MULTI-STATE PROCESSES WITH DURATION-DEPENDENT TRANSITION INTENSITIES: STATISTICAL METHODS AND APPLICATIONS

by

Lihui Zhao

B.Sc., Nankai University, China, 2001

M.Sc., Nankai University, China, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the School
of
Statistics and Actuarial Science

SIMON FRASER UNIVERSITY
Fall 2009

APPROVAL

Name: Lihui Zhao

Degree: Doctor of Philosophy

Title of Thesis: Multi-state Processes with Duration-dependent Transition Intensities: Statistical Methods and Applications

Examining Committee: Dr. Derek Bingham
Chair

____
Dr. X. Joan Hu, Senior Supervisor

____
Dr. Charmaine Dean, Supervisor

____
Dr. Richard Lockhart, Supervisor

____
Dr. Carl Schwarz, Supervisor

____
Dr. Jiguo Cao, SFU Examiner

____
Dr. John Braun, External Examiner,
University of Western Ontario

Date Approved: Sep 16, 2009

Declaration of Partial Copyright Licence

Abstract

Multi-state processes provide a convenient framework for analysis of event history data, which arise in many fields including public health, biomedical and health services research, reliability, business, and social sciences. This thesis develops methods for statistical analyses with various Markov processes in particular, and presents applications of the methodology.

Starting with the homogeneous semi-Markov (HSM) process, a generalization of the classical homogeneous Markov processes, we propose an alternative estimation procedure with right-censored data to the existing approaches to avoid their possible inconsistency in estimating the transition probabilities. Two simulation based algorithms are implemented to construct confidence bands for the HSM kernel and the sojourn time distributions. The modulated semi-Markov (MSM) process extends the HSM process to a Cox regression setting, allowing for general time-dependent covariates but invalidating the usual martingale methods to derive asymptotics. We consider estimation of the regression parameters in the MSM model and establish the consistency, asymptotic normality and efficiency of the estimators, applying the modern empirical process theory. As a further generalization, the nonhomogeneous semi-Markov (NHSM) process assumes its transition intensity involving two time scales, the individual study time since the onset of the process and the duration time in the current state. We provide estimation procedures for the parameters in four model specifications with the NHSM process. The last topic of the thesis is to deal with dependent censoring in event history data analysis. We focus on a particular informative censoring scheme with the observation of a NHSM process, and adapt a copula-based approach for dependent competing risks. Finite sample properties of all the proposed methods are examined via simulation. In addition, with the proposed methods, we conduct analyses of two real data sets, the human sleep data presented in Kneib and Hennerfeind (2008) and the hospitalization data collected by the CAYACS

program (PI: M. McBride) with BC Cancer Centre.

Keywords: Estimation, Event History Information, Likelihood, Markov Process, Right-Censoring

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Professor X. Joan Hu, for her guidance and patience throughout my study at SFU, particularly during the time of conducting this thesis project. This thesis would not have been completed without her.

My special thanks go to my thesis committee members, Professors John Braun, Jiguo Cao, Charmaine Dean, Richard Lockhart, and Carl Schwarz, for their insightful comments and suggestions about the thesis.

I am grateful to the Canadian Cancer Society Research Institute (formerly National Cancer Institute of Canada) and the Canadian Institute of Health Research for the fundings and the BC Cancer Registry, the BC Ministry of Health, the BC Cancer Agency and the BC Children's Hospital Health Records for the databases.

I want to voice my appreciation to the faculty and staff of the Department of Statistics and Actuarial Science for providing such a great place to study. I also thank my fellow graduate students for their support and friendship. Particular mention goes to: Yunfeng Dai, Celes Ying, Chunfang Lin, Wilson Lu, Zhijian Chen, Xin Feng, Na Lei, Lucy Liu, Suli Ma, Simon Bonner, Jean Shin, Elizabeth Juarez-Colunga, Carolyn Huston, Darby Thompson, and Saman Muthukumarana.

Finally, I thank my family for their support, understanding, and love. I would like to particularly thank my mother and my mother-in-law for taking care of my children when mostly needed.

Contents

Approval ii

Abstract iii

Acknowledgments v

Contents vi

List of Tables x

List of Figures xii

Notation Index xvi

1 Introduction **1**

1.1 Introduction 1

1.2 General Formulation 2

1.2.1 Multi-state Processes 2

1.2.2 Sojourn Time Formulation 5

1.2.3 Counting Process Formulation 8

1.2.4 Processes Resulting from Right-Censoring 8

1.2.5 Likelihood Function 10

1.3 Outline of Thesis 10

2 Homogeneous Semi-Markov Process **12**

2.1 Introduction 12

2.2 Confidence Bands for Semi-Markov Kernel 13

2.2.1 Bootstrap Approach 15

2.2.2 Lin-Wei-Ying's Resampling Approach 16

2.3 Robust Inference Procedure for Transition Probabilities and Sojourn Time Distributions . . . 16
2.3.1 Existing Estimation Procedures and Their Limitations . . . 17
2.3.2 Robust Estimation Procedure . . . 18
2.3.2.1 Preliminaries . . . 18
2.3.2.2 Robust Inference for Transition Probabilities . . . 20
2.3.2.3 Robust Inference for Sojourn Time Distributions . . . 21
2.4 Simulation . . . 22
2.4.1 Settings . . . 22
2.4.2 Semi-Markov Kernel Estimation . . . 23
2.4.3 Transition Probabilities and Sojourn Time Distributions . . . 30
2.5 Summary . . . 41

3 Modulated Semi-Markov Process **42**
3.1 Introduction . . . 42
3.2 Models and Estimation Procedures . . . 43
3.3 Asymptotic Properties . . . 47
3.3.1 Preliminaries . . . 47
3.3.2 Consistency and Asymptotic Normality . . . 49
3.3.3 Asymptotic Efficiency . . . 57
3.4 Simulation . . . 60
3.5 Summary . . . 61

4 Nonhomogeneous Semi-Markov Process **64**
4.1 Introduction . . . 64
4.2 Piecewise Constant Approach . . . 65
4.3 Nonparametric Estimation . . . 67
4.3.1 Estimation Procedure . . . 67
4.3.2 Asymptotic Properties . . . 69
4.4 Nonparametric Multiplicative Model . . . 75
4.4.1 Model . . . 75
4.4.2 Algorithm . . . 76
4.5 Semiparametric Estimation . . . 77
4.6 Simulation . . . 78
4.7 Extensions . . . 96

5 Semi-Markov Process with Informative Censoring **97**
5.1 Introduction 97
5.2 Modeling Informative Censoring 98
5.2.1 Assumptions 98
5.2.2 Copula Models 101
5.3 Estimation Procedure 102
5.4 Simulation 104
5.5 Remarks 110

6 Applications **111**
6.1 Human Sleep Data 111
6.1.1 Description 111
6.1.2 Analyses of Human Sleep Data 112
6.1.2.1 Analysis with Homogeneous Semi-Markov Model . . 112
6.1.2.2 Analysis with Modulated Semi-Markov Model 114
6.1.2.3 Analysis with Nonhomogeneous Semi-Markov Model 114
6.1.2.4 Model Checking 116
6.1.3 Concluding Remarks 116
6.2 Hospitalization Data 130
6.2.1 Description 130
6.2.2 Analyses of Hospitalization Data 131
6.2.2.1 Analysis with Homogeneous Semi-Markov Model . . 131
6.2.2.2 Analysis with Modulated Semi-Markov Model 132
6.2.2.3 Analysis with Nonhomogeneous Semi-Markov Model 133
6.2.3 Sensitivity Analysis for Informative Censoring 134
6.2.4 Concluding Remarks 134

7 Discussion **146**
7.1 Summary 146
7.2 Further Investigations 148
7.2.1 Interval Transition Probabilities 148
7.2.2 Goodness-of-fit Tests 149
7.2.3 General Modulated Semi-Markov Models 150
7.2.4 Alternative Observation Schemes 151
7.2.5 Other Further Investigations 151

7.2.5.1 Robustness of the Semiparametric Approach 151
7.2.5.2 Nonparametric Additive Model 152
7.2.5.3 Bandwidth Selection 152
7.2.5.4 Estimation of Marginal Quantities 152

Bibliography **153**

List of Tables

2.1 Empirical coverage probabilities of the 95% confidence intervals for the semi-Markov kernel in simulation setting 2.1 24

2.2 Empirical coverage probabilities of the 95% confidence intervals for the semi-Markov kernel in simulation setting 2.2 25

2.3 Empirical coverage probabilities of the 95% confidence intervals for the semi-Markov kernel in simulation setting 2.3 26

2.4 Empirical coverage probabilities of the 95% confidence bands for the semi-Markov kernel in simulation setting 2.1 27

2.5 Empirical coverage probabilities of the 95% confidence bands for the semi-Markov kernel in simulation setting 2.2 28

2.6 Empirical coverage probabilities of the 95% confidence bands for the semi-Markov kernel in simulation setting 2.3 29

2.7 Sample mean, bias, and standard deviation of the estimated transition probabilities of the embedded Markov chain in simulation setting 2.1 32

2.8 Sample mean, bias, and standard deviation of the estimated transition probabilities of the embedded Markov chain in simulation setting 2.2 33

2.9 Sample mean, bias, and standard deviation of the estimated transition probabilities of the embedded Markov chain in simulation setting 2.3 34

2.10 Empirical coverage probabilities and sample mean lengths of the 95% confidence intervals for the transition probabilities of the embedded Markov chain in simulation setting 2.1 35

2.11 Empirical coverage probabilities and sample mean lengths of the 95% confidence intervals for the transition probabilities of the embedded Markov chain in simulation setting 2.2 36

2.12 Empirical coverage probabilities and sample mean lengths of the 95% confidence intervals for the transition probabilities of the embedded Markov chain in simulation setting 2.3 37

2.13 Empirical coverage probabilities of the 90% and 95% confidence bands for the attainable sojourn time distributions in simulation setting 2.1 38
2.14 Empirical coverage probabilities of the 90% and 95% confidence bands for the attainable sojourn time distributions in simulation setting 2.2 39
2.15 Empirical coverage probabilities of the 90% and 95% confidence bands for the attainable sojourn time distributions in simulation setting 2.3 40

3.1 Estimated regression parameters in simulation setting 3.1 62
3.2 Estimated regression parameters in simulation setting 3.2 63

4.1 Some commonly used kernel functions 66

6.1 Frequency of the observed transitions, and median and mean of the sojourn times 112
6.2 Estimates and 95% confidence intervals of the transition probabilities for the human sleep data (1: Awake, 2: Non-REM, 3: REM) 113
6.3 Point estimate, standard error and confidence interval of the regression parameters for the human sleep data 114
6.4 Frequency of the number of hospital admissions 130
6.5 Frequency of the observed transitions, and median and mean of the sojourn times (1: out of hospital, 2: in hospital, 3: dead) 130
6.6 Estimates and 95% confidence intervals of the transition probabilities for the hospitalization data (1: out of hospital, 2: in hospital, 3: dead) 131
6.7 Point estimate, standard error and confidence interval of the regression parameters for the hospitalization data (1: out of hospital, 2: in hospital, 3: dead) 133

List of Figures

1.1 A two-state model for survival process 2
1.2 A three-state model for the hospitalization process 3
1.3 A three-state model for the human sleep process 3

4.1 Truth and sample means of estimated transition rate functions from state 1 to state 2 in simulation setting 4.1 80
4.2 Truth and sample means of estimated transition rate functions from state 1 to state 3 in simulation setting 4.1 81
4.3 Truth and sample means of estimated transition rate functions from state 2 to state 1 in simulation setting 4.1 82
4.4 Truth and sample means of estimated transition rate functions from state 2 to state 3 in simulation setting 4.1 83
4.5 Truth and sample mean of estimated transition rate functions for fixed t in simulation setting 4.1 84
4.6 Truth and sample mean of estimated transition rate functions for fixed τ in simulation setting 4.1 85
4.7 Truth and sample mean of estimated semi-Markov kernel for fixed t in simulation setting 4.1 86
4.8 Truth and sample mean of estimated semi-Markov kernel for fixed τ in simulation setting 4.1 87
4.9 Truth and sample means of estimated transition rate functions from state 1 to state 2 in simulation setting 4.2 88
4.10 Truth and sample means of estimated transition rate functions from state 1 to state 3 in simulation setting 4.2 89
4.11 Truth and sample means of estimated transition rate functions from state 2 to state 1 in simulation setting 4.2 90

4.12 Truth and sample means of estimated transition rate functions from state 2 to state 3 in simulation setting 4.2 91
4.13 Truth and sample mean of estimated transition rate functions for fixed t in simulation setting 4.2 92
4.14 Truth and sample mean of estimated transition rate functions for fixed τ in simulation setting 4.2 93
4.15 Truth and sample mean of estimated semi-Markov kernel for fixed t in simulation setting 4.2 94
4.16 Truth and sample mean of estimated semi-Markov kernel for fixed τ in simulation setting 4.2 95

5.1 Truth and sample mean of estimated semi-Markov kernels in simulation setting 5.1 106
5.2 Sample standard deviation of estimated semi-Markov kernels in simulation setting 5.1 107
5.3 Truth and sample mean of estimated semi-Markov kernels in simulation setting 5.2 108
5.4 Sample standard deviation of estimated semi-Markov kernels in simulation setting 5.2 109

6.1 Estimated transition probabilities for human sleep process 113
6.2 Realizations of two individual sleep processes and corresponding nocturnal cortisol secretion 117
6.3 Estimated semi-Markov kernel for human sleep data 118
6.4 Estimated distribution of sojourn times for human sleep data 119
6.5 Estimated transition rate functions from Awake to Non-REM 120
6.6 Estimated transition rate functions from Awake to REM 121
6.7 Estimated transition rate functions from Non-REM to Awake 122
6.8 Estimated transition rate functions from Non-REM to REM 123
6.9 Estimated transition rate functions from REM to Awake 124
6.10 Estimated transition rate functions from REM to Non-REM 125
6.11 Estimated semi-Markov kernel for fixed t 126
6.12 Estimated semi-Markov kernel for fixed τ 127

6.13 The positive parts of the difference between the estimated cumulative transition rate functions associated with the nonparametric model and the nonparametric multiplicative model 128
6.14 The negative parts of the difference between the estimated cumulative transition rate functions associated with the nonparametric model and the nonparametric multiplicative model 129
6.15 Estimated transition probabilities for the hospitalization process . . . 131
6.16 Estimated sojourn time distributions for hospitalization data based on normalized estimate 135
6.17 Estimated semi-Markov kernel for hospitalization data 136
6.18 Estimated distribution of sojourn times for hospitalization data . . . 137
6.19 Estimated transition rate functions from "out of hospital" to "in hospital" 138
6.20 Estimated transition rate functions from "out of hospital" to "death" 139
6.21 Estimated transition rate functions from "in hospital" to "out of hospital" 140
6.22 Estimated transition rate functions from "in hospital" to "death" . . 141
6.23 Estimated semi-Markov kernel for fixed t 142
6.24 Estimated semi-Markov kernel for fixed τ 143
6.25 Estimated semi-Markov kernel of the two-state HSM process 144
6.26 Estimated semi-Markov kernel of the two-state NHSM process 145

Notation Index

Notation	*Description*	*Page*
$\mathcal{S}(\cdot)$	a multi-state process	2
$\mathcal{E}$	state space of a multi-state process	2
$X(t-)$	$\lim_{s\uparrow t} X(s)$	2
$\rho_{hj}(t)$	transition intensity of a multi-state process	2
$B(t)$	time since the last transition time before t	4, 8
J_m	the mth state visited by a multi-state process	5
T_m	the mth transition time of a multi-state process	5
X_m	the mth sojourn time of a multi-state process	5
$Q_{hj}(\tau)$, $Q_{hj}(\tau;t)$	semi-Markov kernel of a semi-Markov process	5
$F_{hj}(\tau)$, $F_{hj}(\tau;t)$	distribution function of the sojourn time in state h that starts at study time t and finishes at state j	6
$H_h(\tau)$, $H_h(\tau;t)$	distribution function of the sojourn time in state h that starts at study time t	7
$S_h(\tau)$, $S_h(\tau;t)$	survival function of the sojourn time in state h that starts at study time t	7
P_{hj}, $P_{hj}(t)$	transition probability of the embedded Markov chain of a semi-Markov process	6
$\alpha_{hj}(\tau)$, $\alpha_{hj}(\tau;t)$	cause specific hazard function of a semi-Markov process	6
$\tilde{N}^{hj}(\cdot)$	counting process on the study time scale, 8	8
$N^{hj}(\cdot)$, $N^{hj}(\cdot;m)$	counting process on the duration time scale	9, 44
$\tilde{Y}^h(\cdot)$	'at risk' indicator on the study time scale, 8	9
$Y^h(\cdot)$, $Y^h(\cdot;m)$	'at risk' indicator on the duration time scale	9, 44
$\Delta W(t)$	$W(t) - W(t-)$, 13, 103	
$z^{\otimes l}$	for a vector z, $z^{\otimes 0} = 1$, $z^{\otimes 1} = z$, $z^{\otimes 2} = zz'$, where z' is the transpose of z	45

Chapter 1

Introduction

1.1 Introduction

Event history data arise in studies where a collection of individuals are followed over time, and information on the types of certain events and the times of occurrence is collected. Classical survival analysis focuses on the time to the occurrence of a single event, and can be too simplistic when multiple events are of interest. Multi-state processes provide a convenient framework for event history data analysis (Andersen *et al.*, 1993; Commenges, 1999; Andersen and Keiding, 2002).

Multi-state models are often specified in terms of transition intensities, which may involve two time scales: the *(individual) study time* since the origin of the process and the *duration time* in the current state. Classical Markov models, in which the transition intensities depend on the history only through the current state and the study time since the origin of the process, have been widely used due to the simplicity of the model interpretation and the ease of computation (Andersen *et al.*, 1993). However, because of their memoryless property, the classical Markov models can not deal with duration dependence. They have been found inadequate in many practical applications (Andersen *et al.*, 2000; Kang and Lagakos, 2007).

The literature of inferences with duration-dependent multi-state models for event history analysis is still lacking. This thesis attempts to fill in the gap to some extent. Perhaps the simplest duration-dependent multi-state model is the homogeneous semi-Markov model (Lagakos *et al.*, 1978; Gill, 1980). It assumes that the transition intensities depend on the history through the current state and the duration time in the current state. We develop methods with the homogeneous semi-Markov model and its two generalizations. One generalization incorporates time-dependent covariates through a Cox regression form, and the other allows the dependence of transition intensities on both the duration and the study time scales. In addition, we propose an approach to handling a particular type of informative censoring.

1.2 General Formulation

1.2.1 Multi-state Processes

A *multi-state process* $\mathcal{S}(\cdot) = \{\mathcal{S}(t) : t \geq 0\}$ is a stochastic process with right continuous sample paths which takes values in a finite state space, say, $\mathcal{E} = \{1, 2, \ldots, r\}$ with $r < \infty$. With respect to the *history* of the process, $\{\mathcal{F}_t : t \geq 0\}$, where $\mathcal{F}_t$ is the σ-algebra generated by $\{\mathcal{S}(u) : 0 \leq u < t\}$, the *transition probabilities* are defined as

$$P_{hj}(s, t; \mathcal{F}_s) = P\{\mathcal{S}(t) = j | \mathcal{S}(s) = h, \mathcal{F}_s\}$$

for $h, j \in \mathcal{E}$ and $s \leq t$. Denote $\mathcal{S}(t-)$ as $\lim_{s \uparrow t} \mathcal{S}(s)$. The *transition intensities* are defined by

$$\rho_{hj}(t; \mathcal{F}_t) = \lim_{\Delta t \downarrow 0} \frac{P\{\mathcal{S}(t + \Delta t) = j | \mathcal{S}(t-) = h, \mathcal{F}_t\}}{\Delta t}$$

for $h \neq j \in \mathcal{E}$, which we assume exist. A state $h \in \mathcal{E}$ is *absorbing* if $\rho_{hj}(t; \mathcal{F}_t) = 0$ for all $t \geq 0$ and $j \in \mathcal{E}$ with $j \neq h$. No further transitions can occur from an absorbing state.

Multi-state processes can be graphically illustrated by diagrams with boxes representing the states and arrows among the boxes representing the possible transitions. For example, the survival process can be viewed as a two-state process with one transient state "alive" and one absorbing state "dead", as shown in Figure 1.1.

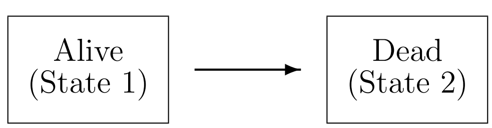

Figure 1.1: A two-state model for survival process

Two practical examples that we consider throughout this thesis are as follows.

Example 1.1 (Hospitalization process)**.** A medical study collected the hospitalization information during 1986-2000 for 1374 over five-year cancer survivors who were diagnosed in British Columbia between the ages of 0 to 19 years between 1981 and 1995. Details about this study can be found in Ying (2006) and Hu *et al.* (2008). The study's primary goal was to assess long-term resource needs of the childhood cancer survivors and to develop strategies to improve access and effectiveness of medical care. As shown in Figure 1.2, we formulate the hospitalization process into a multi-state process with 3 states: out of hospital, in hospital, and dead.

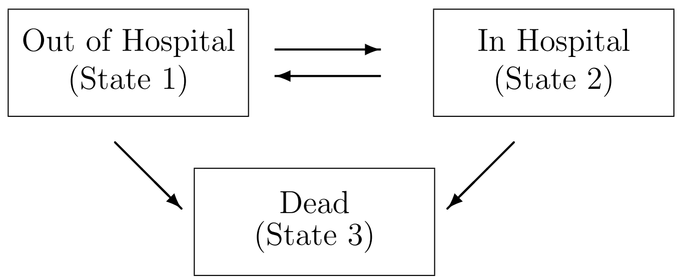

Figure 1.2: A three-state model for the hospitalization process

Example 1.2 (Human sleep process)**.** Kneib and Hennerfeind (2008) analyze the human sleep data collected at the Max-Planck Institute for Psychiatry in Munich, Germany. The study's major goal was to obtain a valid description of the dynamics underlying the sleep process of the 70 participants. The sleep process of each participant was monitored for one whole night, and was recorded by electroencephalographic (EEG) measurements which were afterwards classified into three states: Awake, REM (rapid eye movement), and Non-REM. The three-state model we consider for the data is shown in Figure 1.3. In addition to EEG measures taken every 30 seconds, the nocturnal cortisol secretion was measured approximately every 10 minutes for each participant. It was also of interest to investigate whether the level of cortisol affects the transition intensities.

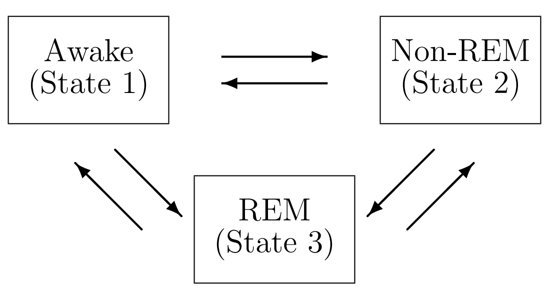

Figure 1.3: A three-state model for the human sleep process

Except for event history with a simple structure such as the survival process, the transition intensities may depend on the history of the process in complex ways. To be feasible, it is often assumed that only part of the history information is relevant to the future evolution of the process. This partial history information often includes, for example, the current state, the study time since the origin of the process, the duration time in the current state, and the total number of transitions occurred.

The classical Markov models are widely used because of the simplicity of model interpretation and the ease of computation. The *nonhomogeneous Markov model* assumes that, given the current state and the study time, the coming transition is independent of the rest of the history. That is, $\rho_{hj}(t; \mathcal{F}_t) \equiv \rho_{hj}(t)$ for all $h, j \in \mathcal{E}$ and $t \geq 0$. The *homogeneous Markov model* further assumes $\rho_{hj}(t) \equiv \rho_{hj}$ for all $h, j \in \mathcal{E}$ and $t \geq 0$. For a nonhomogeneous Markov process, its transition probabilities and transition intensities are linked by the Kolmogorov backward and forward differential equations. By solving these equations, the transition probabilities can be expressed as a function of transition intensities in the form of product integral as a generalization of the Kaplan-Meier estimator for survival function (cf. Andersen *et al.*, 1993).

