An Analysis on the Efficiency of Non-Parametric Estimation of Duration Dependence by Hazard-Based Duration Model

Toshiyuki Yamamoto, Nagoya Univ.
Noritake Nakagawa, Nagoya Univ.
Ryuichi Kitamura, Kyoto Univ.
Jean-Loup Madre, INRETS

Outline

- Backgrounds
- Hazard-based duration models
  - Parametric hazard model
  - Non-parametric hazard model
- Simulation analysis
- Empirical analysis
- Conclusions

Decision-making in a temporal context

- Hazard-based duration models have been applied to represent the time until the occurrence of a choice or an event.
  - Trip timing
  - Activity engagement and duration
  - Vehicle holding duration

- Duration dependence is one of the key elements in the analysis.
  - The snowballing effects
  - The inertial effects
  - The memoriless effects

Types of hazard-based duration model

- There are several types of models in treating time dependency (baseline hazard) and the effects of exogenous variables (covariates).

Model	Time dependency	Exogenous variables
Non-parametric model	*Non-parametric*	*Non-parametric*
Semi-parametric model	*Non-parametric*	*Parametric*
Parametric model	*Parametric*	*Parametric*

Types of hazard-based duration model

- There are several types of models in treating time dependency (baseline hazard) and the effects of exogenous variables (covariates).

Model	Time dependency	Exogenous variables
Non-parametric model	*Non-parametric*	*Non-parametric*
Non-parametric model	*Non-parametric*	*Parametric*
Parametric model	*Parametric*	*Parametric*

Two approaches to estimate the non-parametric hazard models

- The partial likelihood approach (Cox, 1972)
  - Treat the baseline hazard as a nuisance parameter, and only the coefficient estimates of covariates are provided.

- The ordered-response approach (Prentice and Gloeckler, 1978)
  - Provides both the coefficient estimates of covariates and the baseline hazard.

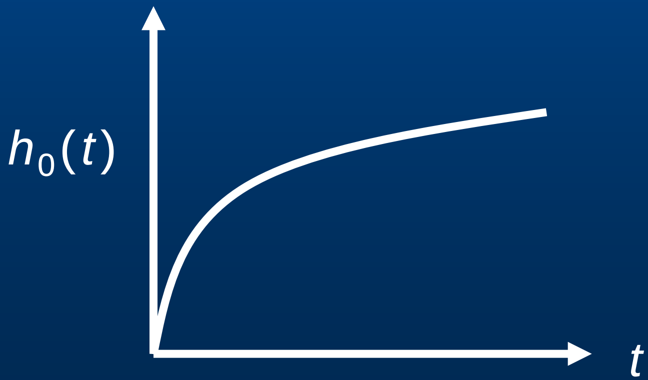
The purpose of the study is to examine the efficiency of estimating the baseline hazard, thus the ordered-response approach is examined here.

Parametric and non-parametric hazard models

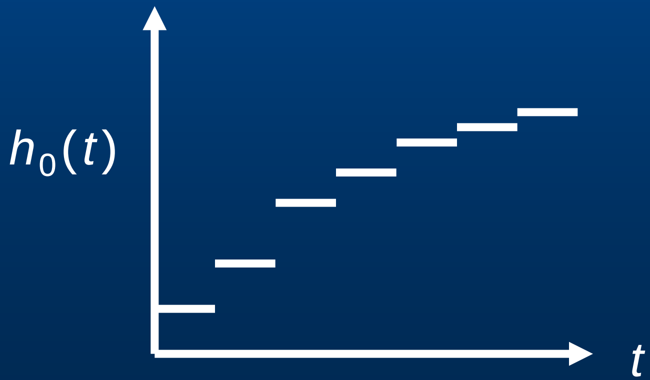
- The hazard function represents the conditional probability density that an event will occur between time t and t+dt, given that the event has not occurred up to time t.

$$h(t|X) = h_0(t)\exp(-\beta X)$$

Parametric hazard



Non-parametric hazard



Parametric and non-parametric hazard models

- The parametric hazard models result in inconsistent estimation of the baseline hazard function and the coefficients of the covariates when the assumed parametric distribution form is incorrect.

- The non-parametric hazard models are regarded to provide consistent coefficient estimates of the covariates regardless of the true baseline hazard.

The non-parametric hazard models are more robust estimator of the coefficients of the covariates.

