Liver Cancer: Investigating the Risk Factors

Eric Roseren
School of Mathematics
University of Edinburgh

April 28, 2021

Content

Aim

Task:

- Investigating the risk factors of HCC in patients undergoing screening from 2009 to 2017 in the South East of Scotland.

Research questions:

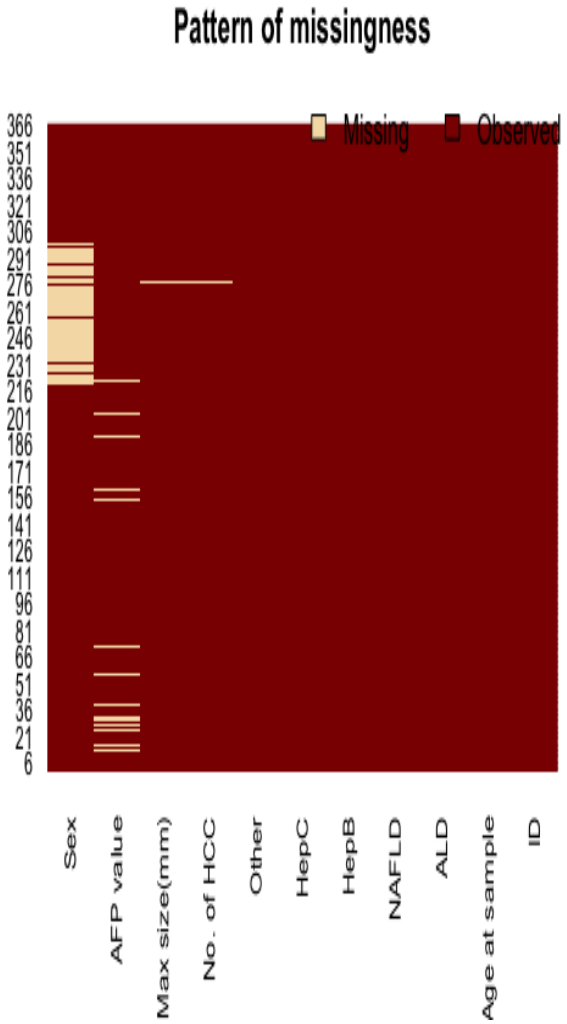
- What is the risk of HCC development in populations with different causes of liver disease?
- Is the male/female gender bias in HCC development stronger in some causes of liver disease?
- How strong is the association between AFP levels and the risk of HCC?
- Can the observed AFP levels provide prediction on survival probabilities?

The dataset

Screening and HCC cohort

1. Patient in screening (1506)
2. Patient in HCC (240)
3. log_{10}(AFP) measures [1]
4. Covariates
   - Age at sample
   - Observation time
   - Time to event
   - Gender
   - Aetiology
   - Status

Data in HCC cohort



The modelling approach

1. Cox proportional hazards model as survival function estimator to estimate the survival probability over time [2].
2. Linear mixed-effect model is used to fit and predict the evolution of the biomarker over time for each patient.
3. Combine both method to obtain a joint model which can measure the association of the AFP measures and the risk of HCC [3].

Univariate Cox regression

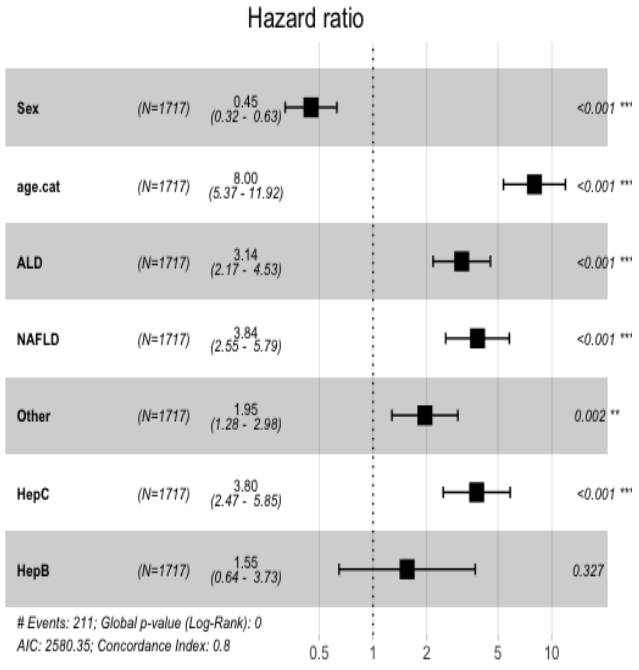
$$h_i(t \mid w_i) = \lim_{dt \to 0} \frac{\Pr[t \leq T^* < t + dt \mid T^* \geq t, w_i]}{dt}$$
$$= h_0(t) e^{\gamma^T w_i}$$