In many applications, however, the Markov assumption is not plausible, for the transition intensities may depend on the duration time in the current state. In the hospitalization study, for instance, the intensity of the transition from "in hospital" to "out of hospital" likely depends on the duration of the current hospitalization. The *homogeneous semi-Markov (HSM) model* assumes that the transition intensities depend on the history through the current state and the duration time in the current state. That is, $\rho_{hj}(t; \mathcal{F}_t) \equiv \rho_{hj}(B(t))$ for all $h, j \in \mathcal{E}$ and $t \geq 0$, where $B(t)$ is the gap time between time t and time of the last transition before t, which is known as the left continuous version of the *backward recurrence time*. Note that we have slightly abused the notation: the functions ρ_{hj} in the above equation are different on the left and the right hand sides.

Nonhomogeneous Markov models and HSM models are comparable in terms of flexibility. To choose between the two models depends on which time scale (study or duration) is more important in a given application. However, both time scales can be important in some applications. For example, in the hospitalization study, in addition to the duration time in the current state, the intensity of the transition from "in hospital" to "out of hospital" can also depend on the study time, i.e., the total time since a subject entered the study. The *nonhomogeneous semi-Markov (NHSM)* *model* allows for the dependence of the transition intensities on both of the study and duration time scales. It assumes that $\rho_{hj}(t;\mathcal{F}_t) \equiv \rho_{hj}(t, B(t))$ for all $h, j \in \mathcal{E}$ and $t \geq 0$, again with a slight abuse of notation. Note that the NHSM model includes both the nonhomogeneous Markov model and the HSM model as special cases.

1.2.2 Sojourn Time Formulation

The duration time plays an important role in both homogeneous and nonhomogeneous semi-Markov processes. To better address the duration dependence, people characterize a multi-state process $\mathcal{S}(\cdot)$ as a two-dimensional process $(\mathbf{J}, \mathbf{T}) = \{(J_m, T_m) : m = 0, 1, \ldots\}$, where the sequence $\{J_0, J_1, \ldots\}$ gives the consecutive states visited by the process, and the sequence $\{T_0, T_1, \ldots\}$ is the set of corresponding transition times. Let $X_m = T_m - T_{m-1}$ be the sojourn time of the mth transition.

Definition 1.2.1. The two-dimensional process $(\mathbf{J}, \mathbf{T})$ is called a *nonhomogeneous Markov renewal (NHMR) process* if it satisfies

$$\begin{aligned} &P\{J_{m+1} = j, X_{m+1} \leq \tau \big| (J_m, T_m) = (h, t), \ldots, (J_0, T_0)\} \\ &= P\{J_{m+1} = j, X_{m+1} \leq \tau \big| (J_m, T_m) = (h, t)\} \\ &\stackrel{def}{=} Q_{hj}(\tau; t) \end{aligned} \tag{1.2.1}$$

for $h, j \in \mathcal{E}$, $t, \tau \geq 0$, and $m \in \mathbb{N}$, the set of all nonnegative integers. Here $\mathbf{Q} = \{Q_{hj}(\tau; t) : h, j \in \mathcal{E}, t, \tau \geq 0\}$ is called the *nonhomogeneous semi-Markov (NHSM) kernel* of the process. If the NHSM kernel does not depend on t, i.e., $Q_{hj}(\tau; t) \equiv Q_{hj}(\tau)$, $(\mathbf{J}, \mathbf{T})$ is a *homogeneous Markov renewal (HMR) process*. Correspondingly, $\mathbf{Q} = \{Q_{hj}(\tau) : h, j \in \mathcal{E}, \tau \geq 0\}$ is the *homogeneous semi-Markov (HSM) kernel*.

Remark 1.2.1. The NHSM kernel $\mathbf{Q}$ is a set of quantities of interest in both of the study and duration time scales. It, together with the initial law of the process, completely determines the stochastic behavior of a NHMR process.

Remark 1.2.2. In this thesis, unless stated otherwise, we assume that the conditional probability in (1.2.1) is free of m, the total number of past transitions. If this assumption is questionable, we may stratify the data based on m, and conduct analyses with the proposed methods for each stratum separately.

Remark 1.2.3. The usual survival process shown in Figure 1.1 corresponds to $(\mathbf{J}, \mathbf{T}) = \{(J_m, T_m) : m = 0, 1\}$ with $\mathcal{E} = \{1(alive), 2(dead)\}$, state 2 as an absorbing state, and $T_1 = T$, the survival time. Thus it is a HMR process, and the only unknown in the semi-Markov kernel is $Q_{12}(\tau; 0)$, which is the cumulative distribution function of the survival time.

Proposition 1.2.1. *$\mathcal{S}(\cdot)$ is a nonhomogeneous (homogeneous) semi-Markov process if and only if its corresponding sojourn time formulation $(\mathbf{J}, \mathbf{T})$ is a nonhomogeneous (homogeneous) Markov renewal process.*

The consecutive states of a HMR process, $\mathbf{J} = \{J_0, J_1, \ldots\}$, form a homogeneous Markov chain with transition probability matrix given by $\mathbf{P} = (P_{hj})_{r\times r}$, where

$$P_{hj} = P\{J_{m+1} = j | J_m = h\} = \lim_{\tau\to\infty} Q_{hj}(\tau), \tag{1.2.2}$$

for $h, j \in \mathcal{E}$. Note that P_{hh} is 0 if state h is transient, and 1 if state h is absorbing. $\mathbf{J}$ is called the *embedded Markov chain* of the HMR process. Given $\mathbf{J}$, the sojourn times $\mathbf{X} = \{X_1, X_2, \ldots\}$ are independent with distributions depending only on the adjoining states. The distribution function of the sojourn time in state h that finishes at state j ($\neq h$) is given by

$$F_{hj}(\tau) = P\{X_{m+1} \leq \tau \big| J_m = h, J_{m+1} = j\}. \tag{1.2.3}$$

We can show that $Q_{hj}(\tau) = P_{hj}F_{hj}(\tau)$.

More generally, given the sequence of the transition times $\mathbf{T} = \{T_0, T_1, \ldots\}$, the consecutive states of a NHMR process, $\mathbf{J} = \{J_0, J_1, \ldots\}$, form a nonhomogeneous Markov chain with the transition probability matrix $\mathbf{P}_m \equiv \mathbf{P}(T_m) = \big(P_{hj}(T_m)\big)_{r\times r}$ at the $(m+1)$th transition, where

$$P_{hj}(t) = P\{J_{m+1} = j | J_m = h, T_m = t\} = \lim_{\tau\to\infty} Q_{hj}(\tau; t) \tag{1.2.4}$$

for $h, j \in \mathcal{E}$. The distribution function of the sojourn time in state h that starts at study time t and finishes at state j ($\neq h$) is given by

$$F_{hj}(\tau; t) = P\{X_{m+1} \leq \tau \big| J_m = h, J_{m+1} = j, T_m = t\}. \tag{1.2.5}$$

Consequently, $Q_{hj}(\tau; t) = P_{hj}(t)F_{hj}(\tau; t)$.

The homogeneous cause-specific hazard function is defined as

$$\alpha_{hj}(\tau) = \lim_{\Delta\tau\downarrow 0} \frac{1}{\Delta\tau} P\{J_{m+1} = j, X_{m+1} \in [\tau, \tau + \Delta\tau) \big| J_m = h, X_{m+1} \geq \tau\} \tag{1.2.6}$$

for a HMR process with $h \neq j \in \mathcal{E}$. Correspondingly, the nonhomogeneous *cause-specific hazard* function is defined as

$$\alpha_{hj}(\tau; t) = \lim_{\Delta\tau\downarrow 0} \frac{1}{\Delta\tau} P\{J_{m+1} = j, X_{m+1} \in [\tau, \tau + \Delta\tau) \big| J_m = h, T_m = t, X_{m+1} \geq \tau\} \tag{1.2.7}$$

for a NHMR process with $h \neq j \in \mathcal{E}$. The cause-specific hazard function of a Markov renewal process and the transition intensity function of the corresponding semi-Markov process are closely related as follows. In the rest of this thesis, we will use the terms cause-specific hazard function and transition intensity function interchangeably.

Proposition 1.2.2. *With a NHMR process, we have* $\alpha_{hj}(\tau;t) = \rho_{hj}(t+\tau,\tau)$*; we have* $\alpha_{hj}(\tau) = \rho_{hj}(\tau)$ *in the homogeneous case.*

There is a one-to-one correspondence between the semi-Markov kernel and the set of the cause-specific hazard functions. Denote

$$P\{X_{m+1} \leq \tau | J_m = h, T_m = t\} = \sum_{j \neq h} Q_{hj}(\tau;t)$$

by $H_h(\tau;t)$. Note that, as a function of τ, $H_h(\tau;t)$ is the distribution function of the sojourn time in state h that starts at study time t. Let $S_h(\tau;t) = 1 - H_h(\tau;t)$ be the corresponding survival function. Then,

$$S_h(\tau;t) = \exp\{-\int_0^\tau \sum_{j \neq h} \alpha_{hj}(u;t)du\}, \tag{1.2.8}$$

and

$$Q_{hj}(\tau;t) = \int_0^\tau \alpha_{hj}(u;t) S_h(u;t) du \tag{1.2.9}$$

for all $h \neq j \in \mathcal{E}$. Thus we can estimate the semi-Markov kernel through estimating the cause-specific hazard functions of the Markov renewal process, or the transition intensity functions of the semi-Markov process.

On the other hand, provided that $Q_{hj}(\tau;t)$ is absolutely continuous with respect to Lebesgue measure as a function of τ, with the partial derivative denoted by

$$q_{hj}(\tau;t) = \frac{\partial Q_{hj}(\tau;t)}{\partial \tau},$$

for all $h \neq j \in \mathcal{E}$,

$$\alpha_{hj}(\tau;t) = \begin{cases} q_{hj}(\tau;t)/S_h(\tau;t) & \text{if } S_h(\tau;t) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

1.2.3 Counting Process Formulation

A counting process representation for multi-state processes has been shown very useful (Andersen *et al.*, 1993). Let

$$\tilde{N}_*^{hj}(t) = \#\{m \geq 1 : T_m \leq t, J_{m-1} = h, J_m = j\}$$

be the total number of $h \to j$ transitions in the time interval $(0, t]$. With respect to to the history of the process, $\mathcal{F}_t$, the multivariate counting process $\tilde{\mathbf{N}}_* = \{\tilde{N}_*^{hj}(t) : h \neq j \in \mathcal{E}, t \geq 0\}$ has intensity function $\{\lambda_*^{hj}(t) : h \neq j \in \mathcal{E}, t \geq 0\}$ with

$$\lambda_*^{hj}(t) = \tilde{Y}_*^h(t)\rho_{hj}(t; \mathcal{F}_t),$$

where $\tilde{Y}_*^h(t) = I\{\mathcal{S}(t-) = h\}$ is the 'at risk' indicator for whether the process has the potential of experiencing a transition from state h at time t, and $\rho_{hj}(t; \mathcal{F}_t)$ is the transition intensity function of the multi-state process. Let $\tilde{N}_*(t) = \sum_{h,j} \tilde{N}_*^{hj}(t)$ be the total number of transitions in $(0, t]$. The backward recurrence time $B(t)$ equals $t - T_{\tilde{N}_*(t-)}$.

If the multi-state process $\mathcal{S}(\cdot)$ is a NHSM process, the intensity function of $\tilde{N}_*^{hj}(t)$ is

$$\lambda_*^{hj}(t) = \tilde{Y}_*^h(t)\rho_{hj}(t, B(t)) = \tilde{Y}_*^h(t)\alpha_{hj}(B(t); T_{\tilde{N}_*(t-)}).$$

It is further simplified to

$$\lambda_*^{hj}(t) = \tilde{Y}_*^h(t)\rho_{hj}(B(t)) = \tilde{Y}_*^h(t)\alpha_{hj}(B(t))$$

when $\mathcal{S}(\cdot)$ is a HSM process.

1.2.4 Processes Resulting from Right-Censoring

Event history data are rarely observed completely. In this thesis, we focus on right censoring, the most common form of incomplete observation. The counting process formulation is convenient to handle this type of incomplete observation. Let $O(t) = I(C \geq t)$ be the indicator that a multi-state process is under observation at time t, where C is the censoring time. Let the counting process

$$\tilde{N}^{hj}(t) = \int_0^t O(s)d\tilde{N}_*^{hj}(s) = \tilde{N}_*^{hj}(t \wedge C) \tag{1.2.10}$$

be the number of observed $h \to j$ transitions in the time interval $(0, t]$, where $t \wedge C$ is the minimum of t and C. Unless otherwise stated, we assume that the censoring time

C is independent of the multi-state process. Then, conditional on C, the intensity function of the observed multivariate counting process $\tilde{\mathbf{N}} = \{\tilde{N}^{hj}(t) : h \neq j \in \mathcal{E}, t \geq 0\}$ is $\{\lambda^{hj}(t) : h \neq j \in \mathcal{E}, t \geq 0\}$ with

$$\lambda^{hj}(t) = \tilde{Y}^h(t)\rho_{hj}(t; \mathcal{F}_t),$$

where

$$\tilde{Y}^h(t) = I\{\mathcal{S}(t-) = h, C \geq t\} \tag{1.2.11}$$

is the 'at risk' indicator, indicating whether the process at time t is under observation and has the potential of experiencing a transition from state h.

In the following, we introduce another set of processes, which count the number of observed sojourn times. For each $t > 0$, define

$$N^{hj}(u; t) = \#\{m \geq 1 : J_{m-1} = h, J_m = j, X_m \leq u, T_m \leq t\} \tag{1.2.12}$$

as the number of sojourn times in state h that are less equal than u and followed by a transition to state j during the study time window $(0, t]$, and

$$Y^h(u; t) = \#\{m \geq 1 : J_{m-1} = h, X_m \geq u, T_{m-1} + u \leq t\} \tag{1.2.13}$$

as the number of sojourn times in state h that are large equal than u during the study time window $(0, t]$. Denote the corresponding resulting processes due to right-censoring by $N^{hj}(u)$ and $Y^h(u)$, respectively. Then

$$N^{hj}(u) = N^{hj}(u; C), \tag{1.2.14}$$

and

$$Y^h(u) = Y^h(u; C). \tag{1.2.15}$$

The multivariate counting processes $\tilde{\mathbf{N}} = \{\tilde{N}^{hj}(t) : h \neq j \in \mathcal{E}, t \geq 0\}$ on the study time scale and $\mathbf{N} = \{N^{hj}(u) : h \neq j \in \mathcal{E}, u \geq 0\}$ on the duration time scale are linked as follows.

Lemma 1.2.3. *For any bounded measurable function f on $[0, \infty)$,*

$$\int_0^\infty f(u)dN^{hj}(u) = \int_0^\infty f(B(t))d\tilde{N}^{hj}(t). \tag{1.2.16}$$

Proof. Extending the arguments in Gill (1980), we can show that both sides of (1.2.16) are equal to

$$\sum_{\{m: T_{m+1} \leq C\}} f(X_{m+1})I\{J_m = h, J_{m+1} = j\}.$$

□

1.2.5 Likelihood Function

If the right-censoring is noninformative, the contribution to the log-likelihood of a censored multi-state process is given by

$$\sum_{h,j}\left[\int_0^\infty \log \lambda^{hj}(t)d\tilde{N}^{hj}(t) - \int_0^\infty \lambda^{hj}(t)dt\right], \tag{1.2.17}$$

up to some constant not related to $\{\lambda^{hj}(\cdot) : h, j \in \mathcal{E}\}$ (Andersen *et al.*, 1993; Cook and Lawless, 2007).

This thesis focuses on duration-dependent multi-state models, in which the intensity functions depend on the duration in the current state. To this end, we often use Lemma 1.2.3 to transform the log-likelihood given by (1.2.17) from the study time scale to the duration time scale.

1.3 Outline of Thesis

The rest of this thesis is organized as follows. Chapter 2 considers nonparametric estimation with incompletely observed HSM processes. We propose two simulation based algorithms to construct confidence bands for the HSM kernel. We show that the existing estimators for the transition probabilities of the embedded Markov chain and the sojourn time distributions can be biased when a right-censoring is involved. A robust estimation procedure is proposed to address the concern.

Time-dependent covariates, such as the study time variable, can be incorporated in the HSM model through the Cox regression form. The dependence of the baseline transition intensities on the duration time scale makes the model fall outside the framework of Aalen's multiplicative intensity models and invalidates the usual martingale methods. Dabrowska *et al.* (1994) consider the Cox regression in the semi-Markov model with covariates dependent on the duration time in the present state only, which excludes the study time variable. As a generalization, Chapter 3 allows general time-dependent covariates and proposes estimating equations of the regression parameters. We derive the asymptotic properties of the estimators by using empirical process theory, and show that the estimators are asymptotically efficient among regular estimators.

In some situations, the transition intensities may depend on both the study and the duration times. This naturally leads to a NHSM model. In Chapter 4, we propose several estimation procedures for the NHSM model. We start from a piecewise constant approach, where the transition intensities are assumed to be piecewise constant on the study time scale, and can vary arbitrarily on the duration time scale. We then propose a nonparametric estimation procedure based on the kernel method. We show that the nonparametric estimator is a maximum local likelihood estimator. The asymptotic properties of the estimator are then established. In addition, we consider a structured nonparametric model, assuming the transition intensities depend on the study and duration times in a multiplicative form. Finally, the semiparametric approach in Chapter 3 is adapted by incorporating the study time as a time-dependent covariate.

Informative censoring problem is challenging in event history data analysis. Many existing methods assume that the censoring is independent conditional on covariates or some latent variables. In Chapter 5, we consider a particular type of informative right censoring scheme for NHSM processes. Motivated by the competing risks formulation of HSM processes (Lagakos *et al.*, 1978), we model the informative censoring mechanism as another competing risk. Under this model assumption, the censored process becomes a new NHSM process with the censoring included as a new absorbing state of the original process. Thus the large literature of competing risks can be adapted to the setting. In particular, we adapt a copula-based modeling proposed by Zheng and Klein (1995). An advantage of the copula approach is that the marginal distributions need not to be specified, and can be estimated nonparametrically.

Chapter 6 presents analyses of the two real data sets described in Examples 1.1 and 1.2 with the proposed methods. Finally, we provide a summary of this thesis project and outline some extensions for future research in Chapter 7.

Chapter 2

Homogeneous Semi-Markov Process

2.1 Introduction

In this chapter, we consider estimation with homogeneous semi-Markov (HSM) processes, of which the observation is subject to independent right censoring. Particularly, we are interested in the semi-Markov kernel, the transition probability matrix of the embedded Markov chain formed by the consecutive states of the process, and the sojourn time distributions.

Lagakos *et al.* (1978) present the nonparametric maximum likelihood estimation for the semi-Markov kernel. Their approach allows an arbitrary number of states as well as right censored observations. Matthews (1984) and Dinse and Larson (1986) express the semi-Markov kernel as cause-specific hazard functions and show advantages of the reformulation: easier calculation and clearer interpretation. Gill (1980) applies the theory of stochastic integration and counting processes to provide a rigorous derivation of the consistency and weak convergence of the estimator of the semi-Markov kernel proposed by Lagakos *et al.* (1978). However, as pointed by Gill (1980), the asymptotic Gaussian process does not have an independent increment structure, thus it can not be transformed into the standard Brownian bridge or Brownian motion to construct confidence bands for the semi-Markov kernel. In Section 2.2, we propose two simulation based algorithms to construct confidence bands.

To estimate the transition probabilities, Lagakos *et al.* (1978) propose a plug-in estimator and its normalized version based on the nonparametric estimator of the semi-Markov kernel given in (2.3.1) and (2.3.3), which are not necessarily consistent. Phelan (1990b) estimates the transition probabilities of a class of Markov renewal processes whose semi-Markov kernel satisfies $F_{hj}(\cdot) \equiv F_{hk}(\cdot)$ for all h, j, and k. However, all the above estimators can be biased, as will be shown in Sections 2.3 and 2.4. As an alternative, we in Section 2.3 propose a robust approach to estimating the transition probabilities of the embedded Markov chain and sojourn time distributions with general right censored semi-Markov processes.

We examine finite sample performance of the proposed methods via simulation in Section 2.4. Section 2.5 concludes this chapter with a summary and motivates the next chapters.

2.2 Confidence Bands for Semi-Markov Kernel

Recall the two processes in the time scale of duration defined in (1.2.14) and (1.2.15) of Section 1.2.4:

$$N^{hj}(u) = \#\{m \geq 1 : J_{m-1} = h, J_m = j, X_m \leq u, T_m \leq C\}, \tag{2.2.1}$$

and

$$Y^h(u) = \#\{m \geq 1 : J_{m-1} = h, X_m \geq u, T_{m-1} + u \leq C\}. \tag{2.2.2}$$

Let $\{(N_i^{hj}(\cdot), Y_i^h(\cdot)) : i = 1, \ldots, n\}$ be n independent realizations of $(N^{hj}(\cdot), Y^h(\cdot))$. In what follows, we use "$\cdot$" in subscripts or superscripts to represent summation over the omitted index. For example, $N_\cdot^{hj}(u) \equiv \sum_i N_i^{hj}(u)$. Denote $\Delta W(t)$ as $W(t) - W(t-)$. The nonparametric maximum likelihood estimator of the semi-Markov kernel Q_{hj} proposed by Lagakos *et al.* (1978) can be written as

$$\hat{Q}_{hj}(\tau) = \int_0^\tau (1 - \hat{H}_h(u-)) \frac{dN_\cdot^{hj}(u)}{Y_\cdot^h(u)}, \tag{2.2.3}$$

where

$$\hat{H}_h(u) = 1 - \prod\nolimits_{v \leq u} \left(1 - \frac{\Delta N_\cdot^{h\cdot}(v)}{Y_\cdot^h(v)}\right), \quad u > 0, \tag{2.2.4}$$

is the nonparametric maximum likelihood estimator of $H_h(u)$.

According to Gill (1980), on $\{\tau : Y_\cdot^h(\tau) > 0 \text{ and } 1 - H_h(\tau-) > 0\}$, it can be shown by integration by parts that

$$
\begin{aligned}
n^{1/2}\left[\hat{Q}_{hj}(\tau)-Q_{hj}(\tau)\right] = & \int_0^{\tau}(1-\hat{H}_h(u-))\frac{n}{Y_\cdot^h(u)}\frac{dZ_\cdot^{hj}(u)}{n^{1/2}} \\
& - Q_{hj}(\tau)\int_0^{\tau-}\frac{1-\hat{H}_h(u-)}{1-H_h(u)}\frac{n}{Y_\cdot^h(u)}\frac{dZ_\cdot^{h\cdot}(u)}{n^{1/2}} \\
& + \int_0^{\tau-} Q_{hj}(u)\frac{1-\hat{H}_h(u-)}{1-H_h(u)}\frac{n}{Y_\cdot^h(u)}\frac{dZ_\cdot^{h\cdot}(u)}{n^{1/2}},
\end{aligned}
\tag{2.2.5}
$$

where

$$
Z_i^{hj}(u) = N_i^{hj}(u) - \int_0^u Y_i^h(v)\frac{dQ_{hj}(v)}{1-H_h(v-)}. \tag{2.2.6}
$$

The following lemma is from Theorems 1 and 3 of Gill (1980):

Lemma 2.2.1 (Theorems 1 and 3, Gill 1980)**.** *Assume that the semi-Markov kernel $Q_{hj}(\cdot)$ is continuous for all $h, j \in \mathcal{E}$. Let*

$$
\tau_h = \sup\left\{\tau : P\left(Y^h(\tau) > 0\right) > 0\right\}. \tag{2.2.7}
$$

Then as $n \to \infty$,

$$
\sup_{\tau\in[0,\tau_h]}\left|\hat{H}_h(\tau)-H_h(\tau)\right| \xrightarrow{P} 0,
$$

and

$$
\sup_{\tau\in[0,\tau_h]}\left|\hat{Q}_{hj}(\tau)-Q_{hj}(\tau)\right| \xrightarrow{P} 0.
$$

Furthermore, if ν_h satisfies

$$
P\left(Y^h(\nu_h) > 0\right) > 0,
$$

then $n^{1/2}\left[\hat{Q}_{hj}(\cdot)-Q_{hj}(\cdot)\right]$ converges weakly to a Gaussian process on $[0,\nu_h]$. The weak convergence is in the space $D[0,\nu_h]$ equipped with the Skorohod metric.

