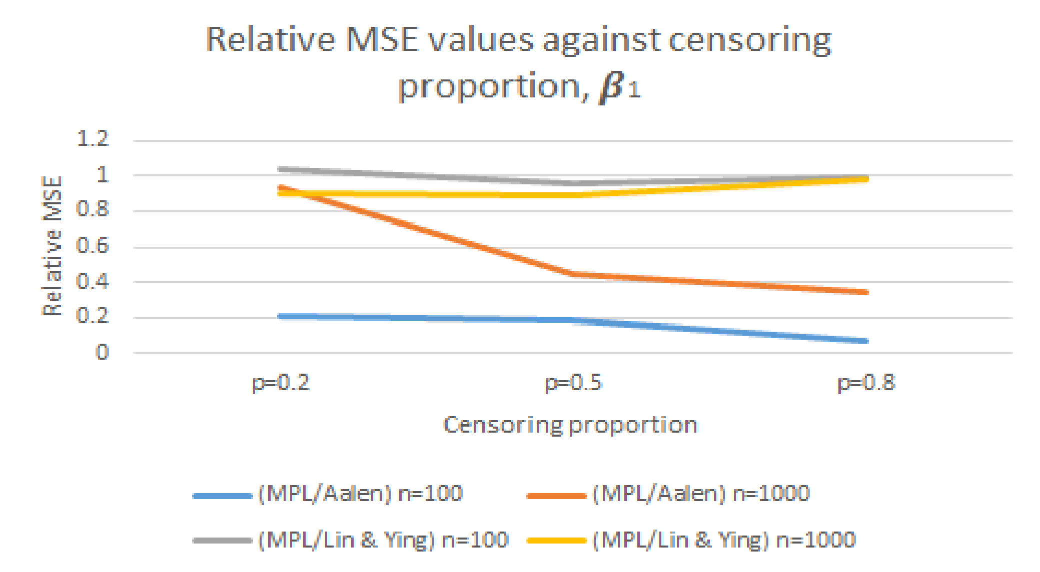
* Here the way the constraints are imposed can make the estimation procedure unstable when the baseline survival estimate approaches zero.