Table: Coefficient estimates of the Cox model

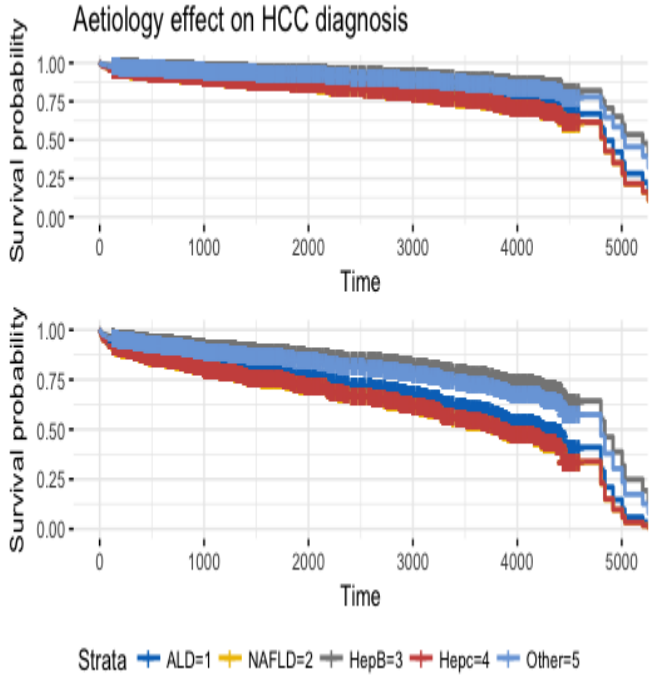
	beta	Hazard Ratio (95% CI)	p-value
Sex	-0.68	0.51 (0.37-0.69)	1.4e-05
Age	1.6	5.2 (3.7-7.1)	0
ALD	0.73	2.1 (1.6-2.7)	2e-07
NAFLD	1.1	3.2 (2.3-4.4)	7.7e-12
HepB	-1.3	0.27 (0.12-0.6)	0.0015
HepC	-0.18	0.83 (0.62-1.1)	0.24