Remark 2.2.1. τ_h defined in (2.2.7) can be interpreted as the "largest observable" sojourn time at state h. When both the semi-Markov kernel and the censoring time distribution are continuous, $P\left(Y^h(\tau) > 0\right)$ is a continuous function of τ. In this case, $P\left(Y^h(\tau_h) > 0\right) = 0$ by the definition of τ_h. Thus Lemma 2.2.1 does not yield the asymptotic distribution of $n^{1/2}\left[\hat{Q}_{hj}(\tau_h)-Q_{hj}(\tau_h)\right]$, although $\hat{Q}_{hj}(\tau_h)$ is consistent for $Q_{hj}(\tau_h)$.

As pointed out by Gill (1980), the weak limit of $n^{1/2}\left[\hat{Q}_{hj}(\cdot)-Q_{hj}(\cdot)\right]$ does not have an independent increment structure, and thus it can not be transformed into the standard Brownian bridge or Brownian motion, which is often used to construct confidence bands for an unknown function. We propose simulation based approaches to approximate the critical values to construct confidence bands in the following.

Various types of confidence bands for $Q_{hj}(\cdot)$ can be constructed using the class of transformed processes

$$G_{hj}(\tau) = n^{1/2} g_{hj}^{(n)}(\tau)[\phi(\hat{Q}_{hj}(\tau)) - \phi(Q_{hj}(\tau))],$$

where ϕ is a known function with non-zero continuous derivative ϕ' and $g_{hj}^{(n)}(\tau)$ is a weight function. The weight $g_{hj}^{(n)}(\cdot)$ determines the shape of the bands, and has a deterministic limit $g_{hj}(\tau)$ in probability. By the functional delta-method (Andersen *et al.*, 1993), the process $G_{hj}(\tau)$ is asymptotically equivalent to

$$g_{hj}^{(n)}(\tau)\phi'(\hat{Q}_{hj}(\tau))n^{1/2}[\hat{Q}_{hj}(\tau) - Q_{hj}(\tau)].$$

We suggest to choose $\phi(x) = \log(-\log(1-x))$, which is analogous to the log-log transformation of the Kaplan-Meier estimator that has been widely used in the literature to construct a confidence band for a survival function. We consider two weight functions, one leading to the equal precision (EP) bands (Nair, 1984) and the other leading to the Hall-Wellner (HW) bands (Hall and Wellner, 1980).

The following presents two simulation based algorithms to construct confidence bands for $Q_{hj}(\cdot)$, the semi-Markov kernel. The first algorithm uses the bootstrap technique and the second algorithm adapts the resampling technique developed in Lin *et al.* (1993).

2.2.1 Bootstrap Approach

We first consider an application of the bootstrap to obtain an approximate $(1-\alpha)$ confidence band for $Q_{hj}(\cdot)$ on $[s_1, s_2] \subseteq [0, \nu_h]$. The algorithm we propose is as follows.

Step 1. Randomly select a sample $\mathcal{M}$ with size n with replacement from $\{1, \ldots, n\}$, and then estimate $Q_{hj}(\cdot)$ based on the data $\{(N_i^{hj}(\cdot), Y_i^h(\cdot)) : i \in \mathcal{M}\}$. Replicate this procedure B times to obtain estimates $\{\hat{Q}_{hj}^{(b)}(\cdot) : b = 1, \ldots, B\}$.

Step 2. For $b = 1, \ldots, B$, let $G_{hj}^{(b)}(\tau) = n^{1/2} g_{hj}^{(n)}(\tau)\phi'(\hat{Q}_{hj}(\tau))[\hat{Q}_{hj}^{(b)}(\tau) - \hat{Q}_{hj}(\tau)]$, and $q_{hj}^{(b)}(s_1, s_2) = \sup_{\tau \in [s_1, s_2]} |G_{hj}^{(b)}(\tau)|$. Finally, let $q_{hj}(s_1, s_2)$ be the $(1-\alpha)$ quantile of $\left\{ q_{hj}^{(b)}(s_1, s_2) : b = 1, \ldots, B \right\}$.

Step 3. An approximate $(1-\alpha)$ confidence band for $\phi(Q_{hj}(\tau))$ on $[s_1, s_2]$ is

$$[\phi(\hat{Q}_{hj}(\tau)) - n^{-1/2} q_{hj}(s_1, s_2)/g_{hj}^{(n)}(\tau), \phi(\hat{Q}_{hj}(\tau)) + n^{-1/2} q_{hj}(s_1, s_2)/g_{hj}^{(n)}(\tau)]$$

for $\tau \in [s_1, s_2]$, which can be converted to a confidence band for $Q_{hj}(\tau)$ on $[s_1, s_2]$.

2.2.2 Lin-Wei-Ying's Resampling Approach

We now adapt the resampling technique developed by Lin *et al.* (1993). By the uniform consistency of $\hat{H}_h(\cdot)$ and $Y_\cdot^h(\cdot)/n$, we can show that $n^{1/2}\left[\hat{Q}_{hj}(\tau) - Q_{hj}(\tau)\right]$ given in (2.2.5) has the same asymptotic distribution as the sum of a set of independent and identically distributed terms. Replace $Z_i^{hj}(u)$ in (2.2.5) with $\hat{Z}_i^{hj}(u)U_i$, where $\{U_i : i = 1, \ldots, n\}$ are independent standard normal random variables which are also independent of the data, and

$$\hat{Z}_i^{hj}(u) = N_i^{hj}(u) - \int_0^u Y_i^h(v)\frac{d\hat{Q}_{hj}(v)}{1 - \hat{H}_h(v-)} = N_i^{hj}(u) - \int_0^u Y_i^h(v)\frac{dN_\cdot^{hj}(v)}{Y_\cdot^h(v)}.$$

Also replace Q_{hj} and H_h in (2.2.5) with $\hat{Q}_{hj}$ and $\hat{H}_h$, respectively. Denote the resulting quantity by $W_{hj}^{(n)}(\tau)$. The following proposition can be shown by a slight extension of the arguments in Lin *et al.* (1993).

Proposition 2.2.2. *As $n \to \infty$ and conditional on the observed data, $\{W_{hj}^{(n)}(\tau) : h, j \in \mathcal{E}, \tau \in [0, \nu_h]\}$ has the same limiting distribution as*

$$\left\{n^{1/2}\left[\hat{Q}_{hj}(\tau) - Q_{hj}(\tau)\right] : h, j \in \mathcal{E}, \tau \in [0, \nu_h]\right\}.$$

Critical values in the construction of confidence bands for $Q_{hj}(\tau)$ can then be approximated based on simulated realizations of $W_{hj}^{(n)}(\tau)$ using different sets of the standard normal random variables. It gives an algorithm based on the realizations of $W_{hj}^{(n)}(\tau)$ to construct confidence bands for $Q_{hj}(\tau)$, similar to the bootstrap approach in Section 2.2.1.

2.3 Robust Inference Procedure for Transition Probabilities and Sojourn Time Distributions

We first review the existing estimation procedures of the transition probabilities for the embedded Markov chain and the conditional distributions of sojourn times from right censored semi-Markov processes, and point out their limitations. Then we introduce a robust approach.

2.3.1 Existing Estimation Procedures and Their Limitations

With $\{(N_i^{hj}(\cdot), Y_i^h(\cdot)) : i = 1, \ldots, n\}$, n independent realizations of $(N^{hj}(\cdot), Y^h(\cdot))$, the nonparametric maximum likelihood estimator of Q_{hj} is given in (2.2.3). Since $P_{hj} = Q_{hj}(\infty)$, Lagakos *et al.* (1978) suggest a plug-in estimator of the transition probability P_{hj},

$$\hat{P}_{hj} = \hat{Q}_{hj}(\infty), \tag{2.3.1}$$

and correspondingly estimate the sojourn time distribution with

$$\hat{F}_{hj}(\tau) = \hat{Q}_{hj}(\tau)/\hat{Q}_{hj}(\infty). \tag{2.3.2}$$

Since $\hat{Q}_{hj}(\infty)$ is the same as $\hat{Q}_{hj}(\cdot)$ evaluated at the largest observed sojourn time starting from state h, $\sum_j \hat{P}_{hj}$ can be less than 1 when the largest sojourn time from state h is censored. This is not desirable. In such cases, Lagakos *et al.* (1978) propose to use the normalized estimator,

$$\tilde{P}_{hj} = \hat{Q}_{hj}(\infty)/\sum_{k \neq h} \hat{Q}_{hk}(\infty), \tag{2.3.3}$$

and correspondingly to estimate the sojourn time distribution with

$$\tilde{F}_{hj}(\tau) = \hat{Q}_{hj}(\tau)/\tilde{P}_{hj}. \tag{2.3.4}$$

Phelan (1990b) considers estimation of the transition probabilities P_{hj} under the assumption that the distributions $F_{hj}(.)$'s are the same for all j, and thus

$$Q_{hj}(\tau) = P_{hj} F_h(\tau), \quad \forall h, j \in \mathcal{E}. \tag{2.3.5}$$

Suppose the independent and identically distributed copies $(\mathbf{J}_i, \mathbf{T}_i, C_i)$, $i = 1, \ldots, n$, are observed, where $(\mathbf{J}_i, \mathbf{T}_i) = \{(J_i^m, T_i^m) : m \geq 0\}$ and C_i is the right censoring time. For each $h, j \in \mathcal{E}$, and for subject i, let

$$\check{N}_i^{hj} = \#\{m : J_i^{m-1} = h, J_i^m = j, T_i^m \leq C_i\}$$

be the number of observed transitions that are from state h to j, and

$$\check{Y}_i^h = \#\{m : J_i^{m-1} = h, T_i^m \leq C_i\}$$

be the number of observed transitions that are started from state h. Phelan proposes an estimator of P_{hj} as

$$\check{P}_{hj} = \check{N}_{\cdot}^{hj}/\check{Y}_{\cdot}^{h}, \tag{2.3.6}$$

which is the proportion of observed $h \to j$ transitions among observed transitions starting from state h. Building upon the results in Gill (1980), Phelan establishes the consistency and asymptotically normality of $\check{P}_{hj}$ under the assumption given in (2.3.5).

However, as can be seen numerically from the simulations in the following section, the plug-in, the normalized, and the Phelan estimators only work under certain conditions. There are situations where they can be inconsistent. In particular, the plug-in estimator $\hat{P}_{hj}$ does not work well when the censoring time is small relative to the sojourn times. Both the normalized estimator $\tilde{P}_{hj}$ and the Phelan estimator $\check{P}_{hj}$ need the assumption that $F_{hj}(.) = F_h(\cdot)$ for all $j \neq h$. In fact, as we will see later, when the censoring time is small relative to the sojourn times, all the three estimators of P_{hj} can be inconsistent.

2.3.2 Robust Estimation Procedure

In this section, we propose a robust approach to estimating the transition probabilities P_{hj} and the sojourn time distributions $F_{hj}(\cdot)$. We assume that the semi-Markov kernel $Q_{hj}(\cdot)$ is continuous for all $h, j \in \mathcal{E}$.

2.3.2.1 Preliminaries

Let $\{(N_i^{hj}(\cdot), Y_i^h(\cdot)) : i = 1, \ldots, n\}$ be n independent realizations of $(N^{hj}(\cdot), Y^h(\cdot))$. Because $Y_\cdot^h(u)$ is a left continuous function of u, we have

$$\sup\left\{u : Y_\cdot^h(u) > 0\right\} = \max\left\{u : Y_\cdot^h(u) > 0\right\}, \tag{2.3.7}$$

which we denote by $V_h^{(n)}$. Note that $V_h^{(n)}$ is the largest fully or partially observed sojourn time in state h.

Theorem 2.3.1. *Let τ_h be as defined in (2.2.7). As $n \to \infty$, $V_h^{(n)} \to \tau_h$ almost surely.*

Proof. For any $\tau' < \tau_h$, we have $P\left(Y^h(\tau') > 0\right) > 0$, by the definition of τ_h. Thus

$$\begin{aligned} P\left(Y_\cdot^h(\tau') = 0\right) &= \prod_{i=1}^{n} P\left(Y_i^h(\tau') = 0\right) \\ &= \prod_{i=1}^{n} \left[1 - P\left(Y_i^h(\tau') > 0\right)\right] \\ &= \left[1 - P\left(Y^h(\tau') > 0\right)\right]^n \\ &\longrightarrow 0, \quad \text{as } n \to \infty. \end{aligned}$$

So

$$P\left(V_h^{(n)} \geq \tau'\right) = P\left(Y_\cdot^h(\tau') > 0\right) \to 1 \tag{2.3.8}$$

as $n \longrightarrow \infty$.

On the other hand, for any $\tau'' > \tau_h$, we have $P\left(Y^h(\tau'') = 0\right) = 1$ by the definition of τ_h. Thus

$$P\left(Y_\cdot^h(\tau'') = 0\right) = \prod_{i=1}^{n} P\left(Y_i^h(\tau'') = 0\right) = 1.$$

By the definition of $V_h^{(n)}$,

$$P\left(V_h^{(n)} \leq \tau''\right) \geq P\left(Y_\cdot^h(\tau'') = 0\right).$$

Thus

$$P\left(V_h^{(n)} \leq \tau''\right) = 1. \tag{2.3.9}$$

Combining (2.3.8) and (2.3.9), $V_h^{(n)} \to \tau_h$ almost surely as $n \to \infty$. □

By Lemma 2.2.1, $\hat{Q}_{hj}(\tau_h)$ is consistent for $Q_{hj}(\tau_h)$. Note that $\hat{Q}_{hj}(\cdot)$ does not change after $V_h^{(n)}$, and thus $\hat{Q}_{hj}(V_h^{(n)}) = \hat{Q}_{hj}(\infty)$, the plug-in estimator for P_{hj}. Since $V_h^{(n)} \to \tau_h$ almost surely as $n \to \infty$ by Theorem 2.3.1, the plug-in estimator $\hat{P}_{hj} = \hat{Q}_{hj}(\infty) = \hat{Q}_{hj}(V_h^{(n)})$ is consistent for $Q_{hj}(\tau_h)$. It is consistent for P_{hj} only if

$$Q_{hj}(\tau_h) = Q_{hj}(\infty) = P_{hj}. \tag{2.3.10}$$

The normalized estimator,

$$\tilde{P}_{hj} = \hat{Q}_{hj}(\infty) / \sum_{k \neq h} \hat{Q}_{hk}(\infty),$$

is consistent for $Q_{hj}(\tau_h) / \sum_{k \neq h} Q_{hk}(\tau_h)$, which is equal to P_{hj} if

$$F_{hj}(\cdot) = F_{hk}(\cdot), \quad \forall j, k \in \mathcal{E}. \tag{2.3.11}$$

Otherwise, $\tilde{P}_{hj}$ is not necessarily consistent for P_{hj}. Correspondingly, the estimators $\hat{F}_{hj}(\cdot)$ in (2.3.2) and $\tilde{F}_{hj}(\cdot)$ in (2.3.4) are not necessarily consistent for the sojourn time distribution $F_{hj}(\cdot)$ without the assumptions (2.3.10) and (2.3.11), respectively.

Denote $P^L_{hj} = Q_{hj}(\tau_h)$ and $P^U_{hj} = 1 - \sum_{j' \neq j} Q_{hj'}(\tau_h)$. Note that

$$P_{hj} \in [P^L_{hj}, P^U_{hj}]. \tag{2.3.12}$$

The available data provide information about P_{hj} through the two limits P^L_{hj} and P^U_{hj}. The corresponding bound for $F_{hj}(\cdot)$ is then

$$\left[F^L_{hj}(\cdot), F^U_{hj}(\cdot)\right], \tag{2.3.13}$$

where $F^L_{hj}(\cdot) = Q_{hj}(\cdot)/P^U_{hj}$ and $F^U_{hj}(\cdot) = Q_{hj}(\cdot)/P^L_{hj}$. This forms the basis of our robust inference procedure.

2.3.2.2 Robust Inference for Transition Probabilities

We have shown that $\hat{P}^L_{hj} = \hat{Q}_{hj}(\infty)$ and $\hat{P}^U_{hj} = 1 - \sum_{j' \neq j} \hat{Q}_{hj'}(\infty)$ consistently estimate P^L_{hj} and P^U_{hj} respectively. Thus the corresponding estimated bounds for $F_{hj}(t)$ are $\hat{Q}_{hj}(\cdot)/\hat{P}^U_{hj}$ and $\hat{Q}_{hj}(\cdot)/\hat{P}^L_{hj}$. These bounds can be used to construct confidence intervals for P_{hj} and confidence bands for $F_{hj}(\cdot)$.

A confidence interval for P_{hj} with level at least $1 - \alpha$ can be constructed as

$$[\hat{P}^L_{hj} - c_1, \hat{P}^U_{hj} + c_2],$$

where c_1 and c_2 are chosen such that

$$P\left([P^L_{hj}, P^U_{hj}] \in [\hat{P}^L_{hj} - c_1, \hat{P}^U_{hj} + c_2]\right) = 1 - \alpha. \tag{2.3.14}$$

We can choose c_1 and c_2 such that $c_1 + c_2$ is minimized to obtain the confidence interval with the shortest length. The distribution of

$$(\hat{P}^L_{hj}, \hat{P}^U_{hj})' - (P^L_{hj}, P^U_{hj})' \tag{2.3.15}$$

is needed to determine c_1 and c_2. Since $P\left(Y^h(\tau_h) > 0\right)$ can be 0, Lemma 2.2.1 does not in general give us the limiting distribution of (2.3.15) from Remark 2.2.1. We propose to use the bootstrap to determine c_1 and c_2 in (2.3.14).

2.3.2.3 Robust Inference for Sojourn Time Distributions

To obtain a set of confidence bands for the sojourn time distributions $F_{hj}(\cdot)$, we consider a bootstrap approach. We first randomly select a sample $\mathcal{M}$ of size n with replacement from $\{1,\ldots,n\}$, and evaluate $\hat{Q}_{hj}(\cdot)$, $\hat{P}^L_{hj}$, and $\hat{P}^U_{hj}$ based on the data $\{(N_i^{hj}(\cdot), Y_i^h(\cdot)) : i \in \mathcal{M}\}$. This procedure is replicated B times to obtain estimates $\{\hat{Q}^{(b)}_{hj}(\cdot), \hat{P}^{(b)L}_{hj}, \hat{P}^{(b)U}_{hj} : b = 1,\ldots,B\}$. We then have $\left[\hat{F}^{(b)L}_{hj}(\cdot), \hat{F}^{(b)U}_{hj}(\cdot)\right]$ by using $\hat{F}_{hj}(\cdot) = \hat{Q}_{hj}(\cdot)/\hat{P}_{hj}$.

Define

$$H^L_{hj}(\tau) = n^{1/2} h^{1n}_{hj}(\tau)[\hat{F}^L_{hj}(\tau) - F^L_{hj}(\tau)], \quad \tau > 0,$$

and

$$H^U_{hj}(\tau) = n^{1/2} h^{2n}_{hj}(\tau)[\hat{F}^U_{hj}(\tau) - F^U_{hj}(\tau)], \quad \tau > 0,$$

where $h^{1n}_{hj}(\tau)$ and $h^{2n}_{hj}(\tau)$ are weight functions which determine the shape of the bands, with deterministic limit $g^1_{hj}(\tau)$ and $g^2_{hj}(\tau)$ in probability, respectively. For $b = 1,\ldots,B$, let

$$H^{(b)L}_{hj}(\tau) = n^{1/2} h^{1n}_{hj}(\tau)[\hat{F}^{(b)L}_{hj}(\tau) - \hat{F}^L_{hj}(\tau)],$$

and

$$H^{(b)U}_{hj}(\tau) = n^{1/2} h^{2n}_{hj}(\tau)[\hat{F}^{(b)U}_{hj}(\tau) - \hat{F}^U_{hj}(\tau)].$$

To obtain an approximate $(1-\alpha)$ confidence band for $F_{hj}(\cdot)$ on a prechosen interval $[s_1, s_2]$, let $q^{(b)L}_{hj}(s_1, s_2) = \sup_{\tau\in[s_1,s_2]} H^{(b)L}_{hj}(\tau)$ and $q^{(b)U}_{hj}(s_1, s_2) = \inf_{\tau\in[s_1,s_2]} H^{(b)U}_{hj}(\tau)$ for $b = 1,\ldots,B$. Determine $q^L_{hj}(s_1, s_2)$ and $q^U_{hj}(s_1, s_2)$ such that $100(1-\alpha)\%$ of $b \in \{1, 2,\ldots,B\}$ satisfy

$$q^{(b)L}_{hj}(s_1, s_2) < q^L_{hj}(s_1, s_2), \text{ and } q^{(b)U}_{hj}(s_1, s_2) > q^U_{hj}(s_1, s_2).$$

Note that $q^L_{hj}(s_1, s_2)$ and $q^U_{hj}(s_1, s_2)$ may be chosen to optimize the width of the confidence bands. An approximate $(1-\alpha)$ confidence band for the attainable bound $(F^L_{hj}(\cdot), F^U_{hj}(\cdot))$ on $[s_1, s_2]$, as a robust confidence band of $F_{hj}(\cdot)$, is then

$$[\hat{F}^L_{hj}(\tau) - n^{-1/2} q^L_{hj}(s_1, s_2)/h^{1n}_{hj}(\tau), \hat{F}^U_{hj}(\tau) + n^{-1/2} q^U_{hj}(s_1, s_2)/h^{2n}_{hj}(\tau)], \quad \tau \in [s_1, s_2].$$

Since $F_{hj}(\cdot)$ is always within $[0, 1]$, we may need to transform it into a quantity ranging $(-\infty,\infty)$, say, $g\big(F_{hj}(\cdot)\big)$, obtain a confidence band, and then transform the confidence band back to obtain $F_{hj}(\cdot)$'s confidence band. This may improve the coverage of the confidence bands of $F_{hj}(\cdot)$, and ensure the confidence band lies entirely within $[0, 1]$.

2.4 Simulation

2.4.1 Settings

We conducted a simulation with the state space $\mathcal{E} = \{1, 2, 3\}$, where state 3 is an absorbing state. A total of n independent and identically distributed semi-Markov processes were simulated, each of which started from state 1 or 2 with equal probabilities, and was observed up to a noninformative right censoring time C. Specifically, we set the transition probabilities of the embedded Markov chain as

$$P_{12} = 0.7, P_{13} = 0.3, P_{21} = P_{23} = 0.5,$$

and simulated the sojourn time distributions from one of the following three settings:

Setting 2.1. $F_{12} \sim exp(2), F_{13} \sim exp(2), F_{21} \sim exp(1), F_{23} \sim exp(1)$.

Setting 2.2. $F_{12} \sim exp(1), F_{13} \sim exp(2), F_{21} \sim exp(1), F_{23} \sim exp(1)$.

Setting 2.3. $F_{12} \sim exp(1), F_{13} \sim unif(0, 2), F_{21} \sim unif(0, 2), F_{23} \sim exp(2)$.