Efficiency of the non-parametric hazard model (ordered-response estimation)

Meyer (1987, 1995) found

- The loss of the efficiency, resulting from the disregarding the information associated with the baseline hazard distribution, may not be very substantial for coefficient estimates of the covariates.

But,

- The efficiency of the estimation of the baseline hazard is not accounted for.

Efficiency of the non-parametric hazard model (ordered-response estimation)

Meyer (1987) also found

- Efficiency loss may be sometimes substantial for <u>time-varying covariates</u>.

According to Hensher and Mannering (1994) and Bhat (2000),

- The effects of duration dependence and those of <u>time-varying covariates</u> are difficult to disentangle.

change its values along the time

Objective

- If the distributional shape of the baseline hazard can be appropriately assumed, selection depends on the efficiency in estimation of the baseline hazard function as well as the coefficients for the covariates.

↓

- The efficiency of the two estimation models in estimating the baseline hazard function in a case with time-varying covariates is examined.

Simulation analysis

- The Weibull distribution is assumed as the true distribution.
- 3 covariates: 2 time-invariant (continuous and dummy) and 1 time-varying (dummy, changes its value at $t=8$)
- Sample size: 100, 300, 500 and 1000
- 10 samples for each sample size

Coefficient estimates for covariates

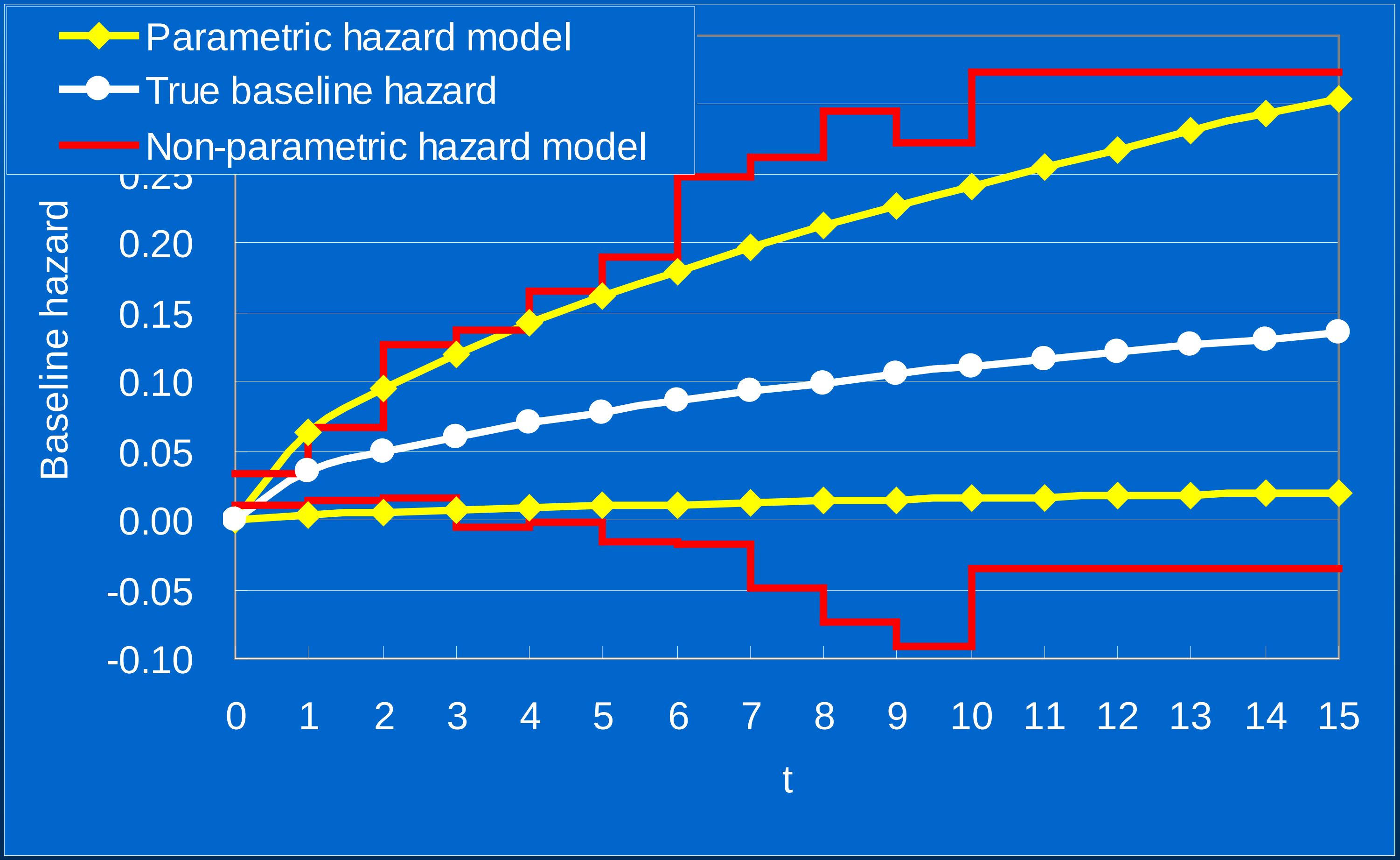
N = 100

	Par.	Non-par.
Time-invariant cont.		
E(s.e.)	0.27	0.29
MSE	0.066	0.081
Time-invariant dummy		
E(s.e.)	0.13	0.15
MSE	0.014	0.017
Time-varying dummy		
E(s.e.)	0.33	0.57
MSE	0.055	0.152