Multivariate Cox regression

Multivariate Cox regression



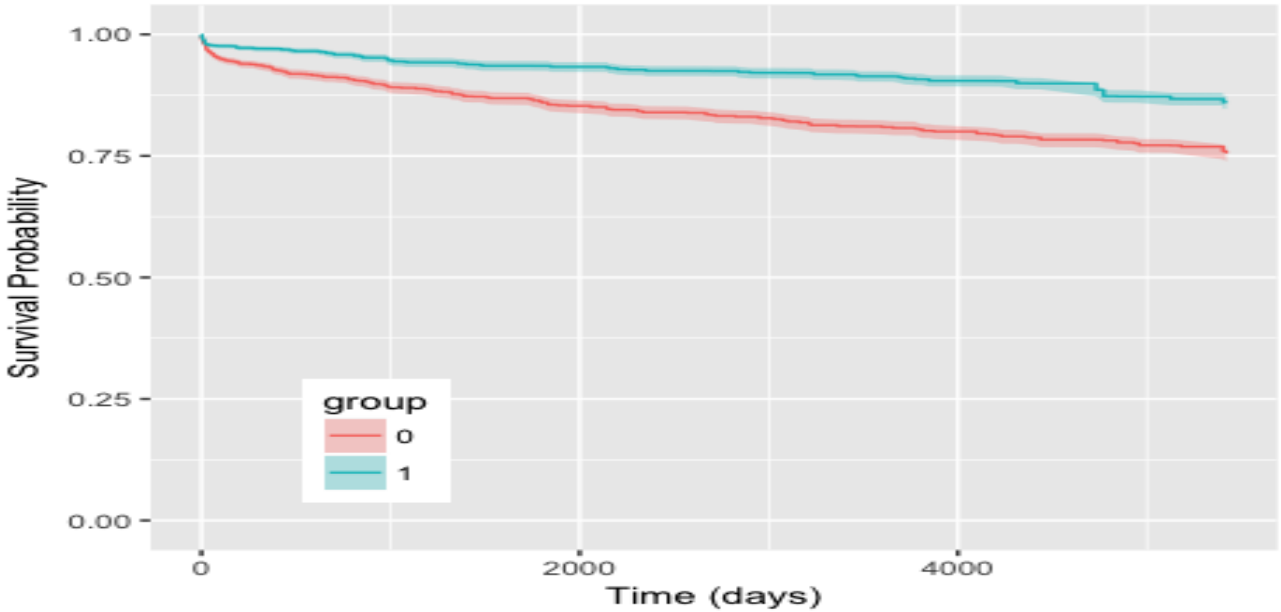
Survival curve for each aetiology



- ALD,NAFLD and Hepatitis C suffering patients more at risk of developing HCC.

Multivariate Cox regression (continue)

Survival probabilities difference between male and female



- Female survival outcome better than males.
- Median survival rate of women between 1 to 3 years greater than men.
- Gender bias for HCC more pronounced in ALD, NAFLD and Hepatitis C compared to Hepatitis B and autoimmune diseases.

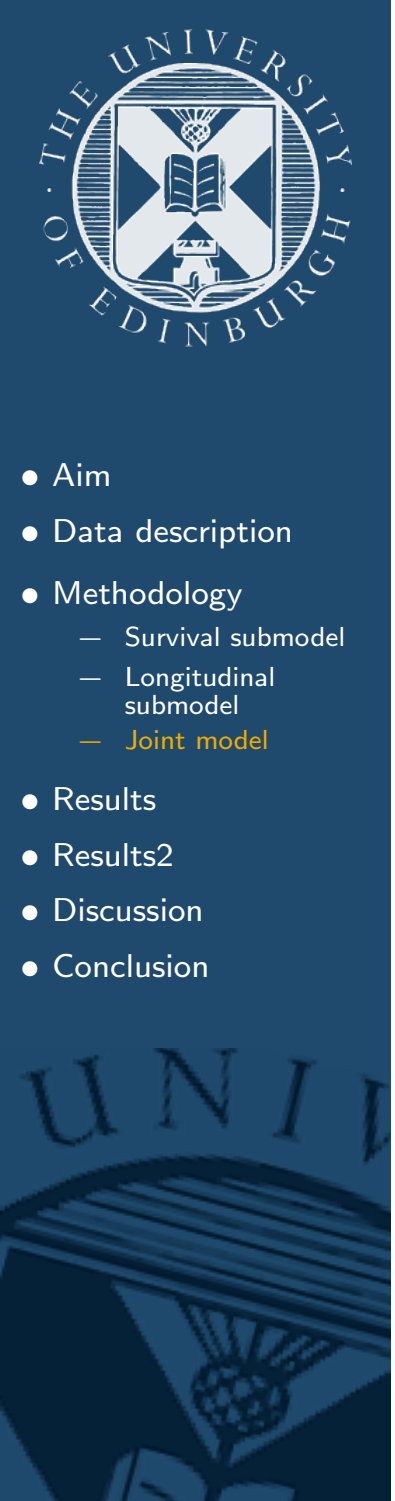
Linear mixed-effects model

$$\begin{cases} y_i(t) = m_i(t) + \epsilon_i(t), \\ m_i(t) = x_i^T(t)\beta + z_i^T b_i, \\ \quad b_i \sim N(0, D), \quad \epsilon_i(t) \sim N(0, \sigma^2), \end{cases}$$

where

- $x_i(t)$ and β are the Fixed-effects part of the model
- $z_i(t)$ and b_i are the Random-effects part of the model

1. $\beta + b_i$ describes individual response trajectories
2. Can obtain different average longitudinal evolution per aetiology.
3. can accommodate the fact that patients have different number of repeated measurements.