We designed the simulation settings to study the behaviors of the proposed estimators in various situations. In Setting 2.1, both the pairs (F_{12}, F_{13}) and (F_{21}, F_{23}) have the same entries. Thus both Lagakos *et al.*'s normalized and Phelan's estimators work well for estimating the transition probabilities, but Lagakos *et al.*'s plug-in estimator can be biased since (2.3.10) does not hold in this setting. In Setting 2.2, F_{21} and F_{23} are the same but F_{12} and F_{13} are different. So we anticipate that both Lagakos *et al.*'s normalized and Phelan's estimators will work well for the estimation of P_{21} and P_{23}, but not for P_{12} and P_{13}. Similarly to it in the Setting 2.1, Lagakos *et al.*'s plug-in estimator can be biased for the transition probabilities. Neither of the pairs (F_{12}, F_{13}) and (F_{21}, F_{23}) has the same entries in Setting 2.3, thus both Lagakos *et al.*'s normalized and Phelan's estimators can be biased for the transition probabilities. Since (2.3.10) holds for the cases of $h = 1$, $j = 3$ and $h = 2$, $j = 1$, Lagakos *et al.*'s plug-in estimator can perform well in the estimation of P_{13} and P_{21}.

In each setting, we used sample size $n = 50$, 100, or 200, and the censoring time $C = 2$ or 5 to have 6 scenarios. In what follows, we summarize the simulation results based on $RP = 1000$ repetitions in each scenario for the estimation of the semi-Markov kernel, the transition probabilities, and the sojourn time distributions.

2.4.2 Semi-Markov Kernel Estimation

We first evaluated the confidence intervals for the semi-Markov kernel at 5 fixed time points $t = 0.2$, 0.6, 1.0, 1.4, and 1.8. The empirical coverage of the 95% confidence intervals in the three simulation settings are summarized in Tables 2.1 to 2.3. Note that the coverage of the standard confidence intervals without transformation tends to be lower than the nominal level, and can be very poor for small t. On the other hand, the transformed confidence intervals perform well in terms of coverage frequency, even for small sample size $n = 50$.

We applied both the bootstrap and the resampling approaches to construct confidence bands for the semi-Markov kernel. The confidence bands are restricted to [0.5,1.5]. The coverage frequencies of confidence bands with nominal levels 90% and 95% are summarized in Tables 2.4 to 2.6. Note that the coverages of the bands based on bootstrap and simulation approaches are close. The coverage frequency of the confidence bands without transformation is lower than the nominal level, while the coverage frequency of the confidence bands constructed by applying the transformation are close to the nominal level, especially with small sample size $n = 50$. The improvement brought about by the transformation is more substantial with the equal precision bands than with the Hall-Wellner bands.

In the context of classical survival analysis, the substantial improvement in performance based on transformations has been found for the confidence intervals and bands for the survival function (Borgan and Liestøl, 1990). Another advantage of transformed confidence intervals and bands is that they are always between 0 and 1.

Table 2.1: Empirical coverage probabilities of the 95% confidence intervals for the semi-Markov kernel at different time points based on 1000 simulations in setting 2.1

		Standard CI					Transformed CI				
n	C	$t=0.2$	0.6	1.0	1.4	1.8	$t=0.2$	0.6	1.0	1.4	1.8
$Q_{12}(t)$											
50	2	0.90	0.92	0.93	0.93	0.93	0.96	0.95	0.94	0.95	0.94
	5	0.87	0.92	0.94	0.93	0.92	0.96	0.94	0.95	0.95	0.95
100	2	0.91	0.93	0.94	0.95	0.95	0.97	0.95	0.95	0.95	0.95
	5	0.91	0.93	0.94	0.94	0.94	0.95	0.94	0.93	0.94	0.94
200	2	0.93	0.95	0.94	0.94	0.94	0.96	0.95	0.95	0.95	0.94
	5	0.93	0.94	0.94	0.94	0.94	0.95	0.94	0.95	0.95	0.94
$Q_{13}(t)$											
50	2	0.69	0.89	0.92	0.92	0.93	0.97	0.96	0.96	0.97	0.97
	5	0.75	0.88	0.92	0.94	0.94	0.97	0.96	0.97	0.95	0.95
100	2	0.90	0.92	0.92	0.94	0.94	0.97	0.97	0.95	0.95	0.95
	5	0.90	0.94	0.94	0.94	0.95	0.97	0.97	0.96	0.96	0.95
200	2	0.92	0.93	0.94	0.95	0.96	0.97	0.96	0.95	0.96	0.95
	5	0.91	0.93	0.94	0.95	0.94	0.96	0.95	0.94	0.95	0.96
$Q_{21}(t)$											
50	2	0.88	0.92	0.94	0.92	0.93	0.97	0.96	0.95	0.95	0.95
	5	0.90	0.93	0.94	0.94	0.94	0.96	0.96	0.96	0.95	0.95
100	2	0.93	0.94	0.95	0.94	0.95	0.97	0.96	0.96	0.96	0.95
	5	0.93	0.95	0.94	0.94	0.94	0.96	0.96	0.95	0.95	0.94
200	2	0.94	0.95	0.94	0.95	0.94	0.96	0.96	0.95	0.95	0.95
	5	0.94	0.94	0.94	0.94	0.94	0.95	0.94	0.95	0.94	0.94
$Q_{23}(t)$											
50	2	0.88	0.92	0.92	0.92	0.93	0.96	0.95	0.94	0.94	0.94
	5	0.92	0.94	0.94	0.93	0.94	0.96	0.95	0.95	0.95	0.95
100	2	0.93	0.93	0.94	0.94	0.95	0.97	0.95	0.95	0.95	0.95
	5	0.93	0.94	0.95	0.94	0.94	0.95	0.95	0.95	0.95	0.94
200	2	0.94	0.94	0.94	0.94	0.95	0.95	0.95	0.94	0.95	0.95
	5	0.95	0.95	0.96	0.95	0.95	0.94	0.95	0.96	0.95	0.95

Table 2.2: Empirical coverage probabilities of the 95% confidence intervals for the semi-Markov kernel at different time points based on 1000 simulations in setting 2.2

		Standard CI					Transformed CI				
n	C	$t = 0.2$	0.6	1.0	1.4	1.8	$t = 0.2$	0.6	1.0	1.4	1.8
$Q_{12}(t)$											
50	2	0.90	0.94	0.92	0.93	0.94	0.96	0.95	0.94	0.94	0.95
	5	0.90	0.92	0.93	0.94	0.94	0.95	0.93	0.94	0.94	0.95
100	2	0.92	0.94	0.94	0.94	0.94	0.96	0.95	0.94	0.95	0.95
	5	0.92	0.93	0.93	0.93	0.92	0.94	0.94	0.94	0.93	0.92
200	2	0.95	0.94	0.95	0.94	0.95	0.96	0.94	0.95	0.94	0.95
	5	0.93	0.95	0.93	0.93	0.92	0.94	0.95	0.93	0.93	0.92
$Q_{13}(t)$											
50	2	0.71	0.88	0.91	0.93	0.93	0.96	0.97	0.97	0.96	0.96
	5	0.78	0.90	0.91	0.93	0.92	0.96	0.96	0.97	0.96	0.95
100	2	0.92	0.92	0.92	0.93	0.94	0.97	0.97	0.95	0.95	0.95
	5	0.85	0.92	0.93	0.95	0.94	0.97	0.96	0.96	0.96	0.96
200	2	0.89	0.93	0.94	0.95	0.95	0.96	0.95	0.96	0.96	0.96
	5	0.92	0.93	0.94	0.94	0.94	0.97	0.96	0.95	0.95	0.95
$Q_{21}(t)$											
50	2	0.91	0.93	0.93	0.92	0.92	0.95	0.96	0.95	0.94	0.94
	5	0.91	0.93	0.93	0.92	0.92	0.97	0.94	0.94	0.92	0.93
100	2	0.93	0.93	0.94	0.94	0.94	0.97	0.94	0.96	0.94	0.95
	5	0.92	0.93	0.92	0.93	0.93	0.95	0.94	0.94	0.93	0.94
200	2	0.93	0.94	0.93	0.93	0.94	0.96	0.95	0.94	0.94	0.95
	5	0.93	0.93	0.92	0.92	0.93	0.94	0.94	0.92	0.92	0.93
$Q_{23}(t)$											
50	2	0.93	0.93	0.93	0.93	0.93	0.97	0.95	0.94	0.93	0.94
	5	0.92	0.93	0.94	0.94	0.94	0.97	0.95	0.94	0.95	0.94
100	2	0.95	0.94	0.94	0.94	0.93	0.96	0.95	0.95	0.94	0.94
	5	0.95	0.94	0.94	0.94	0.94	0.96	0.94	0.94	0.94	0.94
200	2	0.94	0.94	0.94	0.93	0.95	0.95	0.95	0.94	0.94	0.95
	5	0.95	0.94	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95

Table 2.3: Empirical coverage probabilities of the 95% confidence intervals for the semi-Markov kernel at different time points based on 1000 simulations in setting 2.3

		Standard CI					Transformed CI				
n	C	$t = 0.2$	0.6	1.0	1.4	1.8	$t = 0.2$	0.6	1.0	1.4	1.8
$Q_{12}(t)$											
50	2	0.91	0.94	0.94	0.94	0.92	0.96	0.95	0.95	0.96	0.94
	5	0.92	0.93	0.92	0.92	0.93	0.96	0.95	0.93	0.94	0.94
100	2	0.93	0.94	0.94	0.94	0.92	0.94	0.95	0.94	0.95	0.93
	5	0.93	0.94	0.94	0.93	0.94	0.94	0.95	0.94	0.94	0.94
200	2	0.94	0.94	0.95	0.94	0.94	0.95	0.94	0.95	0.94	0.94
	5	0.93	0.95	0.95	0.94	0.95	0.94	0.95	0.95	0.94	0.95
$Q_{13}(t)$											
50	2	0.71	0.88	0.90	0.91	0.93	0.96	0.98	0.96	0.96	0.95
	5	0.79	0.89	0.91	0.92	0.93	0.96	0.97	0.96	0.95	0.95
100	2	0.92	0.92	0.93	0.94	0.93	0.97	0.97	0.94	0.95	0.94
	5	0.84	0.91	0.94	0.92	0.94	0.96	0.95	0.96	0.94	0.95
200	2	0.89	0.94	0.94	0.94	0.95	0.97	0.96	0.95	0.95	0.95
	5	0.89	0.93	0.93	0.94	0.92	0.96	0.95	0.94	0.93	0.94
$Q_{21}(t)$											
50	2	0.89	0.88	0.91	0.92	0.93	0.96	0.95	0.94	0.95	0.95
	5	0.88	0.92	0.93	0.93	0.92	0.97	0.95	0.95	0.95	0.94
100	2	0.91	0.92	0.93	0.94	0.95	0.95	0.94	0.95	0.95	0.95
	5	0.92	0.94	0.95	0.95	0.94	0.97	0.96	0.95	0.95	0.94
200	2	0.94	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.96
	5	0.94	0.93	0.95	0.94	0.93	0.94	0.95	0.95	0.95	0.94
$Q_{23}(t)$											
50	2	0.90	0.92	0.92	0.94	0.94	0.96	0.96	0.95	0.95	0.95
	5	0.88	0.93	0.92	0.93	0.95	0.96	0.96	0.94	0.94	0.96
100	2	0.94	0.92	0.94	0.93	0.95	0.97	0.95	0.96	0.95	0.95
	5	0.92	0.94	0.95	0.95	0.95	0.97	0.96	0.94	0.95	0.95
200	2	0.93	0.94	0.95	0.94	0.95	0.97	0.95	0.96	0.95	0.95
	5	0.92	0.94	0.94	0.95	0.95	0.95	0.94	0.94	0.95	0.95

Table 2.4: Empirical coverage probabilities of the equal precision (EP), Hall-Wellner (HW), transformed equal precision (TEP), and transformed Hall-Wellner (THW) 95% confidence bands for the semi-Markov kernel based on 1000 simulations in setting 2.1

		L-W-Y Resampling				Bootstrap			
n	C	EP	HW	TEP	THW	EP	HW	TEP	THW
$Q_{12}(\cdot)$									
50	2	0.86	0.91	0.92	0.91	0.86	0.91	0.92	0.92
	5	0.88	0.91	0.92	0.92	0.87	0.91	0.92	0.92
100	2	0.90	0.92	0.93	0.93	0.90	0.92	0.93	0.93
	5	0.90	0.91	0.92	0.92	0.90	0.91	0.92	0.92
200	2	0.92	0.94	0.94	0.94	0.92	0.93	0.94	0.94
	5	0.92	0.93	0.94	0.93	0.91	0.93	0.93	0.94
$Q_{13}(\cdot)$									
50	2	0.82	0.91	0.94	0.92	0.81	0.90	0.93	0.92
	5	0.81	0.91	0.93	0.91	0.82	0.91	0.94	0.92
100	2	0.86	0.92	0.94	0.93	0.86	0.92	0.94	0.93
	5	0.88	0.93	0.94	0.94	0.88	0.93	0.94	0.93
200	2	0.91	0.94	0.94	0.94	0.90	0.93	0.94	0.94
	5	0.90	0.93	0.93	0.93	0.90	0.92	0.93	0.93
$Q_{21}(\cdot)$									
50	2	0.88	0.90	0.93	0.93	0.88	0.91	0.93	0.93
	5	0.89	0.93	0.93	0.93	0.88	0.93	0.94	0.93
100	2	0.91	0.92	0.95	0.95	0.91	0.92	0.94	0.94
	5	0.91	0.91	0.93	0.94	0.90	0.91	0.93	0.93
200	2	0.92	0.93	0.94	0.94	0.92	0.93	0.93	0.93
	5	0.91	0.93	0.93	0.94	0.91	0.93	0.93	0.94
$Q_{23}(\cdot)$									
50	2	0.86	0.91	0.93	0.93	0.86	0.90	0.92	0.93
	5	0.90	0.92	0.92	0.91	0.90	0.92	0.92	0.91
100	2	0.90	0.92	0.94	0.94	0.90	0.92	0.94	0.94
	5	0.93	0.93	0.94	0.94	0.93	0.93	0.93	0.93
200	2	0.93	0.94	0.92	0.92	0.92	0.94	0.92	0.92
	5	0.95	0.95	0.95	0.94	0.94	0.94	0.95	0.94

Table 2.5: Empirical coverage probabilities of the equal precision (EP), Hall-Wellner (HW), transformed equal precision (TEP), and transformed Hall-Wellner (THW) 95% confidence bands for the semi-Markov kernel based on 1000 simulations in setting 2.2

		L-W-Y Resampling				Bootstrap			
n	C	EP	HW	TEP	THW	EP	HW	TEP	THW
$Q_{12}(\cdot)$									
50	2	0.88	0.90	0.93	0.93	0.88	0.90	0.93	0.93
	5	0.88	0.89	0.90	0.91	0.88	0.89	0.90	0.91
100	2	0.90	0.92	0.92	0.93	0.90	0.92	0.92	0.93
	5	0.90	0.91	0.92	0.92	0.91	0.91	0.92	0.93
200	2	0.92	0.93	0.93	0.93	0.92	0.93	0.93	0.94
	5	0.92	0.92	0.93	0.92	0.92	0.93	0.93	0.93
$Q_{13}(\cdot)$									
50	2	0.81	0.89	0.93	0.91	0.82	0.89	0.93	0.92
	5	0.84	0.90	0.94	0.92	0.83	0.90	0.94	0.92
100	2	0.85	0.91	0.94	0.92	0.85	0.91	0.94	0.92
	5	0.87	0.93	0.94	0.93	0.87	0.92	0.94	0.93
200	2	0.90	0.94	0.95	0.94	0.90	0.94	0.94	0.93
	5	0.90	0.92	0.94	0.94	0.91	0.92	0.94	0.93
$Q_{21}(\cdot)$									
50	2	0.87	0.90	0.93	0.92	0.86	0.90	0.93	0.93
	5	0.87	0.90	0.91	0.91	0.87	0.91	0.91	0.91
100	2	0.89	0.91	0.93	0.93	0.89	0.91	0.93	0.93
	5	0.90	0.90	0.93	0.93	0.89	0.91	0.93	0.92
200	2	0.91	0.91	0.93	0.93	0.91	0.91	0.93	0.93
	5	0.90	0.91	0.92	0.91	0.90	0.91	0.92	0.92
$Q_{23}(\cdot)$									
50	2	0.89	0.92	0.92	0.92	0.89	0.91	0.92	0.92
	5	0.88	0.92	0.94	0.93	0.88	0.92	0.94	0.93
100	2	0.91	0.92	0.94	0.93	0.92	0.92	0.93	0.93
	5	0.92	0.93	0.93	0.93	0.92	0.94	0.93	0.93
200	2	0.92	0.93	0.93	0.93	0.92	0.92	0.93	0.93
	5	0.93	0.94	0.94	0.94	0.92	0.94	0.93	0.94

Table 2.6: Empirical coverage probabilities of the equal precision (EP), Hall-Wellner (HW), transformed equal precision (TEP), and transformed Hall-Wellner (THW) 95% confidence bands for the semi-Markov kernel based on 1000 simulations in setting 2.3

		L-W-Y Resampling				Bootstrap			
n	C	EP	HW	TEP	THW	EP	HW	TEP	THW
$Q_{12}(\cdot)$									
50	2	0.88	0.90	0.93	0.93	0.88	0.91	0.92	0.94
	5	0.89	0.90	0.91	0.91	0.89	0.90	0.92	0.92
100	2	0.90	0.92	0.93	0.93	0.90	0.92	0.93	0.93
	5	0.92	0.92	0.93	0.93	0.91	0.92	0.92	0.93
200	2	0.91	0.93	0.92	0.93	0.92	0.93	0.92	0.93
	5	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
$Q_{13}(\cdot)$									
50	2	0.77	0.89	0.94	0.92	0.78	0.90	0.93	0.92
	5	0.81	0.91	0.93	0.92	0.82	0.91	0.93	0.92
100	2	0.85	0.91	0.94	0.92	0.85	0.90	0.94	0.92
	5	0.85	0.91	0.93	0.93	0.85	0.91	0.93	0.92
200	2	0.91	0.92	0.94	0.93	0.90	0.93	0.94	0.93
	5	0.90	0.93	0.93	0.92	0.90	0.92	0.92	0.92
$Q_{21}(\cdot)$									
50	2	0.80	0.89	0.93	0.92	0.80	0.88	0.93	0.92
	5	0.86	0.91	0.93	0.92	0.86	0.91	0.93	0.92
100	2	0.88	0.92	0.93	0.93	0.88	0.92	0.94	0.93
	5	0.91	0.93	0.93	0.93	0.90	0.93	0.93	0.93
200	2	0.92	0.93	0.95	0.94	0.92	0.93	0.94	0.94
	5	0.92	0.93	0.93	0.93	0.92	0.93	0.93	0.94
$Q_{23}(\cdot)$									
50	2	0.85	0.92	0.93	0.92	0.84	0.91	0.93	0.92
	5	0.88	0.92	0.94	0.92	0.88	0.92	0.94	0.92
100	2	0.88	0.92	0.94	0.94	0.88	0.92	0.94	0.93
	5	0.92	0.95	0.94	0.93	0.91	0.95	0.94	0.93
200	2	0.92	0.94	0.94	0.94	0.92	0.93	0.94	0.93
	5	0.92	0.94	0.93	0.94	0.91	0.93	0.93	0.93

2.4.3 Transition Probabilities and Sojourn Time Distributions

Lagakos *et al.*'s plug-in and normalized estimates and Phelan's estimate for the transition probabilities of the embedded Markov chain in the three simulation settings are summarized in Tables 2.7 to 2.9. Note from the sample means associated with the estimators that the plug-in estimators are biased for the transition probabilities in all the simulation settings except for P_{13} and P_{21} in Setting 2.3, which satisfy (2.3.10). Both the normalized and Phelan's estimators are verified to be consistent for all the transition probabilities in Setting 2.1, and with P_{21} and P_{23} in Setting 2.2. This is due to the same corresponding sojourn time distributions in the situations. They are biased at other settings. In Setting 2.3, the sample biases of the normalized and Phelan's estimators for P_{21} and P_{23} are larger than for P_{12} and P_{13}, since the difference between $exp(2)$ and $unif(0,2)$ is larger than the difference between $exp(1)$ and $unif(0,2)$. In all simulation settings, the biases become smaller when the censoring time C increases from 2 to 5. With increased sample size n, the standard deviations decrease, but the biases do not shrink.

We constructed confidence intervals for the transition probabilities P_{ij}'s at each simulation setting. For each simulated data set, we resampled $B = 500$ times to get the bootstrap estimate for the distribution of (2.3.15) to construct confidence intervals for P_{ij}'s. Tables 2.10 to 2.12 present the coverage frequencies and the sample mean lengths of the estimated 95% confidence intervals in the three simulation settings for the transition probabilities based on the three existing methods and the proposed robust approach. Note that the coverages of the confidence intervals based on the three existing methods can be rather low when the corresponding point estimates are biased, especially with the smaller censoring time $C = 2$. By contrast, the robust confidence intervals contain the attainable values of the transition probabilities (2.3.12) at approximately the nominal level, and thus cover the true transition probabilities at least at the nominal level. However, the confidence intervals constructed by the robust approach are wider than the other three existing approaches, especially with small censoring time $C = 2$. With the larger censoring time $C = 5$, the coverage becomes close to the nominal level, and the length of the confidence intervals are comparable with the ones based on the three existing approaches.

We also evaluated the confidence bands for the attainable sojourn time distributions (2.3.13). The simulation outcomes are summarized in Tables 2.13 to 2.15 for the three simulation settings respectively. Note that the coverage frequency of the confidence bands is lower than the nominal level without transformation, while the confidence bands constructed by applying transformation have coverage frequency closer to the nominal level, especially for small sample size $n = 50$.