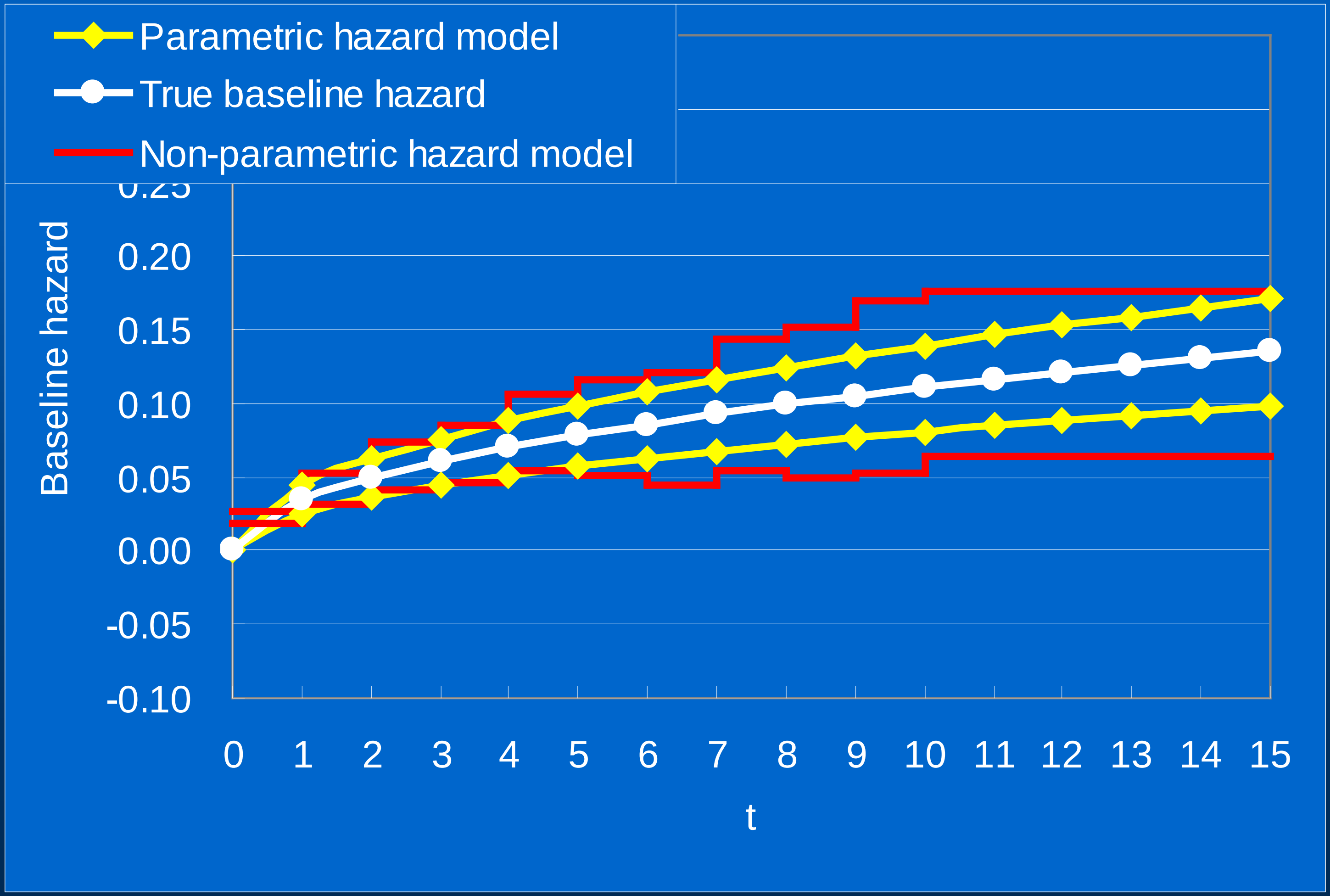
N = 1000

	Par.	Non-par.
Time-invariant cont.		
E(s.e.)	0.081	0.087
MSE	0.0019	0.0030
Time-invariant dummy		
E(s.e.)	0.040	0.044