Joint model

The intuitive idea behind Joint models

- The evolution of the biomarker over time, $m_i(t)$ is described by the longitudinal model.
- We can use the estimated evolutions in a Cox model.
- The two models are combined to estimate their joint distribution.

The longitudinal submodel

$$
\begin{cases}
y_i(t) = m_i(t) + \epsilon_i(t), \\
m_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) B_n(t, 3) + \beta_2 \, Sex \\
\qquad + \beta_3 \, Age_i + \beta_4 \, ALD_i + \beta_5 \, NAFLD_i \\
\qquad + \beta_6 \, HepB_i + \beta_7 \, HepC_i, \\
b_i \sim N(0, D), \quad \epsilon_i(t) \sim N(0, \sigma^2),
\end{cases}
$$

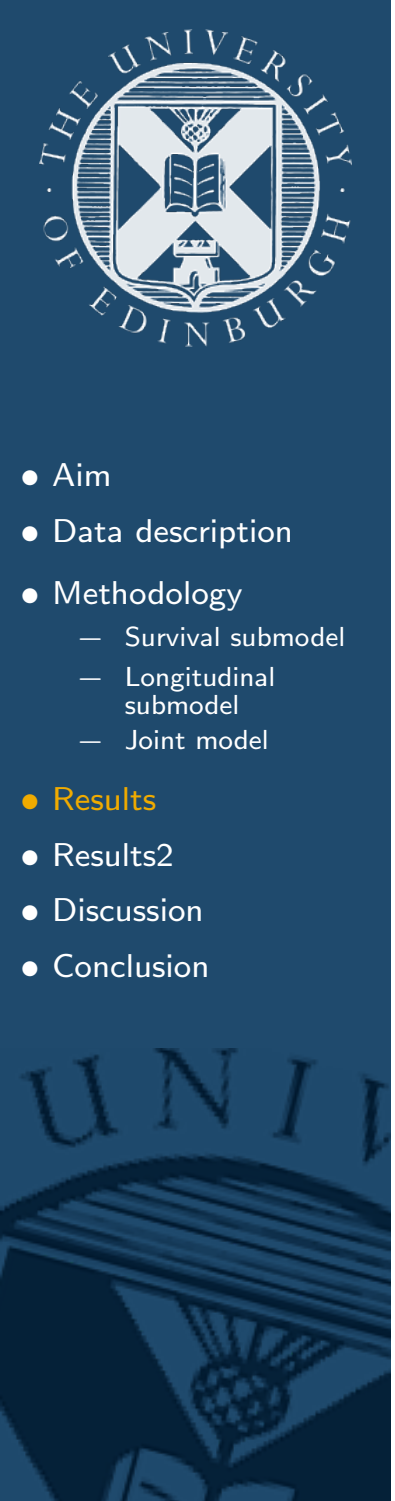
Joint model (continue)

The survival submodel

$$h_i(t \mid M_i(t), w_i) = h_0(t) e^{\gamma^T \mathbf{w_i} + \alpha_1 m_i(t) + \alpha_2 \frac{d\, m_i(t)}{dt}} \quad \text{where}$$

$$\gamma^T \mathbf{w_i} = \gamma_1 + \gamma_2\, Age_i + \gamma_3\, ALD_i + \gamma_4\, NAFLD_i + \gamma_5\, HepB_i + \gamma_6\, HepC_i$$

- $M_i(t)$ is the longitudinal measurement history of the biomarker.
- α quantifies the strength of the association between log(AFP) levels and risk of HCC.
- w_i,the different predictors at baseline.