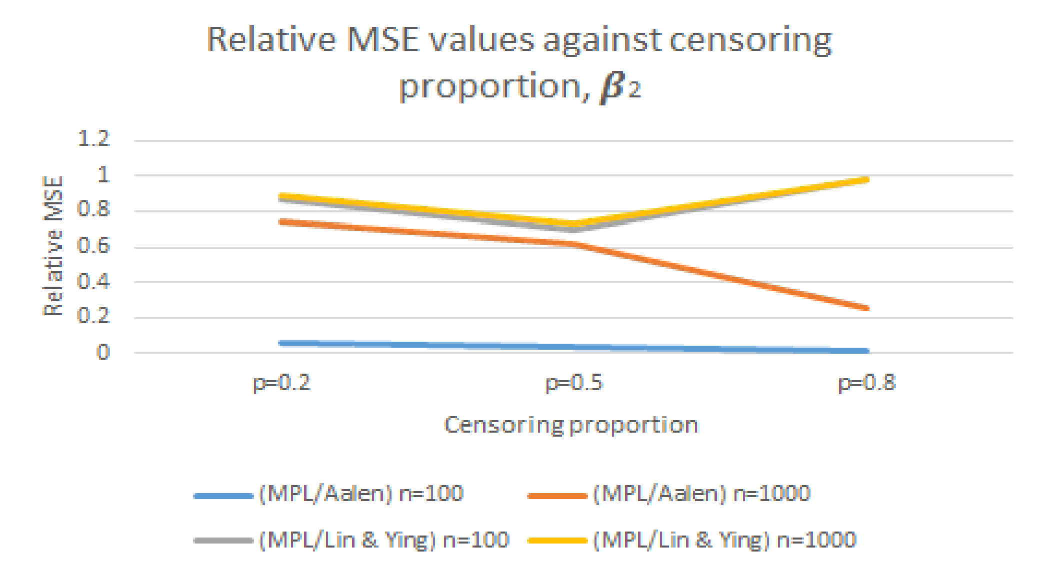
Literature Review

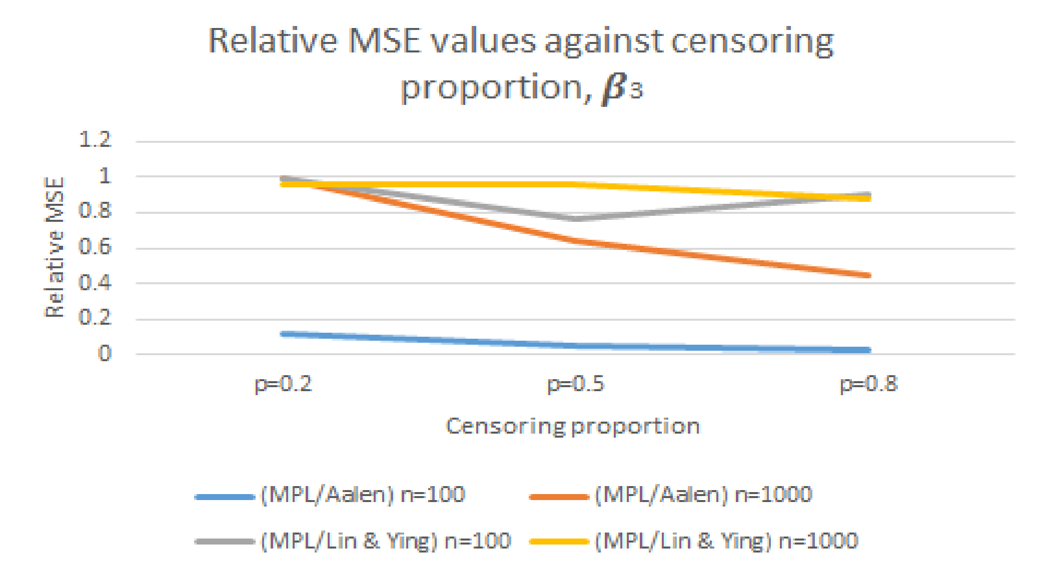
- **Generalized Linear Model (GLM) approach**
  - Farrington (1996) fits the additive hazards model for interval censored data using a generalized linear model (GLM) approach
    - The occurrences left, right and interval censored observations are assumed to be from indepedent Bernoulli trials