Table 2.7: Sample mean, bias, and standard deviation (SD) of the estimated transition probabilities of the embedded Markov chain based on 1000 simulations in setting 2.1

		Plug-in			Normalized			Phelan		
n	C	Mean	Bias	SD	Mean	Bias	SD	Mean	Bias	SD
$P_{12} = 0.7$										
50	2	0.44	-0.26	0.09	0.70	-0.00	0.10	0.70	-0.00	0.10
	5	0.64	-0.06	0.07	0.70	-0.00	0.07	0.70	-0.00	0.07
100	2	0.44	-0.26	0.06	0.70	-0.00	0.07	0.70	-0.00	0.07
	5	0.64	-0.06	0.05	0.70	-0.00	0.05	0.70	-0.00	0.05
200	2	0.44	-0.26	0.04	0.70	-0.00	0.05	0.70	-0.00	0.05
	5	0.64	-0.06	0.04	0.70	-0.00	0.04	0.70	-0.00	0.04
$P_{13} = 0.3$										
50	2	0.19	-0.11	0.07	0.30	0.00	0.10	0.30	0.00	0.10
	5	0.28	-0.02	0.06	0.30	0.00	0.07	0.30	0.00	0.07
100	2	0.19	-0.11	0.05	0.30	0.00	0.07	0.30	0.00	0.07
	5	0.27	-0.03	0.05	0.30	0.00	0.05	0.30	0.00	0.05
200	2	0.19	-0.11	0.03	0.30	0.00	0.05	0.30	0.00	0.05
	5	0.28	-0.02	0.03	0.30	0.00	0.04	0.30	0.00	0.04
$P_{21} = 0.5$										
50	2	0.43	-0.07	0.09	0.50	-0.00	0.09	0.50	-0.00	0.09
	5	0.49	-0.01	0.07	0.49	-0.01	0.07	0.49	-0.01	0.07
100	2	0.43	-0.07	0.06	0.50	-0.00	0.06	0.50	-0.00	0.06
	5	0.49	-0.01	0.05	0.50	-0.00	0.05	0.50	-0.00	0.05
200	2	0.43	-0.07	0.04	0.50	-0.00	0.05	0.50	-0.00	0.05
	5	0.49	-0.01	0.04	0.50	-0.00	0.04	0.50	-0.00	0.04
$P_{23} = 0.5$										
50	2	0.44	-0.06	0.09	0.50	0.00	0.09	0.50	0.00	0.09
	5	0.50	0.00	0.07	0.51	0.01	0.07	0.51	0.01	0.07
100	2	0.43	-0.07	0.06	0.50	0.00	0.06	0.50	0.00	0.06
	5	0.50	0.00	0.05	0.50	0.00	0.05	0.50	0.00	0.05
200	2	0.43	-0.07	0.04	0.50	0.00	0.05	0.50	0.00	0.05
	5	0.50	-0.00	0.04	0.50	0.00	0.04	0.50	0.00	0.04

Table 2.8: Sample mean, bias, and standard deviation (SD) of the estimated transition probabilities of the embedded Markov chain based on 1000 simulations in setting 2.2

		Plug-in			Normalized			Phelan		
n	C	Mean	Bias	SD	Mean	Bias	SD	Mean	Bias	SD
$P_{12} = 0.7$										
50	2	0.60	-0.10	0.08	0.76	0.06	0.08	0.76	0.06	0.08
	5	0.69	-0.01	0.07	0.72	0.02	0.07	0.73	0.03	0.07
100	2	0.60	-0.10	0.06	0.76	0.06	0.06	0.77	0.07	0.06
	5	0.70	-0.00	0.05	0.72	0.02	0.05	0.73	0.03	0.05
200	2	0.60	-0.10	0.04	0.76	0.06	0.04	0.77	0.07	0.04
	5	0.70	-0.00	0.03	0.72	0.02	0.03	0.73	0.03	0.03
$P_{13} = 0.3$										
50	2	0.19	-0.11	0.07	0.24	-0.06	0.08	0.24	-0.06	0.08
	5	0.27	-0.03	0.07	0.28	-0.02	0.07	0.27	-0.03	0.07
100	2	0.19	-0.11	0.05	0.24	-0.06	0.06	0.23	-0.07	0.06
	5	0.27	-0.03	0.05	0.28	-0.02	0.05	0.27	-0.03	0.05
200	2	0.19	-0.11	0.03	0.24	-0.06	0.04	0.23	-0.07	0.04
	5	0.27	-0.03	0.03	0.28	-0.02	0.03	0.27	-0.03	0.03
$P_{21} = 0.5$										
50	2	0.43	-0.07	0.08	0.50	-0.00	0.09	0.50	-0.00	0.09
	5	0.49	-0.01	0.07	0.50	-0.00	0.07	0.50	-0.00	0.07
100	2	0.43	-0.07	0.06	0.50	-0.00	0.06	0.50	-0.00	0.06
	5	0.50	-0.00	0.05	0.50	-0.00	0.05	0.50	-0.00	0.05
200	2	0.43	-0.07	0.04	0.50	-0.00	0.04	0.50	-0.00	0.04
	5	0.50	-0.00	0.03	0.50	0.00	0.03	0.50	-0.00	0.03
$P_{23} = 0.5$										
50	2	0.44	-0.06	0.08	0.50	0.00	0.09	0.50	0.00	0.09
	5	0.50	-0.00	0.07	0.50	0.00	0.07	0.50	0.00	0.07
100	2	0.44	-0.06	0.06	0.50	0.00	0.06	0.50	0.00	0.06
	5	0.50	-0.00	0.05	0.50	0.00	0.05	0.50	0.00	0.05
200	2	0.43	-0.07	0.04	0.50	0.00	0.04	0.50	0.00	0.04
	5	0.50	-0.00	0.03	0.50	-0.00	0.03	0.50	0.00	0.03

Table 2.9: Sample mean, bias, and standard deviation (SD) of the estimated transition probabilities of the embedded Markov chain based on 1000 simulations in setting 2.3

		Plug-in			Normalized			Phelan		
n	C	Mean	Bias	SD	Mean	Bias	SD	Mean	Bias	SD
$P_{12} = 0.7$										
50	2	0.60	-0.10	0.09	0.67	-0.03	0.09	0.69	-0.01	0.08
	5	0.69	-0.01	0.06	0.70	-0.00	0.06	0.70	-0.00	0.06
100	2	0.60	-0.10	0.06	0.67	-0.03	0.06	0.69	-0.01	0.06
	5	0.70	-0.00	0.05	0.70	-0.00	0.05	0.70	-0.00	0.05
200	2	0.60	-0.10	0.04	0.67	-0.03	0.04	0.69	-0.01	0.04
	5	0.70	-0.00	0.03	0.70	-0.00	0.03	0.70	-0.00	0.03
$P_{13} = 0.3$										
50	2	0.30	0.00	0.08	0.33	0.03	0.09	0.31	0.01	0.08
	5	0.30	0.00	0.06	0.30	0.00	0.06	0.30	0.00	0.06
100	2	0.30	0.00	0.06	0.33	0.03	0.06	0.31	0.01	0.06
	5	0.30	-0.00	0.05	0.30	0.00	0.05	0.30	0.00	0.05
200	2	0.30	0.00	0.04	0.33	0.03	0.04	0.31	0.01	0.04
	5	0.30	-0.00	0.03	0.30	0.00	0.03	0.30	0.00	0.03
$P_{21} = 0.5$										
50	2	0.50	-0.00	0.09	0.61	0.11	0.09	0.60	0.10	0.09
	5	0.50	-0.00	0.07	0.52	0.02	0.07	0.53	0.03	0.07
100	2	0.50	-0.00	0.06	0.61	0.11	0.06	0.60	0.10	0.06
	5	0.50	-0.00	0.05	0.52	0.02	0.05	0.53	0.03	0.05
200	2	0.50	-0.00	0.04	0.61	0.11	0.04	0.60	0.10	0.04
	5	0.50	-0.00	0.03	0.52	0.02	0.03	0.54	0.04	0.04
$P_{23} = 0.5$										
50	2	0.32	-0.18	0.08	0.39	-0.11	0.09	0.40	-0.10	0.09
	5	0.46	-0.04	0.07	0.48	-0.02	0.07	0.47	-0.03	0.07
100	2	0.32	-0.18	0.05	0.39	-0.11	0.06	0.40	-0.10	0.06
	5	0.46	-0.04	0.05	0.48	-0.02	0.05	0.47	-0.03	0.05
200	2	0.32	-0.18	0.04	0.39	-0.11	0.04	0.40	-0.10	0.04
	5	0.46	-0.04	0.03	0.48	-0.02	0.03	0.46	-0.04	0.04

Table 2.10: Empirical coverage probabilities (CP) and sample mean lengths (ML) of the 95% confidence intervals for the transition probabilities of the embedded Markov chain based on 1000 simulations in setting 2.1

		Plug-in		Normalized		Phelan		Robust Approach		
n	C	CP	ML	CP	ML	CP	ML	CP1[1]	CP2[2]	ML
$P_{12} = 0.7$										
50	2	0.17	0.33	0.92	0.39	0.93	0.38	1.00	0.95	0.67
	5	0.89	0.29	0.95	0.28	0.95	0.28	0.99	0.96	0.36
100	2	0.01	0.24	0.93	0.27	0.94	0.27	1.00	0.96	0.58
	5	0.83	0.20	0.95	0.20	0.95	0.20	1.00	0.96	0.28
200	2	0.00	0.17	0.94	0.19	0.94	0.19	1.00	0.95	0.52
	5	0.67	0.14	0.95	0.14	0.94	0.14	1.00	0.96	0.22
$P_{13} = 0.3$										
50	2	0.59	0.26	0.92	0.39	0.93	0.38	1.00	0.95	0.67
	5	0.91	0.26	0.95	0.28	0.95	0.28	0.99	0.96	0.36
100	2	0.38	0.19	0.93	0.27	0.94	0.27	1.00	0.96	0.58
	5	0.91	0.19	0.95	0.20	0.95	0.20	1.00	0.96	0.28
200	2	0.12	0.13	0.94	0.19	0.94	0.19	1.00	0.95	0.52
	5	0.88	0.13	0.95	0.14	0.94	0.14	1.00	0.96	0.22
$P_{21} = 0.5$										
50	2	0.83	0.33	0.93	0.35	0.93	0.35	0.99	0.94	0.46
	5	0.94	0.28	0.94	0.28	0.94	0.28	0.95	0.94	0.29
100	2	0.77	0.23	0.94	0.25	0.94	0.25	1.00	0.95	0.37
	5	0.94	0.20	0.94	0.20	0.94	0.20	0.95	0.94	0.20
200	2	0.62	0.17	0.95	0.18	0.95	0.18	1.00	0.95	0.30
	5	0.94	0.14	0.94	0.14	0.94	0.14	0.95	0.94	0.15
$P_{23} = 0.5$										
50	2	0.85	0.33	0.93	0.35	0.93	0.35	0.99	0.94	0.46
	5	0.94	0.28	0.94	0.28	0.94	0.28	0.95	0.94	0.29
100	2	0.79	0.23	0.94	0.25	0.94	0.25	1.00	0.95	0.37
	5	0.94	0.20	0.94	0.20	0.94	0.20	0.95	0.94	0.20
200	2	0.64	0.17	0.95	0.18	0.95	0.18	1.00	0.95	0.30
	5	0.94	0.14	0.94	0.14	0.94	0.14	0.95	0.94	0.15

[1] CP for the transition probabilities

[2] CP for the attainable values of the transition probabilities

Table 2.11: Empirical coverage probabilities (CP) and sample mean lengths (ML) of the 95% confidence intervals for the transition probabilities of the embedded Markov chain based on 1000 simulations in setting 2.2

		Plug-in		Normalized		Phelan		Robust Approach		
n	C	CP	ML	CP	ML	CP	ML	CP1[1]	CP2[2]	ML
$P_{12} = 0.7$										
50	2	0.78	0.32	0.82	0.31	0.80	0.31	1.00	0.94	0.50
	5	0.94	0.26	0.92	0.26	0.89	0.25	0.96	0.94	0.29
100	2	0.62	0.23	0.77	0.22	0.74	0.22	1.00	0.94	0.41
	5	0.94	0.18	0.92	0.18	0.89	0.18	0.96	0.94	0.21
200	2	0.38	0.16	0.64	0.16	0.57	0.15	1.00	0.95	0.35
	5	0.94	0.13	0.90	0.13	0.83	0.13	0.97	0.95	0.16
$P_{13} = 0.3$										
50	2	0.59	0.26	0.82	0.31	0.80	0.31	1.00	0.94	0.50
	5	0.90	0.25	0.92	0.26	0.89	0.25	0.96	0.94	0.29
100	2	0.39	0.18	0.77	0.22	0.74	0.22	1.00	0.94	0.41
	5	0.89	0.18	0.92	0.18	0.89	0.18	0.96	0.94	0.21
200	2	0.13	0.13	0.64	0.16	0.57	0.15	1.00	0.95	0.35
	5	0.84	0.13	0.90	0.13	0.83	0.13	0.97	0.95	0.16
$P_{21} = 0.5$										
50	2	0.82	0.31	0.92	0.33	0.93	0.32	0.99	0.94	0.44
	5	0.93	0.26	0.93	0.26	0.93	0.26	0.94	0.94	0.27
100	2	0.74	0.22	0.93	0.23	0.94	0.23	1.00	0.95	0.35
	5	0.93	0.18	0.93	0.18	0.94	0.18	0.94	0.94	0.19
200	2	0.58	0.15	0.94	0.17	0.93	0.16	1.00	0.95	0.29
	5	0.94	0.13	0.94	0.13	0.93	0.13	0.96	0.94	0.14
$P_{23} = 0.5$										
50	2	0.82	0.31	0.92	0.33	0.93	0.32	0.99	0.94	0.44
	5	0.93	0.26	0.93	0.26	0.93	0.26	0.94	0.94	0.27
100	2	0.78	0.22	0.93	0.23	0.94	0.23	1.00	0.95	0.35
	5	0.93	0.18	0.93	0.18	0.94	0.18	0.94	0.94	0.19
200	2	0.60	0.15	0.94	0.17	0.93	0.16	1.00	0.95	0.29
	5	0.94	0.13	0.94	0.13	0.93	0.13	0.96	0.94	0.14

[1] CP for the transition probabilities

[2] CP for the attainable values of the transition probabilities

Table 2.12: Empirical coverage probabilities (CP) and sample mean lengths (ML) of the 95% confidence intervals for the transition probabilities of the embedded Markov chain based on 1000 simulations in setting 2.3

		Plug-in		Normalized		Phelan		Robust Approach		
n	C	CP	ML	CP	ML	CP	ML	CP1[1]	CP2[2]	ML
$P_{12} = 0.7$										
50	2	0.81	0.33	0.93	0.33	0.94	0.32	0.98	0.95	0.42
	5	0.95	0.25	0.94	0.25	0.94	0.25	0.95	0.94	0.26
100	2	0.65	0.23	0.91	0.24	0.93	0.23	0.97	0.94	0.32
	5	0.95	0.18	0.95	0.18	0.95	0.18	0.95	0.95	0.18
200	2	0.38	0.16	0.89	0.17	0.94	0.16	0.98	0.95	0.26
	5	0.94	0.13	0.94	0.13	0.94	0.13	0.95	0.94	0.13
$P_{13} = 0.3$										
50	2	0.94	0.31	0.93	0.33	0.94	0.32	0.98	0.95	0.42
	5	0.94	0.25	0.94	0.25	0.94	0.25	0.95	0.94	0.26
100	2	0.93	0.22	0.91	0.24	0.93	0.23	0.97	0.94	0.32
	5	0.94	0.18	0.95	0.18	0.95	0.18	0.95	0.95	0.18
200	2	0.95	0.16	0.89	0.17	0.94	0.16	0.98	0.95	0.26
	5	0.94	0.13	0.94	0.13	0.94	0.13	0.95	0.94	0.13
$P_{21} = 0.5$										
50	2	0.93	0.33	0.74	0.35	0.77	0.34	0.97	0.95	0.50
	5	0.94	0.26	0.93	0.27	0.90	0.27	0.97	0.94	0.30
100	2	0.94	0.23	0.58	0.24	0.63	0.24	0.97	0.95	0.40
	5	0.94	0.18	0.92	0.19	0.88	0.19	0.97	0.94	0.23
200	2	0.95	0.17	0.30	0.17	0.37	0.17	0.98	0.95	0.34
	5	0.94	0.13	0.89	0.13	0.80	0.14	0.96	0.94	0.17
$P_{23} = 0.5$										
50	2	0.37	0.30	0.74	0.35	0.77	0.34	0.97	0.95	0.50
	5	0.90	0.26	0.93	0.27	0.90	0.27	0.97	0.94	0.30
100	2	0.10	0.21	0.58	0.24	0.63	0.24	0.97	0.95	0.40
	5	0.86	0.19	0.92	0.19	0.88	0.19	0.97	0.94	0.23
200	2	0.01	0.15	0.30	0.17	0.37	0.17	0.98	0.95	0.34
	5	0.77	0.13	0.89	0.13	0.80	0.14	0.96	0.94	0.17

[1] CP for the transition probabilities
[2] CP for the attainable values of the transition probabilities

Table 2.13: Empirical coverage probabilities of the equal precision (EP), Hall-Wellner (HW), transformed equal precision (TEP), and transformed Hall-Wellner (THW) 90% and 95% confidence bands for the attainable sojourn time distributions based on 1000 simulations in setting 2.1

		Nominal level 90%				Nominal level 95%			
n	C	EP	HW	TEP	THW	EP	HW	TEP	THW
$F_{12}(\cdot)$									
50	2	0.85	0.80	0.89	0.83	0.92	0.87	0.93	0.91
	5	0.80	0.82	0.87	0.86	0.87	0.90	0.92	0.91
100	2	0.88	0.85	0.89	0.85	0.94	0.91	0.95	0.92
	5	0.84	0.85	0.86	0.86	0.90	0.91	0.92	0.92
200	2	0.90	0.86	0.88	0.85	0.95	0.93	0.95	0.92
	5	0.86	0.86	0.87	0.87	0.91	0.92	0.93	0.93
$F_{13}(\cdot)$									
50	2	0.83	0.78	0.90	0.83	0.88	0.85	0.95	0.92
	5	0.77	0.77	0.90	0.90	0.83	0.85	0.95	0.94
100	2	0.87	0.82	0.92	0.83	0.92	0.89	0.96	0.92
	5	0.84	0.83	0.90	0.90	0.89	0.89	0.95	0.94
200	2	0.89	0.85	0.90	0.85	0.94	0.91	0.95	0.93
	5	0.85	0.85	0.89	0.88	0.92	0.91	0.94	0.94
$F_{21}(\cdot)$									
50	2	0.84	0.84	0.90	0.89	0.89	0.91	0.96	0.94
	5	0.76	0.82	0.85	0.85	0.84	0.89	0.92	0.92
100	2	0.87	0.85	0.89	0.86	0.92	0.91	0.95	0.94
	5	0.82	0.85	0.86	0.87	0.88	0.91	0.91	0.93
200	2	0.89	0.88	0.90	0.87	0.95	0.94	0.94	0.94
	5	0.84	0.86	0.87	0.87	0.91	0.93	0.93	0.94
$F_{23}(\cdot)$									
50	2	0.82	0.82	0.88	0.87	0.88	0.90	0.95	0.94
	5	0.79	0.85	0.88	0.90	0.87	0.92	0.94	0.96
100	2	0.87	0.87	0.90	0.88	0.93	0.92	0.95	0.94
	5	0.84	0.86	0.88	0.89	0.90	0.93	0.94	0.94
200	2	0.88	0.87	0.89	0.87	0.94	0.93	0.94	0.93
	5	0.86	0.88	0.88	0.88	0.91	0.93	0.94	0.94

Table 2.14: Empirical coverage probabilities of the equal precision (EP), Hall-Wellner (HW), transformed equal precision (TEP), and transformed Hall-Wellner (THW) 90% and 95% confidence bands for the attainable sojourn time distributions based on 1000 simulations in setting 2.2

		Nominal level 90%				Nominal level 95%			
n	C	EP	HW	TEP	THW	EP	HW	TEP	THW
$F_{12}(\cdot)$									
50	2	0.85	0.83	0.89	0.84	0.92	0.90	0.95	0.92
	5	0.78	0.82	0.82	0.85	0.86	0.89	0.91	0.91
100	2	0.86	0.83	0.87	0.84	0.93	0.91	0.94	0.91
	5	0.84	0.85	0.87	0.85	0.91	0.92	0.93	0.93
200	2	0.89	0.86	0.87	0.86	0.94	0.93	0.94	0.92
	5	0.85	0.88	0.88	0.88	0.93	0.93	0.94	0.94
$F_{13}(\cdot)$									
50	2	0.81	0.80	0.90	0.88	0.87	0.86	0.95	0.95
	5	0.75	0.78	0.88	0.90	0.83	0.86	0.94	0.95
100	2	0.86	0.83	0.89	0.86	0.92	0.90	0.95	0.93
	5	0.80	0.82	0.88	0.89	0.86	0.88	0.94	0.94
200	2	0.89	0.88	0.90	0.87	0.94	0.94	0.96	0.94
	5	0.83	0.85	0.87	0.88	0.90	0.91	0.93	0.94
$F_{21}(\cdot)$									
50	2	0.84	0.83	0.89	0.87	0.90	0.90	0.95	0.93
	5	0.78	0.83	0.86	0.86	0.86	0.90	0.92	0.93
100	2	0.87	0.87	0.88	0.87	0.92	0.92	0.93	0.93
	5	0.85	0.86	0.89	0.89	0.91	0.91	0.94	0.94
200	2	0.88	0.87	0.88	0.87	0.93	0.92	0.93	0.93
	5	0.85	0.86	0.87	0.87	0.90	0.91	0.92	0.92
$F_{23}(\cdot)$									
50	2	0.84	0.85	0.90	0.90	0.90	0.91	0.95	0.95
	5	0.80	0.85	0.88	0.90	0.87	0.92	0.94	0.95
100	2	0.89	0.88	0.90	0.89	0.93	0.93	0.95	0.94
	5	0.85	0.88	0.88	0.88	0.92	0.93	0.94	0.94
200	2	0.90	0.87	0.90	0.89	0.95	0.94	0.96	0.94
	5	0.87	0.89	0.89	0.90	0.93	0.94	0.95	0.95

Table 2.15: Empirical coverage probabilities of the equal precision (EP), Hall-Wellner (HW), transformed equal precision (TEP), and transformed Hall-Wellner (THW) 90% and 95% confidence bands for the attainable sojourn time distributions based on 1000 simulations in setting 2.3

		Nominal level 90%				Nominal level 95%			
n	C	EP	HW	TEP	THW	EP	HW	TEP	THW
$F_{12}(\cdot)$									
50	2	0.80	0.84	0.88	0.87	0.87	0.91	0.94	0.94
	5	0.79	0.83	0.84	0.87	0.87	0.91	0.92	0.93
100	2	0.86	0.85	0.90	0.87	0.92	0.92	0.95	0.93
	5	0.83	0.84	0.86	0.86	0.90	0.92	0.92	0.93
200	2	0.88	0.86	0.88	0.86	0.94	0.91	0.94	0.92
	5	0.87	0.88	0.88	0.89	0.92	0.94	0.94	0.94
$F_{13}(\cdot)$									
50	2	0.71	0.74	0.89	0.90	0.80	0.82	0.95	0.95
	5	0.71	0.77	0.88	0.89	0.79	0.85	0.94	0.95
100	2	0.80	0.81	0.90	0.90	0.88	0.88	0.96	0.96
	5	0.78	0.81	0.88	0.89	0.85	0.88	0.93	0.95
200	2	0.85	0.85	0.89	0.88	0.91	0.90	0.95	0.95
	5	0.82	0.86	0.87	0.88	0.90	0.92	0.94	0.94
$F_{21}(\cdot)$									
50	2	0.76	0.75	0.87	0.85	0.83	0.83	0.92	0.91
	5	0.79	0.82	0.86	0.87	0.86	0.89	0.93	0.92
100	2	0.85	0.83	0.88	0.87	0.90	0.89	0.94	0.93
	5	0.84	0.86	0.87	0.89	0.91	0.92	0.94	0.94
200	2	0.87	0.86	0.88	0.86	0.92	0.92	0.93	0.92
	5	0.84	0.85	0.86	0.86	0.91	0.91	0.93	0.92
$F_{23}(\cdot)$									
50	2	0.85	0.83	0.91	0.90	0.90	0.89	0.95	0.95
	5	0.82	0.86	0.90	0.90	0.88	0.92	0.95	0.96
100	2	0.86	0.85	0.89	0.87	0.93	0.91	0.95	0.94
	5	0.86	0.88	0.89	0.90	0.93	0.94	0.94	0.95
200	2	0.87	0.85	0.89	0.86	0.93	0.92	0.94	0.93
	5	0.86	0.87	0.88	0.88	0.93	0.94	0.94	0.95

2.5 Summary

In this chapter, we consider estimation with right censored HSM processes, of which the transition intensities only depend on the present state and the duration time in the present state. We propose two simulation based algorithms to overcome the difficulty in constructing confidence bands for the semi-Markov kernel. Moreover, we show that the existing estimators for the transition probabilities of the embedded Markov chain can be inconsistent. We propose robust confidence intervals for the transition probabilities, and robust confidence bands for the sojourn time distributions.

The homogeneity assumption may not hold in many practical situations. In the human sleep process, for instance, the level of cortisol has been found to affect the transition intensities between Non-REM and REM sleep phases (Kneib and Hennerfeind, 2008). In the next chapter, we consider an extension of the HSM model, the modulated semi-Markov model, which handles the nonhomogeneity by incorporating covariates in the Cox regression form.

Chapter 3

Modulated Semi-Markov Process

3.1 Introduction

In practice, the subjects in a study often have different covariate patterns which makes the homogeneous assumption questionable. For instance, the hospitalization processes of cancer survivors diagnosed in different time periods can be rather different. One explanation for this is that treatments for cancer have been evolving over time.

In this chapter, we consider the modulated semi-Markov model (Cox, 1973) which incorporates covariates in the homogeneous semi-Markov model through the Cox regression form. This model differs from the well-studied classical Markov based regression model. It uses the duration time in the current state as the basic time scale in the baseline transition intensity function, instead of the study time since the beginning of the process. The dependence of the baseline transition intensity on the duration time makes the model fall outside the framework of Aalen's multiplicative intensity models and invalidates the usual martingale methods (Gill, 1980; Voelkel and Crowley, 1984; Andersen *et al.*, 1993; Oakes and Cui, 1993; Dabrowska *et al.*, 1994; Dabrowska, 1995).