Joint model (continue)

The joint model distribution [4]

$$p(T_i, \delta_i, y_i) = \int p(y_i \mid b_i)\{h(T_i \mid b_i)^{\delta_i} S(T_i \mid b_i)\} p(b_i)\, db_i$$

- $S(.)$ denotes the survival function and $p(.)$ the density function.
- estimation is done under the Bayesian approach (MCMC)

posterior distribution

$$p(\boldsymbol{\theta}, \boldsymbol{b}) \propto \prod_{i=1}^{n} \prod_{l=1}^{n_i} p(\boldsymbol{y_i} \mid \boldsymbol{b_i}, \boldsymbol{\theta}) p(T_i, \delta_i \mid \boldsymbol{b_i}, \boldsymbol{\theta}) p(\boldsymbol{b_i} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

Learning and controls using GP

$$y = f(x_1, x_2, \ldots, x_p,$$

- $S(.)$ denotes the survival function and $p(.)$ the density function.
- estimation is done under the Bayesian approach (MCMC)

posterior distribution

$$p(\boldsymbol{\theta}, \boldsymbol{b}) \propto \prod_{i=1}^{n} \prod_{l=1}^{n_i} p(\boldsymbol{y_i} \mid \boldsymbol{b_i}, \boldsymbol{\theta}) p(T_i, \delta_i \mid \boldsymbol{b_i}, \boldsymbol{\theta}) p(\boldsymbol{b_i} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

Joint model (continue)

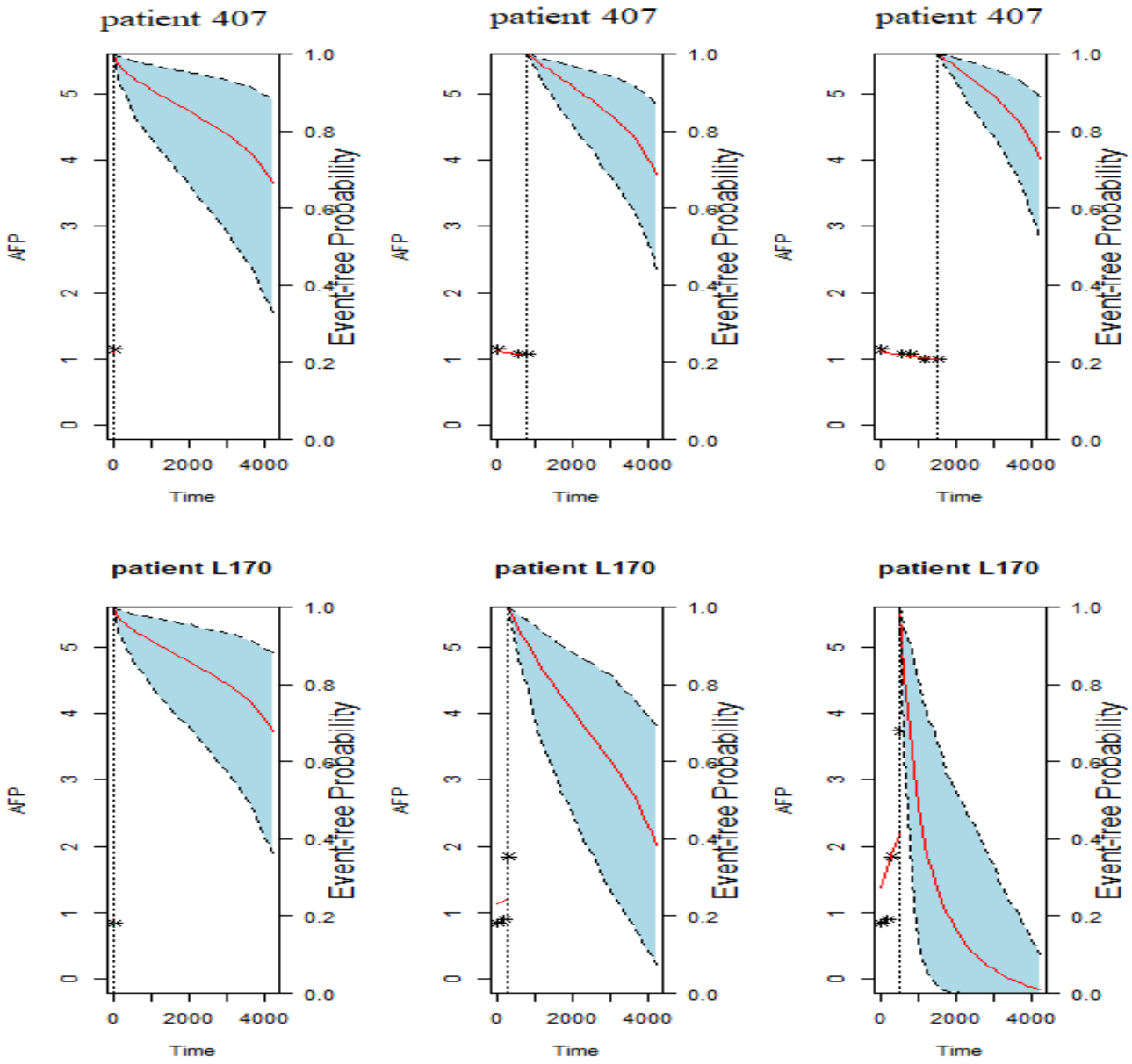
Table: Coefficient estimates for the Cox survival submodel