    - Occurrence probability is related to a linear predictor by a negative log link function
    - Then β and $h_0(.)$ can be estimated by fitting the generalized linear model
    - The baseline hazard $h_0(.)$ is assumed to be piecewise constant over some intervals

*Neither non-negativity nor smoothness of the $h_0(.)$ can be guaranteed

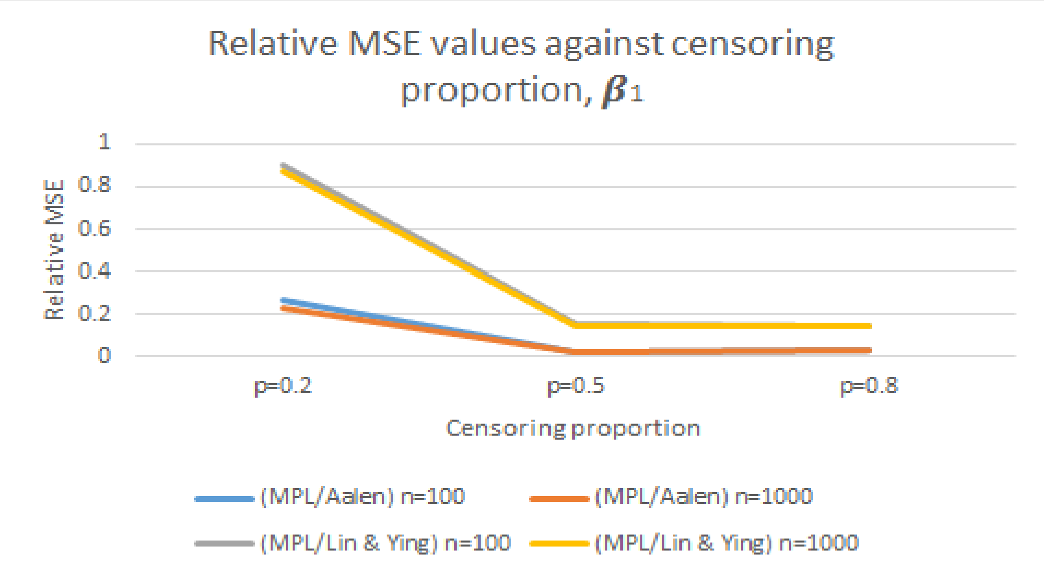
Simulation results : study ONE – Right censored data MSE comparison