MSE	0.0014	0.0015
Time-varying dummy		
E(s.e.)	0.11	0.18
MSE	0.020	0.033

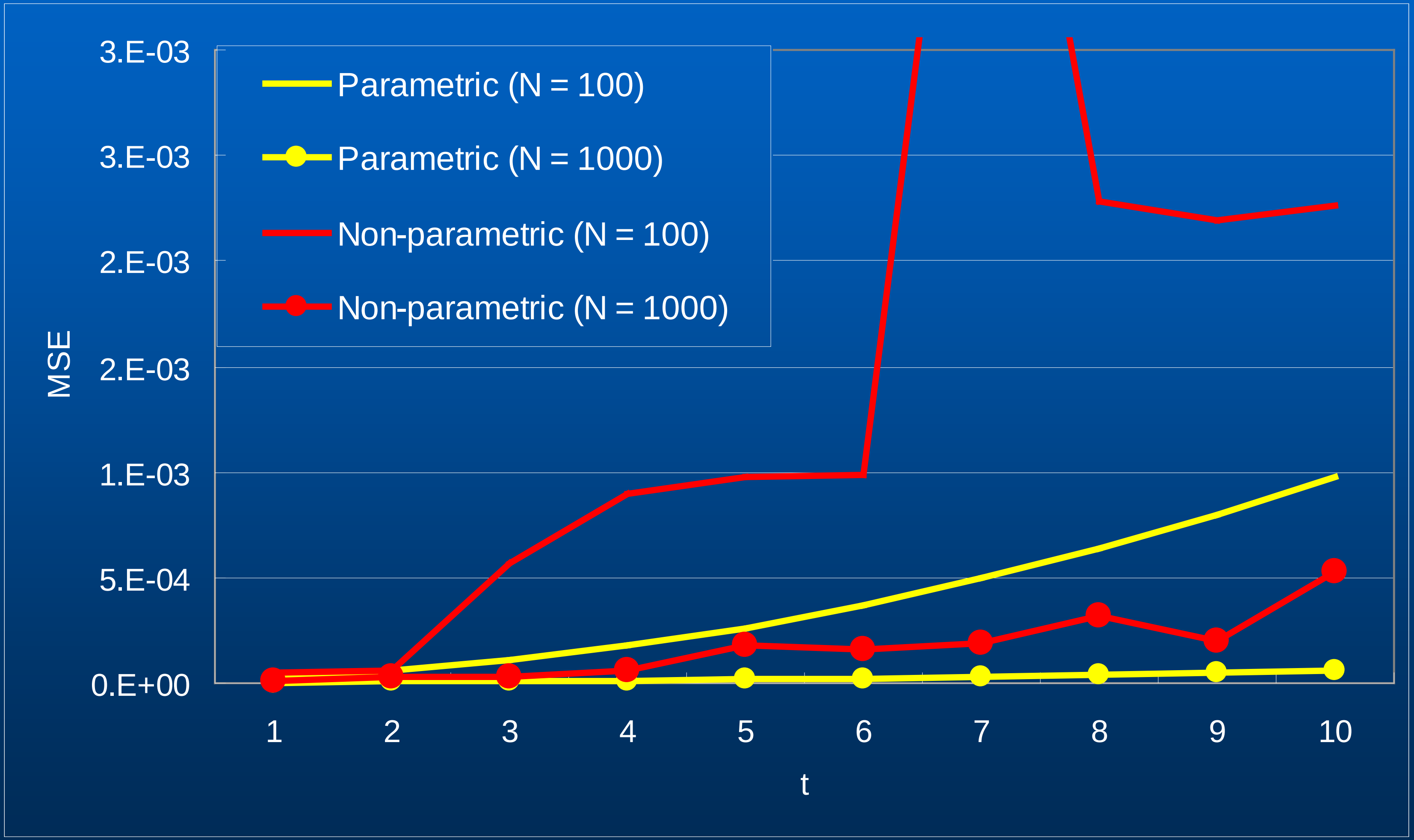
Average confidence intervals of estimated baseline hazard (N = 100)