	Value	Std.Err	Std.Dev	2.5%	97.5%	P
Sex	−0.837	0.013	0.183	−1.188	−0.476	0
Age	0.095	0.002	0.007	0.080	0.106	0
ALD	0.789	0.011	0.159	0.494	1.117	0
NAFLD	0.533	0.013	0.198	0.136	0.922	0.001
HepB	0.298	0.032	0.446	−0.603	1.126	0.510
HepC	0.771	0.016	0.192	0.382	1.154	0
Assoct	1.930	0.005	0.094	1.747	2.111	0
AssoctE	0.252	0.079	3.102	−5.954	6.225	0.009
tauBs	66.612	8.017	71.051	3.522	259.697	

- The intercept and slope of the biomarker trajectory are highly associated with the risk of HCC.
- one unit in log(AFP) increases the risk of HCC by 29%.

Dynamic Prediction

- Aim
- Data description
- Methodology
  - Survival submodel
  - Longitudinal submodel
  - Joint model
- Results
- Results2
- Discussion
- Conclusion



Limitations and further work

Limitations

- Informative censoring mechanism: the probability of a subject being censored depends on the failure process.
- Proportional hazard assumption: explanatory variable acts multiplicatively on the hazard ratio (not directly on the failure time)

Further work

- Multiple Longitudinal Markers (e.g platelet,DCP treatment, etc ...)
- Imputation of missing data
- correction for potential selection bias due to loss to follow up (e.g inverse probability-of-censoring weighted estimation)

Conclusion

- ALD,NAFLD and Hepatitis C suffering patients more at risk of developing HCC.
- median survival time 9.5 years vs 13 years for "Other" category.
- gender bias in developing HCC more pronounced in ALD, NAFLD and Hepatitis C compared to other aetiologies.
- The intercept and slope of the biomarker trajectory are highly associated with the risk of HCC.

References

Irina Teofănescu, Elena Gologan, Gabriela Stefănescu, and Gh Bălan.
Surveillance of cirrhosis for hepatocellular carcinoma–clinical validation of new serological biomarkers for improved diagnosis].
Revista medico-chirurgicală a Societății de Medici și Naturaliști din Iași, 114(1):39–46, 2007.

David R Cox.
Regression models and life-tables.
In *Breakthroughs in statistics*, pages 527–541. Springer, 1992.

Dimitris Rizopoulos, Jeremy M. G. Taylor, Joost Van Rosmalen, Ewout W. Steyerberg, and Johanna J. M. Takkenberg.
Personalized screening intervals for biomarkers using joint models for longitudinal and survival data.
Biostatistics, 17(1):kxv031, 2015.

Anastasios A Tsiatis and Marie Davidian.
Joint modeling of longitudinal and time-to-event data: an overview.
Statistica Sinica, pages 809–834, 2004.

Thank you.
Questions ?