When the underlying process is unidirectional (i.e., if state j can be reached from state h, state h can not be reached from state j), Voelkel and Crowley (1984) propose a random time change to transform the modulated semi-Markov model into the multiplicative intensity model. However, this trick does not work for bidirectional processes due to the renewal nature of the semi-Markov process. Dabrowska *et al.* (1994) and Dabrowska (1995) consider bidirectional modulated semi-Markov processes through the Cox regression form with possible time-dependent covariates.

They allow only the covariates to depend on the duration time in the current state, which excludes the covariates that depend on the study time scale. The study time scale can be very useful both practically and theoretically. For example, by including the study time as a covariate in the modulated semi-Markov model and conducting the hypothesis test on the corresponding regression parameter, we can check whether the time-homogeneity assumption of the homogeneous semi-Markov model is appropriate. In this chapter, we consider the modulated semi-Markov models with general time-dependent covariates.

The rest of this chapter is organized as follows. Section 3.2 describes the modulated semi-Markov models in the Cox regression form with general time-dependent covariates, and the corresponding estimation procedures for the regression parameters and the baseline transition intensities. In Section 3.3, we derive the asymptotic properties of the proposed estimators by using empirical process theory. We examine the methodology by simulation in Section 3.4. Section 3.5 concludes this chapter with some remarks.

3.2 Models and Estimation Procedures

We introduce the modulated semi-Markov models in the counting process formulation. Let $\tilde{N}_*^{hj}(t)$ be the total number of $h \to j$ transitions in the time interval $(0, t]$ without censoring, and $\boldsymbol{\lambda}_* = \{\lambda_*^{hj}(t) : h, j \in \mathcal{E}\}$ be the set of intensity functions of the multivariate counting process $\tilde{\mathbf{N}}_*(t) = \{\tilde{N}_*^{hj}(t) : h, j \in \mathcal{E}\}$ with respect to its self-exciting filtration $\mathcal{F}_t$, the σ-algebra generated by $\{\tilde{N}_*^{hj}(s) : h, j \in \mathcal{E}, 0 \le s \le t\}$. Let $\tilde{N}_*(t) = \sum_{h,j} \tilde{N}_*^{hj}(t)$ be the total number of transitions occurred in $(0, t]$. Suppose there are transition specific time-dependent covariates $Z(t) = \{Z^{hj}(t) : h, j \in \mathcal{E}\}$, whose association with the transition intensities is of interest. We consider the following two different specifications for $\lambda_*^{hj}(t)$.

Model 3.1. (The modulated renewal model; Cox 1973). Assume that

$$\lambda_*^{hj}(t) = \tilde{Y}_*^h(t)\alpha_{0hj}(B(t)) \exp(\theta' Z^{hj}(t)), \tag{3.2.1}$$

where $\tilde{Y}_*^h(t) = I\{\mathcal{S}(t-) = h\}$ is the 'at risk' indicator for whether the process has the potential of experiencing a transition from state h at time t, and $B(t) = t - T_{\tilde{N}_*(t-)}$ is a left continuous version of the backward-recurrence time.

Model 3.2. Assume that

$$\lambda_*^{hj}(t) = \tilde{Y}_*^h(t)\alpha_{0hj}(B(t); \tilde{N}(t-)) \exp(\theta' Z^{hj}(t)), \tag{3.2.2}$$

where $\tilde{Y}_*^h(t)$ and $B(t)$ are as in Model 3.1.

Remark 3.2.1. Dabrowska *et al.* (1994) and Dabrowska (1995) consider a special case of Model 3.1 with covariates depending on the time through the backward recurrence time only. That is,

$$\lambda_*^{hj}(t) = \tilde{Y}_*^h(t)\alpha_{0hj}(B(t)) \exp(\theta' Z_*^{hj}(B(t))).$$

Remark 3.2.2. The dependence of the transition intensities on the backward recurrence time $B(t)$ makes both Model 3.1 and Model 3.2 fall outside of the multiplicative intensity model framework. The transition intensities can not be written as the product of a predictable process times a deterministic function, which is of interest.

Remark 3.2.3. As a more general model than Model 3.1, Model 3.2 allows the baseline transition intensity function to vary after the occurrence of each transition, with the effect of covariates remaining the same.

Due to the dependence of the baseline transition intensities on the backward recurrence time $B(t)$, our proposed estimation procedures with both Model 3.1 and Model 3.2 involve the change of time scale from the study time to the duration time. Recall the two processes in the time scale of duration defined in (2.2.1) and (2.2.2):

$$N^{hj}(u) = \#\{m \geq 1 : J_{m-1} = h, J_m = j, X_m \leq u, T_m \leq C\},$$

and

$$Y^h(u) = \#\{m \geq 1 : J_{m-1} = h, X_m \geq u, T_{m-1} + u \leq C\}.$$

Since $Z^{hj}(t)$ can vary from transition to transition after converting to the time scale of duration (unless $Z^{hj}(t)$ depends only on the time through the duration at the current state, as in Dabrowska *et al.* 1994), we work with processes which count the number of sojourn times for each transition. Specifically, we define

$$N^{hj}(u; m) = 1\{J_{m-1} = h, J_m = j, X_m \leq u, T_m \leq C\}, \tag{3.2.3}$$

which indicates whether the mth transition is from h to j with duration $\leq u$ and occurs before the censoring,

$$Y^h(u; m) = 1\{J_{m-1} = h, X_m \geq u, T_{m-1} + u \leq C\}, \tag{3.2.4}$$

which indicates whether the mth transition is observed to be from state h and takes time $\geq u$. Let

$$Z^{hj}(u;m) = Z^{hj}(T_{m-1}+u). \tag{3.2.5}$$

Suppose we have n i.i.d. replicates of $\{N^{hj}(u;m), Y^h(u;m), Z^{hj}(u;m) : h, j \in \mathcal{E}, u \in [0, \mathcal{T}_0], m \geq 1\}$, say $\{N_i^{hj}(u;m), Y_i^h(u;m), Z_i^{hj}(u;m) : h, j \in \mathcal{E}, u \in [0, \mathcal{T}_0], m \geq 1\}$ for $i = 1, \ldots, n$. Here $\mathcal{T}_0$ is the time of the end of the study. For a vector z, denote $z^{\otimes l}$ as 1, z and zz' for $l = 0$, 1, and 2, respectively. Let

$$S_{hj}^{(l)}(\theta, u; m) = \frac{1}{n}\sum_i Y_i^h(u;m)\exp(\theta' Z_i^{hj}(u;m))Z_i^{hj}(u;m)^{\otimes l}$$

and

$$S_{hj}^{(l)}(\theta, u) = \sum_m S_{hj}^{(l)}(\theta, u; m)$$

for $l = 0, 1, 2$. We consider the following estimating functions for θ with Model 3.1 and Model 3.2, respectively:

$$U(\theta, \mathcal{T}_0) = \sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[Z_i^{hj}(u;m) - \frac{S_{hj}^{(1)}(\theta, u)}{S_{hj}^{(0)}(\theta, u)} \right] dN_i^{hj}(u;m) \tag{3.2.6}$$

and

$$U_2(\theta, \mathcal{T}_0) = \sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[Z_i^{hj}(u;m) - \frac{S_{hj}^{(1)}(\theta, u; m)}{S_{hj}^{(0)}(\theta, u; m)} \right] dN_i^{hj}(u;m). \tag{3.2.7}$$

Remark 3.2.4. Estimating function (3.2.6) for Model 3.1 reduces to the one proposed by Dabrowska *et al.* (1994) if the covariates depend on the time only through the backward recurrence time. Estimating function (3.2.7) can also be used with Model 3.1, but it is less efficient than (3.2.6) under Model 3.1. On the other hand, estimating function (3.2.7) is unbiased under the more general Model 3.2, and thus is more robust than (3.2.6).

Denote the estimators based on the estimating functions $U(\theta, \mathcal{T}_0)$ in (3.2.6) and $U_2(\theta, \mathcal{T}_0)$ in (3.2.7) by $\hat{\theta}$ and $\hat{\theta}_2$ respectively. We can then estimate the cumulative baseline transition intensity function

$$A_{0hj}(\tau) = \int_0^{\tau} \alpha_{0hj}(u)du \tag{3.2.8}$$

of Model 3.1 by

$$\hat{A}_{0hj}(\tau) = \sum_{i,m} \int_0^{\tau} \frac{dN_i^{hj}(u;m)}{nS_{hj}^{(0)}(\hat{\theta}, u)} \tag{3.2.9}$$

for $h, j \in \mathcal{E}$ and $\tau \in [0, \mathcal{T}_0]$. Under Model 3.2, we can estimate the cumulative baseline transition intensity function

$$A_{0hj}(\tau; m) = \int_0^\tau \alpha_{0hj}(u; m)du \tag{3.2.10}$$

with

$$\hat{A}_{0hj}(\tau; m) = \sum_i \int_0^\tau \frac{dN_i^{hj}(u; m)}{nS_{hj}^{(0)}(\hat{\theta}_2, u; m)} \tag{3.2.11}$$

for $h, j \in \mathcal{E}$, $m \in \mathbb{N}$, and $\tau \in [0, \mathcal{T}_0]$.

Estimating functions $U(\theta, \mathcal{T}_0)$ in (3.2.6) and $U_2(\theta, \mathcal{T}_0)$ in (3.2.7) can be justified as follows:

Proposition 3.2.1. *Estimating functions (3.2.6) and (3.2.7) are the the score functions of the profile likelihoods with Model 3.1 and Model 3.2, respectively.*

Proof. The log-likelihood function for the observed data is given by

$$\log L(\theta, \alpha) = \sum_i \sum_{h,j} \left[\int_0^{\mathcal{T}_0} \log \lambda_i^{hj}(t) d\tilde{N}_i^{hj}(t) - \int_0^{\mathcal{T}_0} \lambda_i^{hj}(t) dt \right].$$

Under Model 3.1, the log-likelihood function is specified as

$$\begin{aligned}
\log L(\theta, \alpha) = \sum_i \sum_{h,j} \Big[& \int_0^{\mathcal{T}_0} \log \alpha_{0hj}(B_i(t)) d\tilde{N}_i^{hj}(t) + \int_0^{\mathcal{T}_0} \theta' Z_i^{hj}(t) d\tilde{N}_i^{hj}(t) \\
& - \int_0^{\mathcal{T}_0} \tilde{Y}_i^h(t) e^{\theta' Z_i^{hj}(t)} \alpha_{0hj}(B_i(t)) dt \Big] \\
= \sum_i \sum_{h,j} \Big[& \int_0^{\mathcal{T}_0} \log \alpha_{0hj}(u) dN_i^{hj}(u) + \sum_m \int_0^{\mathcal{T}_0} \theta' Z_i^{hj}(u; m) dN_i^{hj}(u; m) \\
& - \sum_m \int_0^{\mathcal{T}_0} Y_i^h(u; m) e^{\theta' Z_i^{hj}(u;m)} dA_{0hj}(u) \Big],
\end{aligned}$$

by changing the time scale from the study time to the duration time. To maximize the log-likelihood function, $A_{0hj}(.)$ should be a step function which changes only when $N_\cdot^{hj}(\cdot)$ jumps, i.e., at the observed sojourn times of the transitions from state h to state j. With fixed θ, this gives us

$$\hat{A}_{0hj}(u) = \sum_{i,m} \int_0^u \frac{dN_i^{hj}(v; m)}{\sum_{i,m} Y_i^h(v; m) e^{\theta' Z_i^{hj}(v;m)}}.$$

Replacing $A_{0hj}(u)$ with $\hat{A}_{0hj}(u)$, we obtain the profile log-likelihood, $\log \tilde{L}(\theta)$, as

$$\sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[\theta' Z_i^{hj}(u; m) - \log \left(\sum_{i,m} Y_i^h(u; m) e^{\theta' Z_i^{hj}(u;m)} \right) \right] dN_i^{hj}(u; m).$$

Estimating function (3.2.6) is $\partial \log \tilde{L}(\theta)/\partial\theta$.

Under Model 3.2, the log-likelihood function simplifies to

$$\begin{aligned}\log L_2(\theta,\alpha) &= \sum_i \sum_{h,j} \Big[\int_0^{\mathcal{T}_0} \log \alpha_{0hj}(B_i(t); \tilde{N}(t-)) d\tilde{N}_i^{hj}(t) + \int_0^{\mathcal{T}_0} \theta' Z_i^{hj}(t) d\tilde{N}_i^{hj}(t) \\ &\qquad - \int_0^{\mathcal{T}_0} \tilde{Y}_i^h(t) e^{\theta' Z_i^{hj}(t)} \alpha_{0hj}(B_i(t); \tilde{N}(t-)) dt \Big] \\ &= \sum_i \sum_{h,j} \sum_m \Big[\int_0^{\mathcal{T}_0} \log \alpha_{0hj}(u;m) dN_i^{hj}(u;m) \\ &\qquad + \int_0^{\mathcal{T}_0} \theta' Z_i^{hj}(u;m) dN_i^{hj}(u;m) \\ &\qquad - \int_0^{\mathcal{T}_0} Y_i^h(u;m) e^{\theta' Z_i^{hj}(u;m)} dA_{0hj}(u;m) \Big],\end{aligned}$$

by changing the time scale from the study time to the duration time. To maximize the log-likelihood function, $A_{0hj}(.;m)$ should be a step function which changes only when $N_\cdot^{hj}(\cdot;m)$ jumps, i.e., at the observed sojourn times of the mth transitions that are from state h to state j. For fixed θ, this gives us

$$\hat{A}_{0hj}(u;m) = \sum_i \int_0^u \frac{dN_i^{hj}(v;m)}{\sum_i Y_i^h(v;m) e^{\theta' Z_i^{hj}(v;m)}}.$$

Replacing $A_{0hj}(u;m)$ with $\hat{A}_{0hj}(u;m)$, we get the profile log-likelihood, $\log \tilde{L}_2(\theta)$, as

$$\sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[\theta' Z_i^{hj}(u;m) - \log \left(\sum_{i,m} Y_i^h(u;m) e^{\theta' Z_i^{hj}(u;m)} \right) \right] dN_i^{hj}(u;m).$$

Estimating function (3.2.7) is $\partial \log \tilde{L}_2(\theta)/\partial\theta$.

□

3.3 Asymptotic Properties

3.3.1 Preliminaries

The usual derivation of asymptotic properties with multiplicative intensity models, in which the baseline intensity function is a deterministic function of the study time, relies on the counting process martingales with respect to filtrations on the study time scale. However, as indicated in Section 3.2, due to the dependence of the baseline transition intensities on the backward recurrence time $B(t)$, our estimation procedures with both Model 3.1 and Model 3.2 involve the change of time scale from the study time to the duration time. Thus the usual counting process martingale theory can not be applied directly.

In what follows, we transform Model 3.2 to a multiplicative intensity model by random time changes according to each transition (Voelkel and Crowley, 1984; Chang and Hsiung, 1994). The counting process martingales, with random time changes, can then be used to derive the asymptotic properties of the estimator based on estimating function (3.2.7). However, this trick can not be used for the estimator based on estimating function (3.2.6) with Model 3.1, because of the common baseline intensity function shared by all transitions. Instead, we use empirical process theory to derive its asymptotic properties.

Let $\tilde{N}^{hj}(t) = \tilde{N}_*^{hj}(t \wedge C)$ be the total number of observed $h \to j$ transitions in the time interval $(0, t]$ in the presence of censoring. Then the intensity function of the observed multivariate counting process $\tilde{\mathbf{N}}(\cdot) = \{\tilde{N}^{hj}(t) : h \neq j \in \mathcal{E}, t \geq 0\}$, with respect to its self-exciting filtration $\mathcal{F}_t$, is given by $\{\lambda^{hj}(t) : h \neq j \in \mathcal{E}, t \geq 0\}$ with

$$\lambda^{hj}(t) = \tilde{Y}^h(t)\alpha_{0hj}(B(t)) \exp(\theta' Z^{hj}(t))$$

under Model 3.1, and

$$\lambda^{hj}(t) = \tilde{Y}^h(t)\alpha_{0hj}(B(t); \tilde{N}(t-)) \exp(\theta' Z^{hj}(t))$$

under Model 3.2, where

$$\tilde{Y}^h(t) = I\{\mathcal{S}(t-) = h, C \geq t\}$$

indicates whether the process $\mathcal{S}(\cdot)$ is under observation and in state h just before time t.

Define

$$\tilde{M}^{hj}(t) = \tilde{N}^{hj}(t) - \int_0^t \lambda^{hj}(s)ds,$$

which is a counting process martingale with respect to $\mathcal{F}_t$. Then $\tilde{M}^{hj}(T_m + u)$ is an $\mathcal{F}_{T_m+u}$ martingale. This together with the fact that $X_{m+1} = T_{m+1} - T_m$ is an $\mathcal{F}_{T_m+u}$ stopping time ensures that

$$M^{hj}(u; m) = \tilde{M}^{hj}(T_m + u \wedge X_{m+1}) - \tilde{M}^{hj}(T_m)$$

is also an $\mathcal{F}_{T_m+u}$ martingale. By some algebra and Lemma 1.2.3,

$$M^{hj}(u; m) = N^{hj}(u; m) - \int_0^u Y^h(v; m) e^{\theta_0' Z^{hj}(v;m)} \alpha_{0hj}(v)dv,$$

under Model 3.1, and

$$M^{hj}(u;m) = N^{hj}(u;m) - \int_0^u Y^h(v;m) e^{\theta_0' Z^{hj}(v;m)} \alpha_{0hj}(v;m) dv,$$

under Model 3.2, where $N^{hj}(u;m)$ and $Y^h(u;m)$ are defined in (3.2.3) and (3.2.4), respectively. Thus for each $m \in \mathbb{N}$, the intensity of the multivariate counting process $\{N^{hj}(u;m) : h, j \in \mathcal{E}, u \geq 0\}$, with respect to the filtration $\mathcal{F}_{T_m+u}$, has a multiplicative form. The integrand in estimating function (3.2.7) is a predictable process relative to the filtration $\mathcal{F}_{T_m+u}$. Thus the counting process martingale theory can be applied to derive the asymptotic properties of the estimator based on estimating function (3.2.7). However, this approach does not work for estimating function (3.2.6) under Model 3.1, since the integrand in estimating function (3.2.6) is not a predictable process relative to the filtration $\mathcal{F}_{T_m+u}$.

3.3.2 Consistency and Asymptotic Normality

We first introduce some new notation. Let

$$s_{hj}^{(l)}(\theta, u; m) = E\left(S_{hj}^{(l)}(\theta, u; m)\right), \tag{3.3.1}$$

and

$$s_{hj}^{(l)}(\theta, u) = \sum_m s_{hj}^{(l)}(\theta, u; m). \tag{3.3.2}$$

Let also

$$\Sigma(\theta_0, \mathcal{T}_0) = \sum_{h,j} \int_0^{\mathcal{T}_0} \left(s_{hj}^{(2)}(\theta_0, u) - \frac{\left(s_{hj}^{(1)}(\theta_0, u)\right)^{\otimes 2}}{s_{hj}^{(0)}(\theta_0, u)} \right) \alpha_{0hj}(u) du, \tag{3.3.3}$$

and

$$\Sigma_2(\theta_0, \mathcal{T}_0) = \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left(s_{hj}^{(2)}(\theta_0, u; m) - \frac{\left(s_{hj}^{(1)}(\theta_0, u; m)\right)^{\otimes 2}}{s_{hj}^{(0)}(\theta_0, u; m)} \right) \alpha_{0hj}(u; m) du. \tag{3.3.4}$$

We assume the following regularity conditions:

(a) There exists a constant K such that the total variation $|Z_i^{hj}(0)| + \int_0^{\mathcal{T}_0} |dZ_i^{hj}(u)| \leq K$ for all $h, j \in \mathcal{E}$ and $1 \leq i \leq n$, where the two $|.|$'s denote the L_1-norm for a p-dimensional vector and L_1-type total variation for a p–dimensional vector function respectively.

(b) $E\{[N^{hj}(\mathcal{T}_0)]^3\} < \infty$ for all $h, j \in \mathcal{E}$;

(c) $\int_0^{\mathcal{T}_0} \alpha_{0hj}(u)du < \infty$ and $P(Y^h(\mathcal{T}_0) > 0) > 0$ for all $h, j \in \mathcal{E}$;

(d) $\Sigma(\theta_0, \mathcal{T}_0)$ and $\Sigma_2(\theta_0, \mathcal{T}_0)$ are positive definite.

The asymptotic properties of the estimator derived from estimating function (3.2.7) under Model 3.2 can be established by the counting process martingale theory, following the lines of Chang and Hsiung (1994). We omit the details and simply state the results in the following.

Theorem 3.3.1. *The estimator $\hat{\theta}_2$ of θ_0 from estimating function (3.2.7) under Model 3.2 is asymptotically efficient, and $n^{1/2}(\hat{\theta}_2 - \theta_0)$ is asymptotically normal with mean 0 and variance $(\Sigma_2(\theta_0, \mathcal{T}_0))^{-1}$ in the limiting distribution.*

Remark 3.3.1. Estimating function (3.2.7) can also be used for Model 3.1, the asymptotic normality of $\hat{\theta}_2$ still holds. However, it is not asymptotically efficient under Model 3.1.

In what follows, we show the asymptotic properties of the estimator derived from estimating function (3.2.6) with Model 3.1.

Theorem 3.3.2. *Under the regularity conditions (a)–(d), the estimator $\hat{\theta}$ from estimating function (3.2.6) is strongly consistent.*

Proof. Let

$$X_1(\theta, \mathcal{T}_0) = \frac{1}{n} \sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[(\theta - \theta_0)' Z_i^{hj}(u; m) - \log \frac{S_{hj}^{(0)}(\theta, u)}{S_{hj}^{(0)}(\theta_0, u)} \right] dN_i^{hj}(u; m).$$

We first show $X_1(\theta, \mathcal{T}_0)$ has the same limit as

$$X_2(\theta, \mathcal{T}_0) = \frac{1}{n} \sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[(\theta - \theta_0)' Z_i^{hj}(u; m) - \log \frac{s_{hj}^{(0)}(\theta, u)}{s_{hj}^{(0)}(\theta_0, u)} \right] dN_i^{hj}(u; m).$$

In a neighborhood Θ of θ_0, we have, by the uniform strong law of large numbers (Pollard, 1990, Theorem 8.3, page 41),

$$\sup_{u \in [0, \mathcal{T}_0]} \left| S_{hj}^{(0)}(\theta, u) - s_{hj}^{(0)}(\theta, u) \right| \xrightarrow{a.s.} 0, \quad \text{for } \theta \in \Theta.$$

According to condition (c), $s_{hj}^{(0)}(\theta, u)$ is bounded away from 0 on $\Theta \times [0, \mathcal{T}_0]$. Thus

$$\sup_{u \in [0, \mathcal{T}_0]} \left| \log\{S_{hj}^{(0)}(\theta, u)\} - \log\{s_{hj}^{(0)}(\theta, u)\} \right| \xrightarrow{a.s.} 0.$$

Denote $\mu_{hj}(u) = E\{N^{hj}(u)\}$ for $u \in [0, \mathcal{T}_0]$. Then $\mu_{hj}(u)$ is nondecreasing and bounded on $[0, \mathcal{T}_0]$ by condition (b). Applying the uniform strong law of large numbers (Pollard, 1990, Theorem 8.3, page 41), we have

$$\sup_{u \in [0,\mathcal{T}_0]} \left| \frac{1}{n} \sum_i N_i^{hj}(u) - \mu_{hj}(u) \right| \xrightarrow{a.s.} 0.$$

By Lemma 1 of Lin *et al.* (2000), page 724,

$$\left| \int_0^{\mathcal{T}_0} \left[\log\{S_{hj}^{(0)}(\theta, u)\} - \log\{s_{hj}^{(0)}(\theta, u)\} \right] \frac{1}{n} \sum_i dN_i^{hj}(u) \right| \xrightarrow{a.s.} 0.$$

Thus we have shown that $|X_1(\theta, \mathcal{T}_0) - X_2(\theta, \mathcal{T}_0)| \xrightarrow{a.s.} 0$.