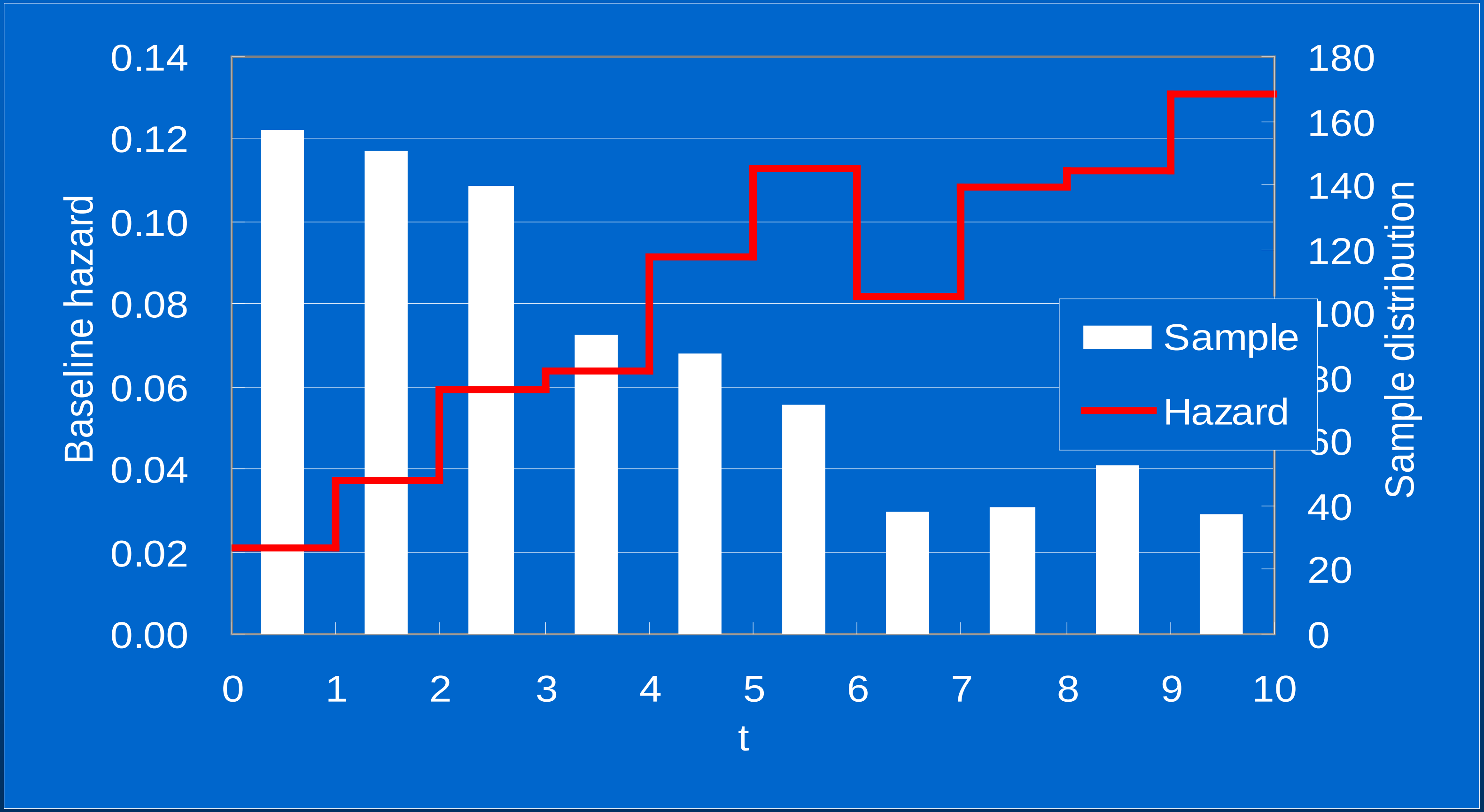
Average confidence intervals of estimated baseline hazard (N = 1000)



MSE of estimated baseline hazard



Example of the sample distribution of the duration and the estimated baseline hazard function by non-parametric hazard model (N = 1000)