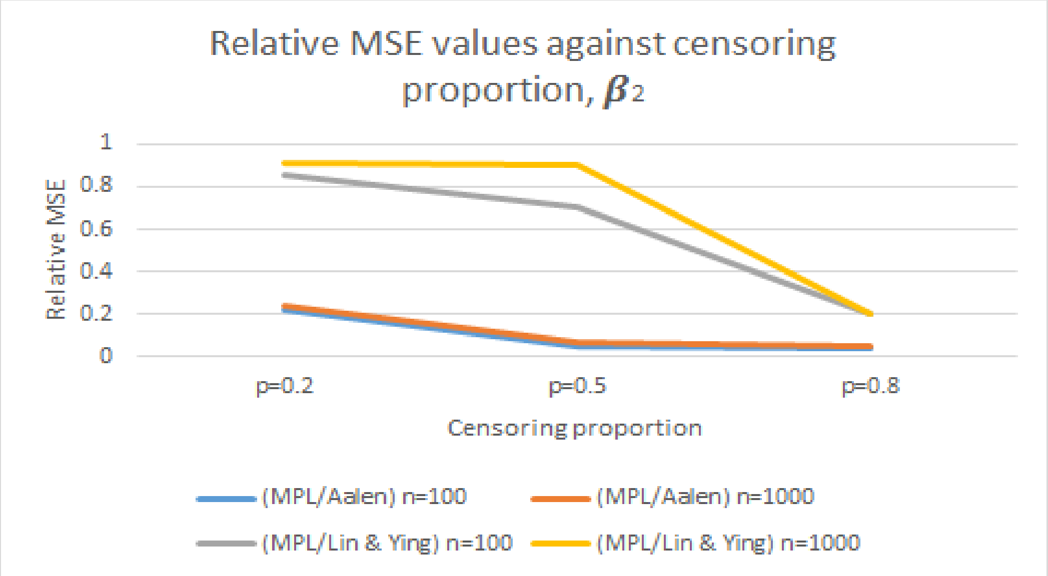


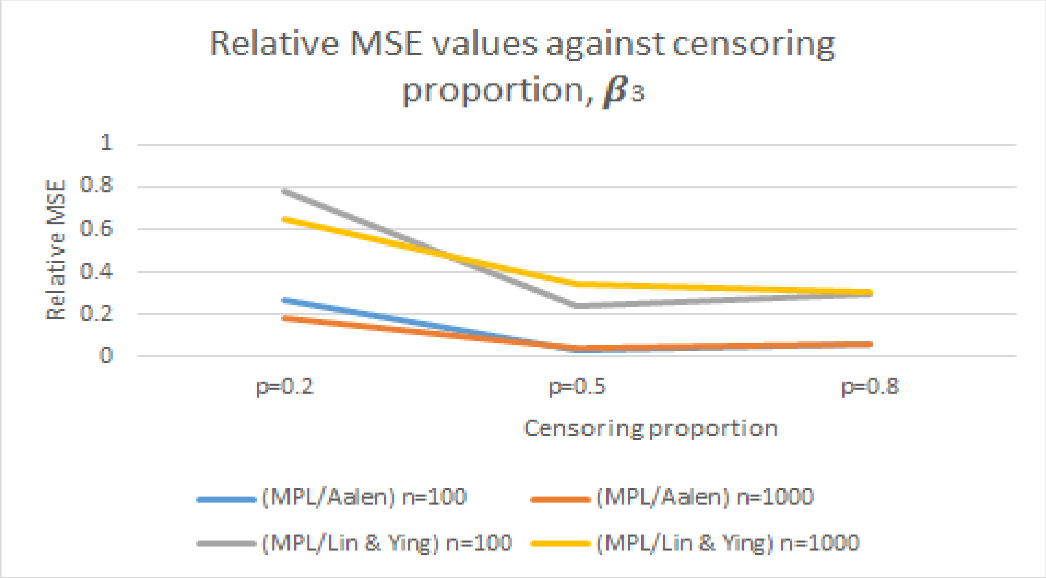




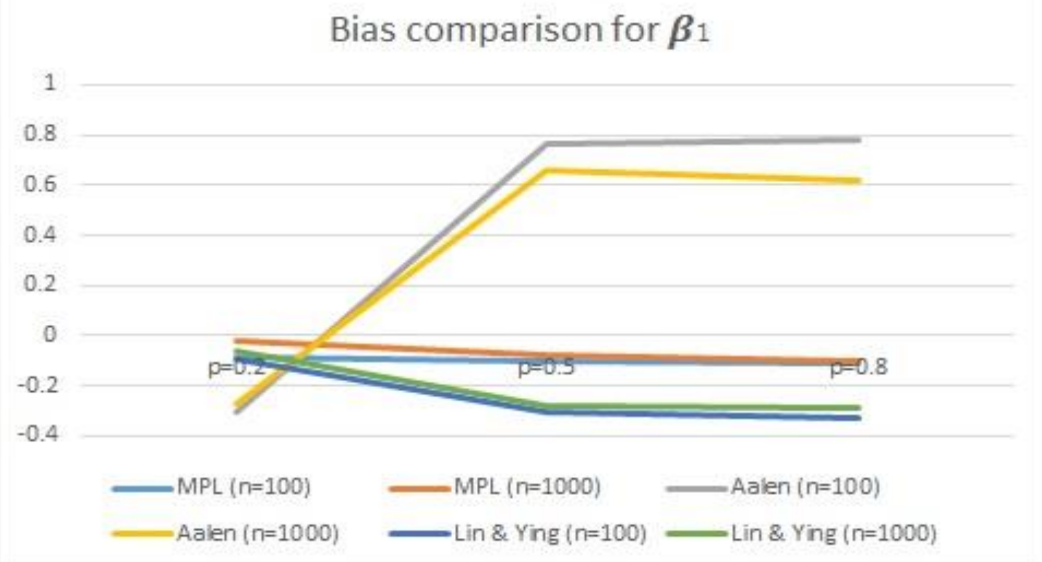
Simulation results : study TWO – Interval censored data MSE comparison