Now denote

$$A_n(\theta, \mathcal{T}_0) = \frac{1}{n} \sum_{h,j} \int_0^{\mathcal{T}_0} \left[(\theta - \theta_0)' S_{hj}^{(1)}(\theta_0, u) - \log \left\{ \frac{s_{hj}^{(0)}(\theta, u)}{s_{hj}^{(0)}(\theta_0, u)} \right\} S_{hj}^{(0)}(\theta_0, u) \right] \alpha_{0hj}(u) du.$$

Then

$$\begin{aligned} &X_2(\theta, \mathcal{T}_0) - A_n(\theta, \mathcal{T}_0) \\ &\quad = \frac{1}{n} \sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[(\theta - \theta_0)' Z_i^{hj}(u; m) - \log \frac{s_{hj}^{(0)}(\theta, u)}{s_{hj}^{(0)}(\theta_0, u)} \right] dM_i^{hj}(u; m), \end{aligned}$$

which converges to 0 almost surely by strong law of large numbers. By condition (c) and uniform convergence of $S_{hj}^{(l)}$ to $s_{hj}^{(l)}$, we know that $A_n(\theta, \mathcal{T}_0)$ converges almost surely to

$$A(\theta, \mathcal{T}_0) = \frac{1}{n} \sum_{h,j} \int_0^{\mathcal{T}_0} \left[(\theta - \theta_0)' s_{hj}^{(1)}(\theta_0, u) - \log \left\{ \frac{s_{hj}^{(0)}(\theta, u)}{s_{hj}^{(0)}(\theta_0, u)} \right\} s_{hj}^{(0)}(\theta_0, u) \right] \alpha_{0hj}(u) du.$$

We can see from their second derivatives that both $X_1(\theta, \mathcal{T}_0)$ and $A(\theta, \mathcal{T}_0)$ are concave functions of θ . In addition, $\partial A(\theta_0, \mathcal{T}_0)/\partial\theta = 0$ and $\partial^2 A(\theta_0, \mathcal{T}_0)/\partial\theta^2 = -\Sigma(\theta_0, \mathcal{T}_0)$, which is strictly negative definite. Thus $A(\theta, \mathcal{T}_0)$ has a unique maximizer θ_0, and $\hat{\theta}$ is strongly consistent with θ_0 (Andersen and Gill, 1982). □

Under Model 3.1, we have

$$U(\theta_0, \mathcal{T}_0) = \sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[Z_i^{hj}(u; m) - \bar{Z}_{hj}(\theta_0, u) \right] dM_i^{hj}(u; m), \tag{3.3.5}$$

where

$$\bar{Z}_{hj}(\theta_0, u) = \frac{\sum_{i,m} Y_i^h(u;m) e^{\theta_0' Z_i^{hj}(u;m)} Z_i^{hj}(u;m)}{\sum_{i,m} Y_i^h(u;m) e^{\theta_0' Z_i^{hj}(u;m)}} = \frac{S_{hj}^{(1)}(\theta_0, u)}{S_{hj}^{(0)}(\theta_0, u)}. \tag{3.3.6}$$

Let

$$W_{hj}^{(0)}(\theta, u) = n^{-1/2} \sum_{i,m} M_i^{hj}(u;m),$$

and

$$W_{hj}^{(1)}(\theta, u) = n^{-1/2} \sum_{i,m} \int_0^t Z_i^{hj}(v;m) dM_i^{hj}(v;m).$$

Then

$$n^{-1/2} U(\theta_0, u) = \sum_{h,j} W_{hj}^{(1)}(\theta_0, u) - \sum_{h,j} \int_0^u \bar{Z}_{hj}(\theta_0, v) W_{hj}^{(0)}(\theta_0, dv). \tag{3.3.7}$$

Lemma 3.3.3. *Under the regularity conditions (a)–(d),* $\{W_{hj}^{(0)}(\theta_0, u), W_{hj}^{(1)}(\theta_0, u) : h, j \in \mathcal{E}, u \in [0, \mathcal{T}_0]\}$ *converges weakly to* $\{\mathcal{W}_{hj}^{(0)}(\theta_0, u), \mathcal{W}_{hj}^{(1)}(\theta_0, u) : h, j \in \mathcal{E}, u \in [0, \mathcal{T}_0]\}$*, which is a mean 0 Gaussian process with continuous sample paths and covariance functions*

$$\text{cov}\left(\mathcal{W}_{hj}^{(0)}(\theta_0, u_1), \mathcal{W}_{kl}^{(0)}(\theta_0, u_2)\right) = \int_0^{u_1 \wedge u_2} s_{hj}^{(0)}(\theta_0, u) \alpha_{0hj}(u) du,$$

$$\text{cov}\left(\mathcal{W}_{hj}^{(0)}(\theta_0, u_1), \mathcal{W}_{kl}^{(1)}(\theta_0, u_2)\right) = \int_0^{u_1 \wedge u_2} s_{hj}^{(1)}(\theta_0, u) \alpha_{0hj}(u) du,$$

and

$$\text{cov}\left(\mathcal{W}_{hj}^{(1)}(\theta_0, u_1), \mathcal{W}_{kl}^{(1)}(\theta_0, u_2)\right) = \int_0^{u_1 \wedge u_2} s_{hj}^{(2)}(\theta_0, u) \alpha_{0hj}(u) du.$$

Proof. Since $W_{hj}^{(0)}$ can be viewed as a component of $W_{hj}^{(1)}$ when the corresponding component of the covariate Z^{hj} is 1, we only need to show the convergence for $W_{hj}^{(1)}(\theta_0, .)$. Let

$$f_i(\theta_0, u) = n^{-1/2} \sum_m \int_0^u Z_i^{hj}(v;m) dM_i^{hj}(v;m),$$

then $W_{hj}^{(1)}(\theta_0, .) = \sum_k f_i(\theta_0, .)$. Without loss of generality, we assume the dimension of Z^{hj} is 1. To apply the functional central limit theorem (Pollard, 1990, Theorem 10.6, page 53) to show the weak convergence, we verify conditions $(i) - (v)$ of the theorem first.

To verify condition (i), note that

$$\begin{aligned} f_i(\theta_0, u) =& n^{-1/2} \sum_m \int_0^u Z_i^{hj}(v;m) dN_i^{hj}(v;m) \\ &- n^{-1/2} \sum_m \int_0^u Z_i^{hj}(v;m) Y_i^h(v;m) e^{\theta_0' Z_i^{hj}(v;m)} \alpha_{0hj}(v) dv. \end{aligned}$$

If $Z_i^{hj}(v;m) \geq 0$, then $f_i(\theta_0, u)$ is the difference of two monotone functions in u and thus manageable (Bilias *et al.*, 1997, Lemma A.1 and A.2, page 679). This is true for general Z^{hj} by writing $Z_i^{hj}(v;m) = Z_i^{hj}(v;m)^+ - Z_i^{hj}(v;m)^-$.

By assumptions (a) and (c), we can use envelopes $F_i = n^{-1/2}(K_1 N_i^{hj}(\mathcal{T}_0) + K_2)$, where K_1 and K_2 are positive constants. Conditions (iii) and (iv) are satisfied by assumption (b). Since f_i's are i.i.d., condition (v) is trivially satisfied. Finally, condition (ii) follows by the multivariate central limit theorem.

□

The following Lemma is adopted from Lemma A.3 of Bilias *et al.* (1997), page 679.

Lemma 3.3.4. *Let f_m and g_m be two sequences of bounded functions such that, for some constant $t_0 > 0$,*

$$\lim_{m\to\infty} \sup_{\tau\in[0,t_0]} \{|f_m(\tau) - f(\tau)| + |g_m(\tau) - g(\tau)|\} = 0,$$

where f is continuous on $[0, t_0]$, and g_m has total variation bounded by a constant K, independent of m. Then

$$\lim_{m\to\infty} \sup_{\tau\in[0,t_0]} \left| \int_0^\tau f_m(u) dg_m(u) - \int_0^\tau f(u) dg(u) \right| = 0, \tag{3.3.8}$$

$$\lim_{m\to\infty} \sup_{\tau\in[0,t_0]} \left| \int_0^\tau g_m(u) df_m(u) - \int_0^\tau g(u) df(u) \right| = 0. \tag{3.3.9}$$

Proof. Since $\{g_m : m \geq 1\}$ converges uniformly to g, and the total variation of g_m is bounded by K for all m, the total variation of g must be also bounded by K. Write

$$\begin{aligned} \int_0^\tau f_m(u) dg_m(u) - \int_0^\tau f(u) dg(u) =& \int_0^\tau [f_m(u) - f(u)] dg_m(u) \\ &+ \left[\int_0^\tau f(u) dg_m(u) - \int_0^\tau f(u) dg(u) \right], \end{aligned}$$

since $\{f_m : m \geq 1\}$ converges uniformly to f and g_m has bounded variation, we have

$$\lim_{m\to\infty} \sup_{\tau\in[0,t_0]} \left| \int_0^\tau [f_m(u) - f(u)] dg_m(u) \right| = 0. \tag{3.3.10}$$

Because f is continuous on $[0, t_0]$, for any $\epsilon > 0$, we can find a partition $0 = \tau_0 < \tau_1 < \cdots < \tau_{m_0} = t_0$ such that

$$\sup_{\tau\in[0,t_0]} |f_\epsilon(\tau) - f(\tau)| < \epsilon,$$

where $f_\epsilon(\tau) = \sum_{h=1}^{m_0} f(\tau_h) I_{\tau\in(\tau_{h-1},\tau_h]}$ for $\tau \in (0, t_0]$ and $f_\epsilon(0) = f(0)$. Note that

$$\begin{aligned}
&\left| \int_0^\tau f(u) dg_m(u) - \int_0^\tau f(u) dg(u) \right| \\
&\qquad \leq \left| \int_0^\tau [f(u) - f_\epsilon(u)] dg_m(u) \right| + \left| \int_0^\tau f_\epsilon(u) [dg_m(u) - dg(u)] \right| \\
&\qquad\qquad + \left| \int_0^\tau [f(u) - f_\epsilon(u)] dg(u) \right| \\
&\qquad \leq 2K\epsilon + 2 \sum_{h=1}^{m_0} |f(\tau_h)| \sup_{\tau\in[0,t_0]} |g_m(\tau) - g(\tau)| \\
&\qquad \longrightarrow 2K\epsilon \text{ as } m \to \infty.
\end{aligned}$$

Since ϵ can be arbitrarily small, we have

$$\lim_{m\to\infty} \sup_{\tau\in[0,t_0]} \left| \int_0^\tau f(u) dg_m(u) - \int_0^\tau f(u) dg(u) \right| = 0 \tag{3.3.11}$$

From (3.3.10) and (3.3.11) we have (3.3.8). Finally, (3.3.9) follows from (3.3.8) after applying integration by parts. □

Theorem 3.3.5. *Under the regularity conditions (a)–(d),* $\left\{n^{-1/2} U(\theta_0, u) : u \in [0, \mathcal{T}_0]\right\}$ *converges in distribution to a mean 0 Gaussian process* $\mathcal{U}(\cdot)$ *with continuous sample paths and the covariance function*

$$\operatorname{cov}\left(\mathcal{U}(u_1), \mathcal{U}(u_2)\right) = \sum_{h,j} \int_0^{u_1\wedge u_2} \left[s_{hj}^{(2)}(\theta_0, u) - \left(s_{hj}^{(1)}(\theta_0, u)\right)^{\otimes 2} / s_{hj}^{(0)}(\theta_0, u) \right] \alpha_{0hj}(u) du.$$

Proof. We have $n^{-1/2} U(\theta_0, u) = \sum_{h,j} W_{hj}^{(1)}(\theta_0, u) - \sum_{h,j} \int_0^u \bar{Z}_{hj}(\theta_0, v) W_{hj}^{(0)}(\theta_0, dv)$ from (3.3.7). Define

$$\bar{z}_{hj}(\theta_0, u) = \frac{s_{hj}^{(1)}(\theta_0, u)}{s_{hj}^{(0)}(\theta_0, u)}.$$

We shall show that $n^{-1/2}U(\theta_0, u)$ converges weakly to

$$\sum_{h,j} \mathcal{W}_{hj}^{(1)}(\theta_0, u) - \sum_{h,j} \int_0^u \bar{z}_{hj}(\theta_0, v)\mathcal{W}_{hj}^{(0)}(\theta_0, dv).$$

By the uniform strong law of large numbers (Pollard, 1990, Theorem 8.3, page 41),

$$\sup_{u \in [0, \mathcal{T}_0]} \left| S_{hj}^{(l)}(\theta_0, u) - s_{hj}^{(l)}(\theta_0, u) \right| \to 0$$

almost surely for $l = 0, 1$. Thus

$$\sup_{u \in [0, \mathcal{T}_0]} \left| \bar{Z}_{hj}(\theta_0, u) - \bar{z}_{hj}(\theta_0, u) \right| \to 0.$$

By Lemma 1 and the almost sure representation theorem (Pollard, 1990, Theorem 9.4, page 45), there is a new probability space such that

$$\left\{ W_{hj}^{(0)}(\theta_0, .), W_{hj}^{(1)}(\theta_0, .), \bar{Z}_{hj}(\theta_0, .) \right\} \longrightarrow \left\{ \mathcal{W}_{hj}^{(0)}(\theta_0, .), \mathcal{W}_{hj}^{(1)}(\theta_0, .), \bar{z}_{hj}(\theta_0, .) \right\}$$

almost surely. By condition (b), $\bar{Z}_{hj}(\theta_0, .)$ has bounded total variation. Applying Lemma 3.3.4, we have

$$\sup_{u \in [0, \mathcal{T}_0]} \left| \int_0^u \bar{Z}_{hj}(\theta_0, v) W_{hj}^{(0)}(\theta_0, dv) - \int_0^u \bar{z}_{hj}(\theta_0, v)\mathcal{W}_{hj}^{(0)}(\theta_0, dv) \right| \to 0$$

for all $h, j \in \mathcal{E}$. Thus $n^{-1/2}U(\theta_0, u)$ converges uniformly to

$$\sum_{h,j} \mathcal{W}_{hj}^{(1)}(\theta_0, u) - \sum_{h,j} \int_0^u \bar{z}_{hj}(\theta_0, v)\mathcal{W}_{hj}^{(0)}(\theta_0, dv).$$

almost surely in the new probability space and thus weakly in the original probability space. The covariance function calculation follows from Lemma 3.3.3. □

Theorem 3.3.6. *Under the regularity conditions (a)–(d),*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N\left(0, \Sigma^{-1}(\theta_0, \mathcal{T}_0)\right),$$

where $\Sigma(\theta_0, \mathcal{T}_0)$ *is given in (3.3.3).*

Proof. By Taylor's theorem,

$$n^{-1/2}U(\theta_0, \mathcal{T}_0) = [-n^{-1}\partial U(\theta^*, \mathcal{T}_0)/\partial\theta] n^{1/2}(\hat{\theta} - \theta_0),$$

where θ^* is on the line segment between $\hat{\theta}$ and θ_0. By the law of large numbers, consistency of $\hat{\theta}$ and continuity of $\Sigma(\theta, \mathcal{T}_0)$ in a neighborhood of θ_0,

$$-n^{-1}\partial U(\theta^*, \mathcal{T}_0)/\partial\theta \xrightarrow{P} \Sigma(\theta_0, \mathcal{T}_0).$$

So $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, \Sigma^{-1}(\theta_0, \mathcal{T}_0))$ by weak convergence of $n^{-1/2}U(\theta_0, \mathcal{T}_0)$. □

Theorem 3.3.7. *Under the regularity conditions (a)–(d), $\sqrt{n}(\hat{A}_{0hj}(\cdot) - A_{0hj}(\cdot))$ converges weakly to a mean 0 Gaussian process on $[0, \mathcal{T}_0]$.*

Proof. Recall that we estimate $A_{0hj}(\tau) = \int_0^\tau \alpha_{0hj}(u)du$ by

$$\hat{A}_{0hj}(\tau) = \sum_{i,m} \int_0^\tau \frac{dN_i^{hj}(u;m)}{nS_{hj}^{(0)}(\hat{\theta}, u)}.$$

We can decompose $\sqrt{n}\left[\hat{A}_{0hj}(\tau) - A_{0hj}(\tau)\right]$ into

$$\begin{aligned}\sqrt{n}&\left[\sum_{i,m} \int_0^\tau \frac{dN_i^{hj}(u;m)}{nS_{hj}^{(0)}(\hat{\theta}, u)} - \sum_{i,m} \int_0^\tau \frac{dN_i^{hj}(u;m)}{nS_{hj}^{(0)}(\theta_0, u)}\right] \\ &+ \sqrt{n}\left[\sum_{i,m} \int_0^\tau \frac{dN_i^{hj}(u;m)}{nS_{hj}^{(0)}(\theta_0, u)} - \int_0^\tau \alpha_{0hj}(u)du\right].\end{aligned} \tag{3.3.12}$$

The second term of (3.3.12) can be written as

$$n^{-1/2}\sum_{i,m} \int_0^\tau \frac{dM_i^{hj}(u;m)}{S_{hj}^{(0)}(\theta_0, u)} - n^{1/2}\int_0^\tau \alpha_{0hj}(u) I\left\{\sum_i Y_i^h(u) = 0\right\} du.$$

Applying Taylor's theorem, the first term of (3.3.12) equals to $-H(\theta^*, \tau)' n^{1/2}(\hat{\theta} - \theta_0)$, where θ^* is on the line segment between $\hat{\theta}$ and θ_0, and

$$H_{hj}(\theta, \tau) = \int_0^\tau \frac{\bar{Z}_{hj}(\theta, u)}{nS_{hj}^{(0)}(\theta, u)} dN^{hj}(u),$$

which converges almost surely to

$$h_{hj}(\theta, \tau) = \int_0^\tau \bar{z}_{hj}(\theta, u)\alpha_{0hj}(u)du.$$

Thus $\sqrt{n}\left[\hat{A}_{0hj}(\tau) - A_{0hj}(\tau)\right]$ is equivalent to

$$\int_0^\tau \frac{\mathcal{W}_{hj}^{(0)}(\theta_0, du)}{s_{hj}^{(0)}(\theta_0, u)} - h_{hj}(\theta_0, \tau)' \Sigma^{-1}(\theta_0, \mathcal{T}_0)\mathcal{U}_{hj}(\theta_0, \mathcal{T}_0),$$

where

$$\mathcal{U}_{hj}(\theta_0, \mathcal{T}_0) = \sum_{h,j} \mathcal{W}_{hj}^{(1)}(\theta_0, \mathcal{T}_0) - \sum_{h,j} \int_0^{\mathcal{T}_0} \bar{z}_{hj}(\theta_0, u)\mathcal{W}_{hj}^{(0)}(\theta_0, du).$$

□

3.3.3 Asymptotic Efficiency

Here we show our estimator for θ derived from estimating function (3.2.6) is asymptotically efficient under Model 3.1. For simplicity, we focus on a scalar regression parameter θ. We adapt the results of Begun *et al.* (1983) and Chang and Hsiung (1994). Their methods involve (i) the notation of a "Hellinger-differentiable (root-) density" to obtain appropriate scores for the regression parameter θ, and (ii) calculation of the "effective score" for θ. Thus the asymptotic lower bounds for estimation of θ in the presence of nuisance parameter are determined by the geometry of the scores.

The data we have, on the study time scale, are

$$\{\tilde{N}_i^{hj}(\cdot), \tilde{Y}_i^h(\cdot), Z_i^{hj}(\cdot) : h, j \in \mathcal{E}; 1 \leq i \leq n\}. \tag{3.3.13}$$

Let $\mathcal{H}$ be the parameter space for the baseline transition rate functions $\boldsymbol{\alpha} = \{\alpha_{0hj}(\cdot) : h, j \in \mathcal{E}\}$. Note that $\mathcal{H}$ is the nuisance parameter space for the estimation of θ. Let $\mathcal{P}^{(\theta,\boldsymbol{\alpha})}$ denote the probability measure specified by $\theta \in \Theta$ and $\boldsymbol{\alpha} \in \mathcal{H}$. For convenience and without lost of generality, assume $0 \in \Theta$ and $\mathcal{H}$ include the case that all $\{\alpha_{0hj}(t) : h, j \in \mathcal{E}\}$ are constant function 1's. Then by the Radon-Nikodym derivative theorem for point processes (cf. Brémaud, 1981, pages 166 and 187),

$$\frac{d\mathcal{P}^{(\theta,\boldsymbol{\alpha})}[0,t]}{d\mathcal{P}^{(0,\mathbf{1})}[0,t]} = L_n(t;\theta,\boldsymbol{\alpha}) = \prod_i \prod_{h,j} L_i^{hj}(t;\theta,\boldsymbol{\alpha}),$$

where $L_i^{hj}(t;\theta,\boldsymbol{\alpha})$ is such that

$$\begin{aligned} l_i^{hj}(t;\theta,\boldsymbol{\alpha}) &\equiv \log L_i^{hj}(t;\theta,\boldsymbol{\alpha}) \\ &= \int_0^t \log \alpha_{0hj}(B_i(s)) d\tilde{N}_i^{hj}(s) + \int_0^t \theta' Z_i^{hj}(s) d\tilde{N}_i^{hj}(s) \\ &\quad + \int_0^t \left(1 - \alpha_{0hj}(B_i(s)) e^{\theta' Z_i^{hj}(s)}\right) \tilde{Y}_i^h(s) ds. \end{aligned} \tag{3.3.14}$$

We use θ_0 and $\boldsymbol{\alpha}_0$ to denote the true value of the regression parameter θ, and the baseline transition rate functions $\boldsymbol{\alpha}$, respectively. To introduce the concept of regular estimator, let $\{\theta_n\}_{n\geq 1}$ and $\{\boldsymbol{\alpha}_n\}_{n\geq 1}$ be such that

$$\lim_{n\to\infty} |\sqrt{n}(\theta_n - \theta_0) - \delta| = 0,$$

and

$$\lim_{n\to\infty} ||\sqrt{n}(\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_0) - \boldsymbol{\beta}|| = 0,$$

where $\delta \in \Theta$, $\boldsymbol{\beta} \in \mathcal{H}$, and $||.||$ is the supremum norm.

Definition 3.3.1. Let $\check{\theta}$ be an estimator for θ_0 based on data (3.3.13). Then $\check{\theta}$ is called a regular estimator at $(\theta_0, \boldsymbol{\alpha}_0)$ if for every sequence of $(\theta_n, \boldsymbol{\alpha}_n)$ given above, $\sqrt{n}(\check{\theta} - \theta_n)$, under $(\theta_n, \boldsymbol{\alpha}_n)$, converges weakly to a distribution which depends on $(\theta_0, \boldsymbol{\alpha}_0)$, but not on any particular sequence $(\theta_n, \boldsymbol{\alpha}_n)$.

Proposition 3.3.8. *The estimator $\hat{\theta}$ based on estimating function (3.2.6) is a regular estimator of θ_0 under Model 3.1.*

The following lemma gives the Hellinger-differential of $L_{n_0}^{1/2}(t;\theta,\boldsymbol{\alpha})$ at $(\theta_0, \boldsymbol{\alpha}_0)$ for any fixed sample size n_0:

Lemma 3.3.9. *Let n_0 be a fixed integer. As n goes to infinity, $\sqrt{n}(L_{n_0}^{1/2}(t;\theta_n,\boldsymbol{\alpha}_n) - L_{n_0}^{1/2}(t;\theta_0,\boldsymbol{\alpha}_0))$ converges almost surely in L_2 to*

$$\frac{1}{2}L_{n_0}^{1/2}(t;\theta_0,\boldsymbol{\alpha}_0)\left\{\sum_{i=1}^{n_0}\sum_{h,j}\int_0^t\left[\delta Z_i^{hj}(s) + \frac{\beta_{hj}(B_i(s))}{\alpha_{0hj}(B_i(s))}\right]d\tilde{M}_i^{hj}(s)\right\}, \tag{3.3.15}$$

where $\tilde{M}_i^{hj}(t) = \tilde{N}_i^{hj}(t) - \int_0^t \tilde{Y}_i^h(s)\alpha_{0hj}(B(s))\exp(\theta' Z_i^{hj}(s))ds$.