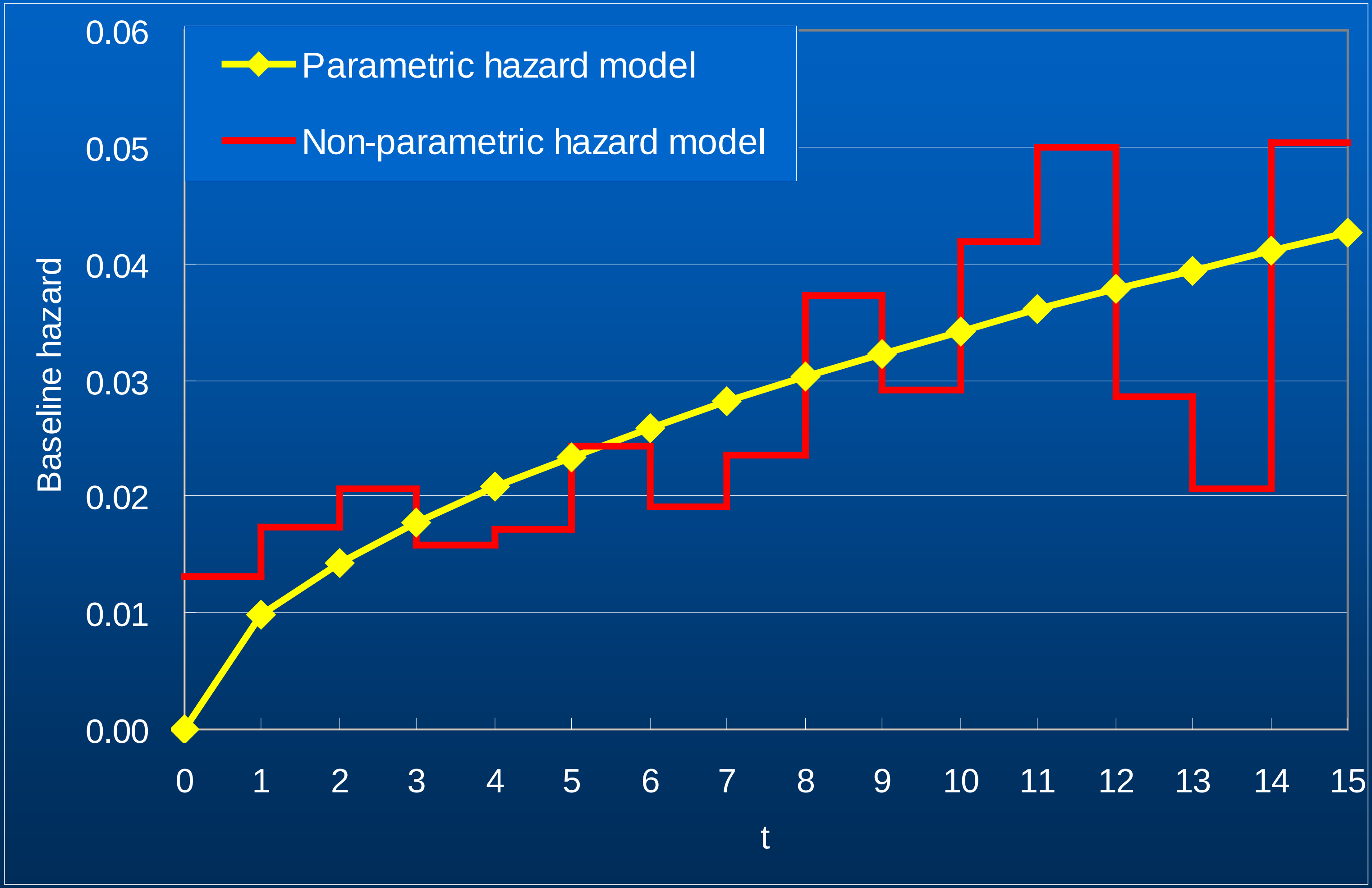


The fluctuation occurs as a result of the probabilistic nature of simulated durations.