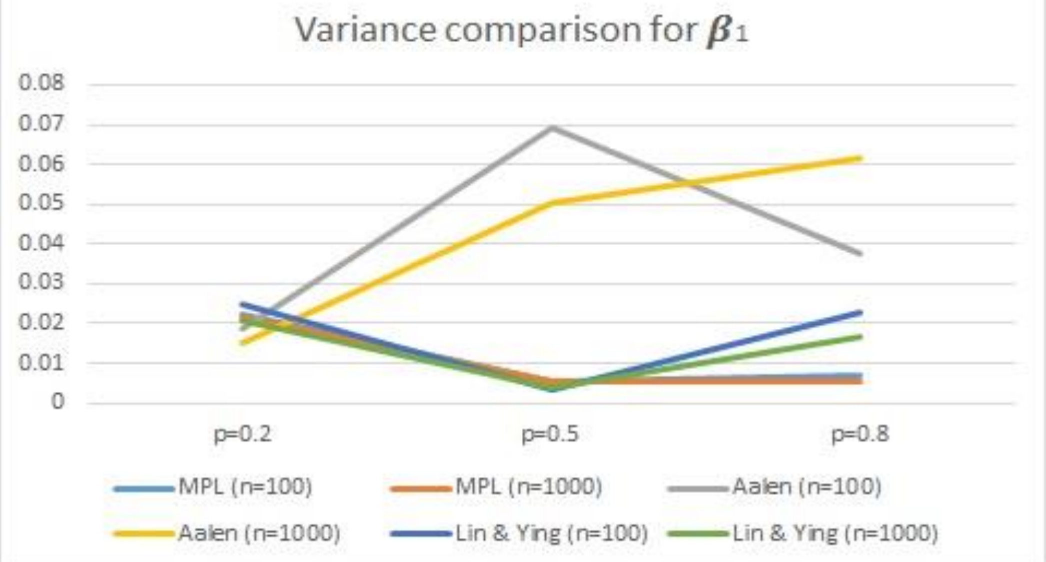


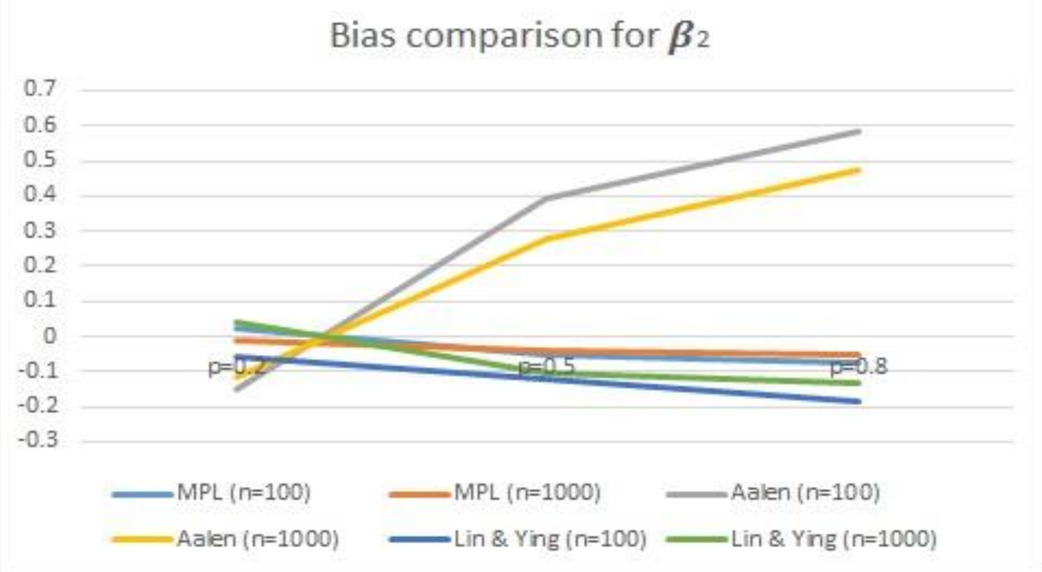


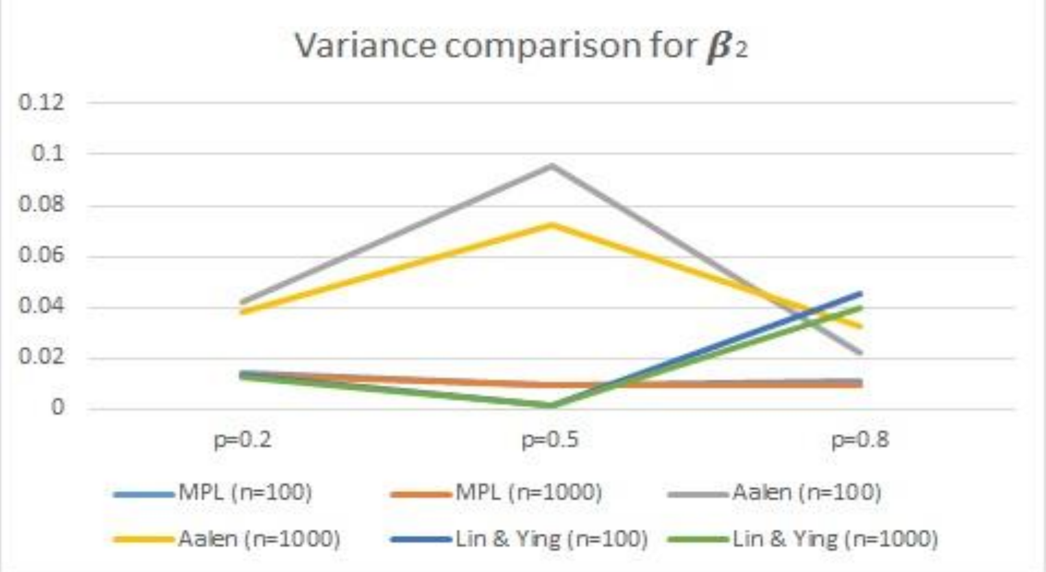


Simulation results : study TWO – Interval censored data
Bias & Variance comparison









Simulation results : study TWO – Interval censored data
Bias & Variance comparison