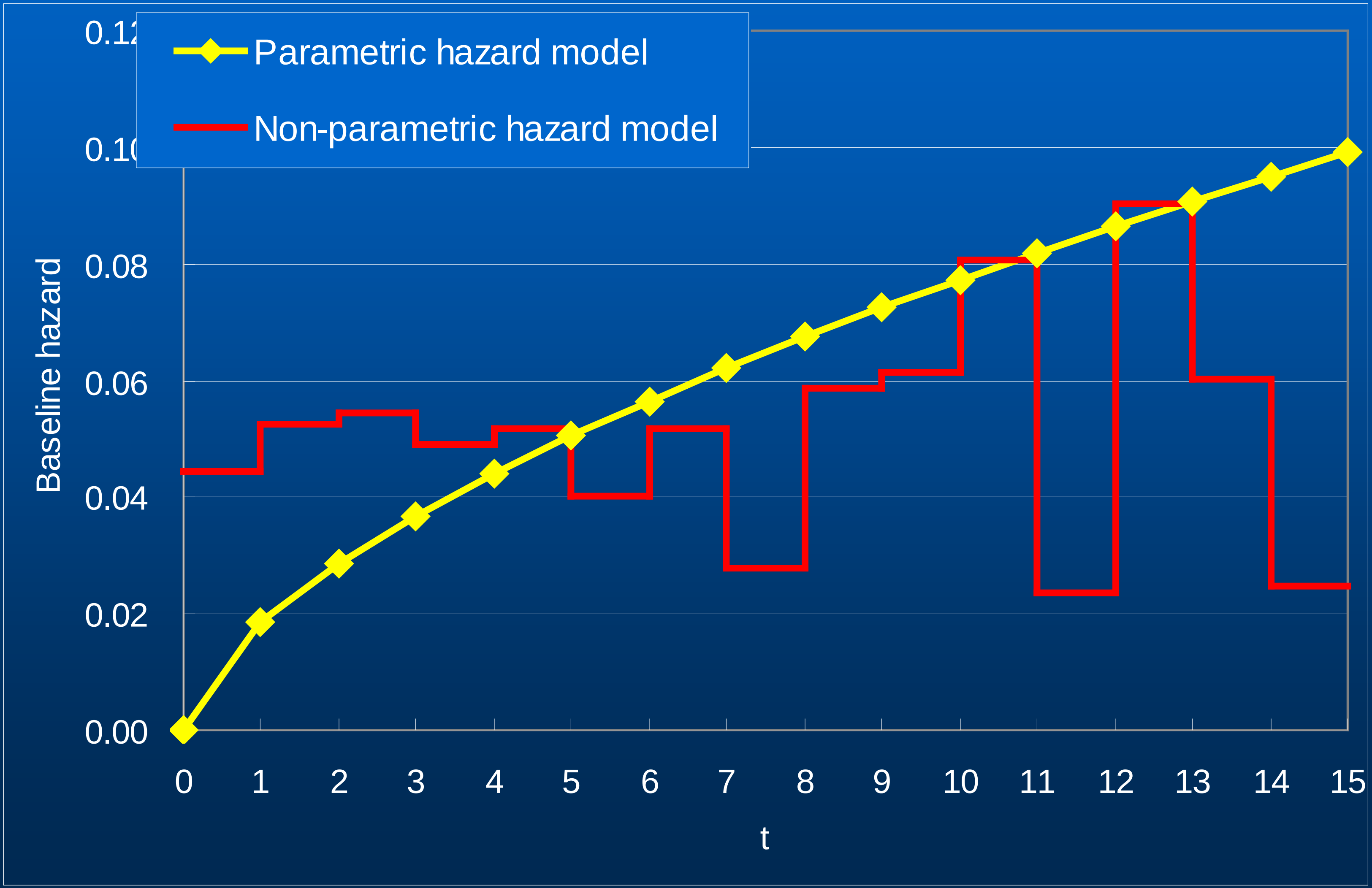
Empirical analysis

- Subset of French households' vehicle ownership panel data called Parc-Auto:
  - Collected in 1984 to 1998
  - 3,638 households and 9,988 vehicles
  - 3,811 replacements, 1,207 disposals and 1,124 acquisitions of vehicles
  - 21 covariates including 7 factors which represent policy measures and those are time-varying covariates.

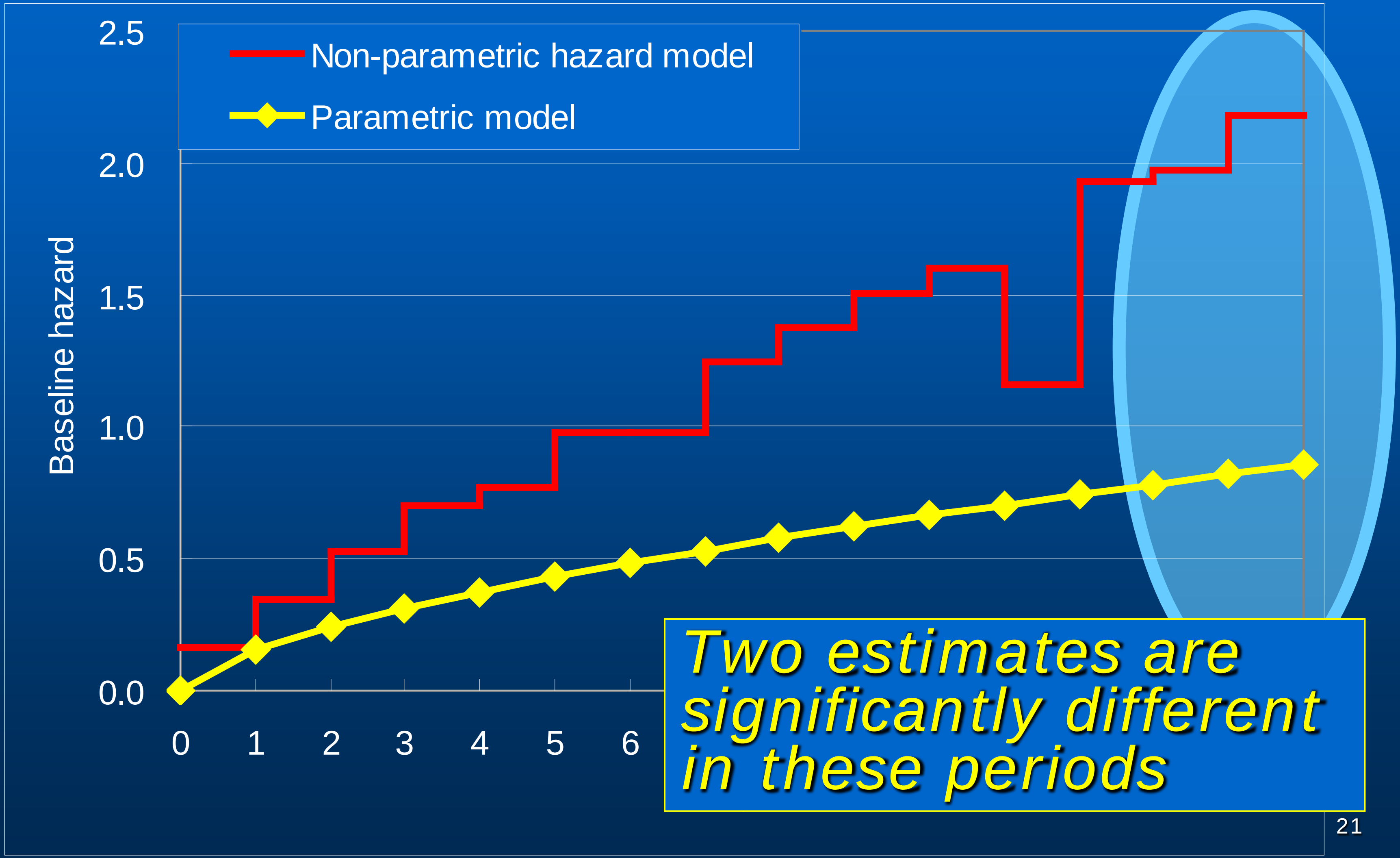
Estimated baseline hazard function of household vehicle disposal model



Estimated baseline hazard function of household vehicle acquisition model



Estimated baseline hazard function of household vehicle replacement model



Coefficient estimates of the covariates

Disposal and acquisition models

- almost all coefficient estimates of the covariates are insignificantly different.

Replacement model

- Almost all coefficient estimates of time-invariant variables are insignificantly different.
- 6 of the 7 factors representing policy measures, which are time-varying, have significantly different coefficient estimates for replacement model.

The last point implies, with the result of the estimates of the baseline hazard, the difficulty in disentangling the two.

Conclusions

- The efficiency of parametric and non-parametric hazard models in estimating the baseline hazard is examined.
- The results suggest that the estimation efficiency is not different regardless of the sample size for time-invariant covariates,
- but is lower with the non-parametric model for time-varying covariates.

Conclusions

- The baseline hazard estimate by the non-parametric model is considerably lower.
- "Over-fitting" tends to occur with the non-parametric model.
- The empirical analysis yielded similar results, although the true distribution is unknown.