Proof. By (3.3.14), we have

$$\begin{aligned}
&\sqrt{n}\left(\frac{L_{n_0}^{1/2}(t;\theta_n,\boldsymbol{\alpha}_n)}{L_{n_0}^{1/2}(t;\theta_0,\boldsymbol{\alpha}_0)} - 1\right)\\
&\qquad = \sqrt{n}\exp\left\{\frac{1}{2}\sum_{i=1}^{n_0}\sum_{h,j}\left[\int_0^t [\log\alpha_{hj;n}(B_i(s)) - \log\alpha_{0hj}(B_i(s))]\,d\tilde{N}_i^{hj}(s)\right.\right.\\
&\qquad\qquad + \int_0^t(\theta_n-\theta_0)' Z_i^{hj}(s)d\tilde{N}_i^{hj}(s)\\
&\qquad \left.\left. + \int_0^t\left(\alpha_{0hj}(B_i(s))e^{\theta' Z_i^{hj}(s)} - \alpha_{hj;n}(B_i(s))e^{\theta_n' Z_i^{hj}(s)}\right)\tilde{Y}_i^h(s)ds\right]\right\}.
\end{aligned}$$

Then (3.3.15) follows by Taylor's theorem.

□

According to Begun *et al.* (1983), the appropriate scores for θ have the form

$$\frac{1}{2}L_{n_0}^{1/2}(t;\theta_0,\boldsymbol{\alpha}_0)\left\{\sum_{i=1}^{n_0}\sum_{h,j}\int_0^t\left[Z_i^{hj}(s) - \frac{\beta_{hj}^*(B_i(s))}{\alpha_{0hj}(B_i(s))}\right]d\tilde{M}_i^{hj}(s)\right\},$$

where $\boldsymbol{\beta}^* = \{\beta_{hj}^*(\cdot) : h,j\in\mathcal{E}\}$ is in $\mathcal{H}$. The effective score of θ has the minimal norm. Thus the asymptotic information for estimation of θ in the presence of $\boldsymbol{\alpha}$ is given by

$$I_* = \inf\left\{E_{(\theta_0,\boldsymbol{\alpha}_0)}(\gamma_1(\mathcal{T}_0,\boldsymbol{\beta}))^2|\boldsymbol{\beta}\in\mathcal{H}\right\},$$

where

$$\gamma_1(\mathcal{T}_0, \boldsymbol{\beta}) \equiv \sum_{h,j} \int_0^{\mathcal{T}_0} \left[Z_1^{hj}(s) + \frac{\beta_{hj}(B_1(s))}{\alpha_{0hj}(B_1(s))} \right] d\tilde{M}_1^{hj}(s).$$

Proposition 3.3.10. *In Model 3.1, the asymptotic information for θ_0 is given by $I_* = \Sigma(\theta_0, \mathcal{T}_0)$.*

Proof. In this proof, all the expectations are taken with respect to $(\theta_0, \boldsymbol{\alpha}_0)$. Note that, under $(\theta_0, \boldsymbol{\alpha}_0)$,

$$\sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[Z_i^{hj}(u;m) - \frac{S_{hj}^{(1)}(\theta_0, u)}{S_{hj}^{(0)}(\theta_0, u)} \right] dM_i^{hj}(u;m)$$

and

$$\sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[\frac{S_{hj}^{(1)}(\theta_0, u)}{S_{hj}^{(0)}(\theta_0, u)} + \frac{\beta_{hj}(u)}{\alpha_{0hj}(u)} \right] dM_i^{hj}(u;m)$$

are uncorrelated and both have mean 0. So

$$\begin{aligned}
&\frac{1}{n} E \left(\sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[Z_i^{hj}(u;m) - \frac{S_{hj}^{(1)}(\theta_0, u)}{S_{hj}^{(0)}(\theta_0, u)} \right] dM_i^{hj}(u;m) \right)^2 \\
&\leq \frac{1}{n} E \left(\sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[Z_i^{hj}(u;m) + \frac{\beta_{hj}(u)}{\alpha_{0hj}(u)} \right] dM_i^{hj}(u;m) \right)^2 \\
&= E \left(\sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[Z_1^{hj}(u;m) + \frac{\beta_{hj}(u)}{\alpha_{0hj}(u)} \right] dM_1^{hj}(u;m) \right)^2 \qquad (3.3.16) \\
&= E \left(\sum_{h,j} \int_0^{\mathcal{T}_0} \left[Z_1^{hj}(s) + \frac{\beta_{hj}(B_1(s))}{\alpha_{0hj}(B_1(s))} \right] d\tilde{M}_1^{hj}(s) \right)^2 \\
&= E \left(\gamma_1(\mathcal{T}_0, \boldsymbol{\beta}) \right)^2 .
\end{aligned}$$

Let $n \to \infty$ in (3.3.16), we obtain that

$$\begin{aligned}
&E \left(\gamma_1(\mathcal{T}_0, \boldsymbol{\beta}) \right)^2 \\
&\geq \lim_{n \to \infty} \frac{1}{n} E \left(\sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[Z_i^{hj}(u;m) - \frac{S_{hj}^{(1)}(\theta_0, u)}{S_{hj}^{(0)}(\theta_0, u)} \right] dM_i^{hj}(u;m) \right)^2 \\
&= \lim_{n \to \infty} \frac{1}{n} E \left(\sum_i \sum_{h,j} \sum_m \int_0^{\mathcal{T}_0} \left[Z_i^{hj}(u;m) - \frac{s_{hj}^{(1)}(\theta_0, u)}{s_{hj}^{(0)}(\theta_0, u)} \right] dM_i^{hj}(u;m) \right)^2 \\
&= \Sigma(\theta_0, \mathcal{T}_0),
\end{aligned}$$

where equality holds when

$$\beta_{hj}(u) = -\alpha_{0hj}(u)\frac{s_{hj}^{(1)}(\theta_0, u)}{s_{hj}^{(0)}(\theta_0, u)}.$$

Thus we have

$$I_* = \inf\left\{E_{(\theta_0,\boldsymbol{\alpha}_0)}(\gamma_1(\mathcal{T}_0, \boldsymbol{\beta}))^2 | \boldsymbol{\beta} \in \mathcal{H}\right\} = \Sigma(\theta_0, \mathcal{T}_0).$$

□

The following proposition is the convolution representation theorem for regular estimators (Begun *et al.*, 1983, Theorem 3.1).

Proposition 3.3.11. *Let $\check{\theta}$ be a regular estimator of θ_0 in Model 3.1. Then the limiting distribution of $\sqrt{n}(\check{\theta}-\theta_n)$ can be represented as the convolution of a $N(0, 1/I_*)$ distribution with another distribution which depends only on $(\theta_0, \boldsymbol{\alpha}_0)$.*

Remark 3.3.2. By the above two propositions, a regular estimator of θ under model 3.1 has variance at least as large as $1/I_*$, where $I_* = \Sigma(\theta_0, \mathcal{T}_0)$. The estimator $\hat{\theta}$ is regular and has asymptotic normal distribution $N(0, 1/\Sigma(\theta_0, \mathcal{T}_0))$. Thus it is asymptotically efficient among the regular estimators.

3.4 Simulation

We simulated n realizations from a three-state modulated semi-Markov process with state 3 absorbing. Associated with each realization, there is an internal time-dependent covariate $Z(t) = t - B(t)$, where $B(t)$ is the back recurrence time. The values for the regression parameters in the two models considered above were

$$\theta_{12} = -0.5, \theta_{13} = -0.5, \theta_{21} = 0.5, \theta_{23} = 1.$$

We simulated the following two settings:

Setting 3.1. $\alpha_{012}(u) = 1.5u, \alpha_{013}(u) = u, \alpha_{021}(u) = 2, \alpha_{023}(u) = 1.$

Setting 3.2. $\alpha_{012}(u; m) = 1.5um, \alpha_{013}(u; m) = um, \alpha_{021}(u; m) = 2m, \alpha_{023}(u; m) = m.$

In each of the two settings, each realization is observed until censoring time $C = 5$. The sample size n varies from 50, 100, to 200. For each of the scenarios, we obtained 1000 replicates.

We used the two estimating functions in Section 3.2 to estimate the regression parameter θ in each of the two settings. Note that Model 3.1 is correctly specified in Setting 3.1, but fails in Setting 3.2. Model 3.2 is correctly specified in both settings. The regression parameter estimates are summarized in Table 3.1 and Table 3.2 in the two settings, respectively.

In Setting 3.1, the sample biases and standard errors of the estimators based on both estimating functions (3.2.6) and (3.2.7) are decreasing while increasing sample size n. The coverage of the 95% confidence intervals for the regression parameters is close to the nominal level, even with a sample size as small as 50. The estimator based on estimating function (3.2.6) is more efficient than the one based on estimating function (3.2.7), as Model 3.1 is correctly specified.

The estimator based on (3.2.7) continues to perform well in Setting 3.2. However, when Model 3.1 fails, the estimator based on estimating function (3.2.6) is biased, and the coverage of the 95% confidence intervals for the regression parameters is poor.

3.5 Summary

This chapter has considered an extension of the HSM model by incorporating covariates in a Cox regression form. The inclusion of the study time as a time-dependent covariate allows us to test whether the HSM model is plausible with the data. The simulation study suggests that the asymptotic approximation is adequate for relatively small sample size.

We consider a NHSM model in the next chapter, as a further extension of the HSM model. It assumes the transition intensities depend on the study time and the duration time in an unspecified way.

Table 3.1: Sample mean, bias, standard deviation (SD), length, and empirical coverage probability (CP) of the 95% confidence intervals of the regression parameters based on 1000 simulations in setting 3.1

	Model 3.1					Model 3.2				
n	Mean	Bias	SD	Length	CP	Mean	Bias	SD	Length	CP
$\theta_{12} = -0.5$										
50	-0.53	-0.03	0.30	1.19	95.4%	-0.55	-0.05	0.85	3.33	95.9%
100	-0.51	-0.01	0.20	0.79	95.3%	-0.53	-0.03	0.49	1.92	95.3%
200	-0.50	-0.00	0.14	0.54	94.8%	-0.50	-0.00	0.31	1.22	94.9%
$\theta_{13} = -0.5$										
50	-0.52	-0.02	0.37	1.46	93.4%	-0.60	-0.10	1.08	4.24	94.8%
100	-0.51	-0.01	0.25	0.97	95.6%	-0.54	-0.04	0.61	2.38	95.0%
200	-0.50	-0.00	0.17	0.67	94.7%	-0.51	-0.01	0.38	1.50	94.4%
$\theta_{21} = 0.5$										
50	0.48	-0.02	0.26	1.01	95.5%	0.56	0.06	0.69	2.71	95.3%
100	0.49	-0.01	0.17	0.68	95.6%	0.50	0.00	0.41	1.60	95.6%
200	0.49	-0.01	0.12	0.47	95.3%	0.50	-0.00	0.26	1.03	94.9%
$\theta_{23} = 1$										
50	1.05	0.05	0.24	0.96	95.5%	1.11	0.11	0.67	2.62	94.8%
100	1.03	0.03	0.16	0.63	96.6%	1.04	0.04	0.38	1.50	95.6%
100	1.01	0.01	0.11	0.43	95.7%	1.03	0.03	0.24	0.94	95.4%

Table 3.2: Sample mean, bias, standard deviation (SD), length, and empirical coverage probability (CP) of the 95% confidence intervals of the regression parameters based on 1000 simulations in setting 3.2

n	Model 3.1					Model 3.2				
	Mean	Bias	SD	Length	CP	Mean	Bias	SD	Length	CP
$\theta_{12} = -0.5$										
50	0.36	0.86	0.31	1.20	21.2%	-0.54	-0.04	0.87	3.41	95.1%
100	0.34	0.84	0.20	0.79	2.4%	-0.51	-0.01	0.51	1.99	94.7%
200	0.32	0.82	0.14	0.53	0.1%	-0.49	0.01	0.32	1.26	94.5%
$\theta_{13} = -0.5$										
50	0.40	0.90	0.37	1.47	32.4%	-0.53	-0.03	1.10	4.29	96.0%
100	0.34	0.84	0.25	0.97	9.6%	-0.54	-0.04	0.63	2.45	95.4%
200	0.33	0.83	0.17	0.65	0.3%	-0.51	-0.01	0.40	1.55	94.8%
$\theta_{21} = 0.5$										
50	1.18	0.68	0.33	1.29	42.4%	0.55	0.05	0.74	2.90	95.1%
100	1.16	0.66	0.22	0.86	15.1%	0.49	-0.01	0.44	1.74	95.5%
200	1.15	0.65	0.15	0.59	0.2%	0.49	-0.01	0.29	1.13	94.3%
$\theta_{23} = 1$										
50	1.69	0.69	0.36	1.42	54.9%	1.13	0.13	0.78	3.04	95.7%
100	1.63	0.63	0.24	0.93	24.3%	1.06	0.06	0.45	1.76	95.4%
100	1.59	0.59	0.16	0.63	3.2%	1.04	0.04	0.29	1.13	95.6%

Chapter 4

Nonhomogeneous Semi-Markov Process

4.1 Introduction

The time-homogeneity assumption of homogeneous semi-Markov models is often too restrictive in practice. For instance, empirical data have demonstrated that human sleep patterns vary during a night. Differences between the first and last thirds of the night are especially noticeable. Welcomed are more flexible models, such as nonhomogeneous semi-Markov models, which allow the transition intensities to depend on both the study and duration times.

Nonhomogeneous semi-Markov (NHSM) processes were introduced by Iosifescu Manu (1972). Janssen and De Dominicis (1984) consider the discrete time case. Lucas *et al.* (2006) propose estimation procedures based on the assumption that the transition rate functions are piecewise constant on both study and duration time scales. Mathieu *et al.* (2007) consider a parametric approach.

In this chapter, we consider several different estimation procedures with NHSM processes. We start with a piecewise constant approach based on the idea of stratifying the population according to the study time scale in Section 4.2. Particularly, we assume the nonhomogeneous cause-specific hazard functions $\alpha_{hj}(\tau; t)$ are piecewise constant with respect to the study time t and can vary arbitrarily with the duration time τ. We then propose a nonparametric estimation procedure based on a kernel method in Section 4.3. In some situations, more parsimonious models, which can lead to more efficient and interpretable inference procedures, are desirable. In Section 4.4, we consider estimation procedures based on a nonparametric multiplicative model where the transition intensities depend on the study and duration times multiplicatively, but their functional forms are left unspecified. Section 4.5 further specifies the effect of one of the two times in a parametric form, ending up to a semiparametric model such as the modulated semi-Markov model studied in Chapter 3.

We study finite sample properties of the proposed methods via simulated data in Section 4.6. Section 4.7 summarizes this chapter with some extensions.

4.2 Piecewise Constant Approach

Assume that $\alpha_{hj}(\tau;t)$ is piecewise constant with respect to t, and varies arbitrarily on τ. Specifically, we assume

$$\alpha_{hj}(\tau;t) = \sum_{l=0}^{L_{hj}-1} \alpha_{hj}(\tau|l) 1_{[t_l^{hj}, t_{l+1}^{hj})}(t),$$

where $0 = t_0^{hj} < t_1^{hj} < \cdots < t_{L_{hj}}^{hj} = \mathcal{T}_0$ is a prefixed or data-dependent partition of time interval $[0, \mathcal{T}_0]$ of interest. Note that the homogeneous SMP can be viewed as a special case with $\alpha_{hj}(\tau|l)$ independent of l.

For ease of presentation, we focus on the case with the partition $0 = t_0^{hj} < t_1^{hj} < \cdots < t_{L_{hj}}^{hj} = \mathcal{T}_0$ being the same for all h and j, denoted by $0 = t_0 < t_1 < \cdots < t_L = \mathcal{T}_0$. Partition the data into L strata, with the lth stratum including transitions that start within $[t_{l-1}, t_l)$ in the study time scale. Note that the data of the lth stratum are from a HSM process with the transition rate function $\alpha_{hj}(\tau|l)$. We can then apply the nonparametric estimation procedures proposed by Lagakos et al (1978) as follows.

First we introduce processes in the time scale of duration as follows. Let

$$N^{hj}(u|l) = \#\{m \geq 1 : J_{m-1} = h, J_m = j, X_m \leq u, T_m \leq C, T_{m-1} \in [t_{l-1}, t_l)\},$$

the number of observed transitions from h to j with duration $\leq u$ and starting within $[t_{l-1}, t_l)$, and

$$Y^h(u|l) = \#\{m \geq 1 : J_{m-1} = h, X_m \geq u, T_{m-1} + u \leq C, T_{m-1} \in [t_{l-1}, t_l)\},$$

the number of transitions from state h observed to take time $\geq u$ and start within $[t_{l-1}, t_l)$. Denote $N^{h\cdot}(u|l) = \sum_j N^{hj}(u|l)$.

Suppose that $\left\{\left(N_i^{hj}(\cdot|l), Y_i^h(\cdot|l)\right) : i = 1, \ldots, n\right\}$, n independent and identically distributed copies of $\left(N^{hj}(\cdot|l), Y^h(\cdot|l)\right)$, are available. We can estimate the semi-Markov kernal $Q_{hj}(\tau; t)$ for $t \in [t_{l-1}, t_l)$ by using the data in the lth stratum. Specifically, we can estimate the cumulative transition rate function with the Nelson-Aalan type estimator

$$\hat{A}_{hj}(\tau; t) = \int_0^\tau \frac{dN_\cdot^{hj}(u|l)}{Y_\cdot^h(u|l)}, \quad t \in [t_{l-1}, t_l)$$

and estimate the semi-Markov kernel with

$$\hat{Q}_{hj}(\tau; t) = \int_0^\tau \left(1 - \hat{H}_h(u-; t)\right) \frac{dN_\cdot^{hj}(u|l)}{Y_\cdot^h(u|l)}, \quad t \in [t_{l-1}, t_l),$$

where

$$\hat{H}_h(u; t) = 1 - \prod_{v \le u} \left(1 - \frac{\triangle N_\cdot^{h\cdot}(v|l)}{Y_\cdot^h(v|l)}\right), \quad t \in [t_{l-1}, t_l).$$

We can also estimate the transition rate functions by kernel smoothing:

$$\hat{\alpha}_{hj}(\tau; t) = \int K_b(\tau - u) \hat{A}_{hj}(du; t),$$

where $K_b(\cdot) = b^{-1}K(\cdot/b)$, $K(\cdot)$ is a kernel function and b is a bandwidth. Table 4.1 lists some commonly used kernel functions adapted from Härdle *et al.* (2004), page 41.

Table 4.1: Some commonly used kernel functions

Kernel	$K(u)$
Uniform	$\frac{1}{2}I(\|u\| \le 1)$
Triangle	$(1 - \|u\|)I(\|u\| \le 1)$
Epanechnikov	$\frac{3}{4}(1 - u^2)I(\|u\| \le 1)$
Biweight (Quartic)	$\frac{15}{16}(1 - u^2)^2 I(\|u\| \le 1)$
Triweight (Tricube)	$\frac{35}{32}(1 - u^2)^3 I(\|u\| \le 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}$
Cosine	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) I(\|u\| \le 1)$

4.3 Nonparametric Estimation

4.3.1 Estimation Procedure

Recall the processes on the duration time scale introduced in (3.2.3) and (3.2.4):

$$N^{hj}(u;m) = 1\{J_{m-1} = h, J_m = j, X_m \le u, T_m \le C\},$$

and

$$Y^h(u;m) = 1\{J_{m-1} = h, X_m \ge u, T_{m-1} + u \le C\}.$$

Suppose we have n i.i.d. replicates of $\{N^{hj}(\cdot;m), Y^h(\cdot;m), Z^{hj}(\cdot;m) : h,j \in \mathcal{E}, m \ge 1\}$, denoted by $\{N_i^{hj}(\cdot;m), Y_i^h(\cdot;m), Z_i^{hj}(\cdot;m) : h,j \in \mathcal{E}, m \ge 1\}$, $i = 1,\ldots,n$. Define

$$\bar{N}_w^{hj}(u;t) = \sum_{i,m} \int_0^u K_w(t - T_{m-1,i}) N_i^{hj}(dv;m)$$

and

$$\bar{Y}_w^h(u;t) = \sum_{i,m} K_w(t - T_{m-1,i}) Y_i^h(u;m),$$

where $K_w(\cdot) = w^{-1}K(\cdot/w)$ with a kernel function $K(\cdot)$ and a bandwidth w.

Remark 4.3.1. Note that $\bar{N}_w^{hj}(u;t)$ is a weighted sum of numbers of observed transitions from h to j with duration $\le u$ and starting around study time t, and $\bar{Y}_w^h(u;t)$ is a weighted sum of numbers of observed transitions from state h to take time $\ge u$ and starting around study time t.

We consider the kernel estimator

$$\hat{A}_{hj}(\tau;t) = \int_0^\tau \frac{\bar{N}_w^{hj}(du;t)}{\bar{Y}_w^h(u;t)} \tag{4.3.1}$$

for the cumulative transition rate function $A_{hj}(\tau;t)$, and estimate the NHSM kernel $Q_{hj}(\tau;t)$ by

$$\hat{Q}_{hj}(\tau;t) = \int_0^\tau \left(1 - \hat{H}_h(u-;t)\right) \frac{\bar{N}_w^{hj}(du;t)}{\bar{Y}_w^h(u;t)},$$

where

$$\hat{H}_h(u;t) = 1 - \prod_{v \le u} \left(1 - \frac{\triangle \bar{N}_w^{h\cdot}(dv;t)}{\bar{Y}_w^h(v;t)}\right).$$

Proposition 4.3.1. *The estimator $\hat{A}_{hj}(\tau;t)$ in (4.3.1) can be viewed as a local maximum likelihood estimator with the local constant assumption on the study time scale t.*

Proof. Recall that the log-likelihood is

$$l(\alpha) = \sum_i \sum_{h,j} \left[\int_0^\infty \log \lambda_i^{hj}(t) d\tilde{N}_i^{hj}(t) - \int_0^\infty \lambda_i^{hj}(t) dt \right],$$

where $\lambda_i^{hj}(t) = \tilde{Y}_i^h(t)\alpha_{hj}(B_i(t); U_i(t))$ with $U_i(t) = t - B_i(t)$ as the last transition time before the study time t.

We define the local likelihood for $\alpha_{hj}(\tau; t)$ at t_0 with the local constant assumption on the chronological time scale t as

$$\begin{aligned} l(\alpha; t_0, \tau) = \sum_i \sum_{h,j} \Big[& \int_0^\infty K_w(t_0 - U_i(t)) \log\left[\alpha_{hj}(B_i(t); t_0)\right] d\tilde{N}_i^{hj}(t) \\ & - \int_0^\infty K_w(t_0 - U_i(t)) \tilde{Y}_i^h(t) \alpha_{hj}(B_i(t); t_0)\, dt \Big]. \end{aligned}$$

To proceed, we change the counting processes from the study time scale to the duration time scale. After some algebra, we have

$$\begin{aligned} l(\alpha; t_0, \tau) = \sum_{i,m} \sum_{h,j} \Big[& \int_0^\infty K_w(t_0 - T_{m-1,i}) \log\left[\alpha_{hj}(u; t_0)\right] dN_i^{hj}(u; m) \\ & - \int_0^\infty K_w(t_0 - T_{m-1,i}) Y_i^h(u; m) \alpha_{hj}(u; t_0) dv \Big]. \end{aligned}$$

So the local maximum likelihood estimator for $A_{hj}(\tau; t_0) = \int_0^\tau \alpha_{hj}(u; t_0) du$ is

$$\hat{A}_{hj}(\tau; t_0) = \sum_{i,m} \int_0^\tau \frac{K_w(t_0 - T_{m-1,i}) dN_i^{hj}(u; m)}{\sum_{m,i} K_w(t_0 - T_{m-1,i}) Y_i^h(u; m)},$$

which is the one given by (4.3.1).

□

Remark 4.3.2. Working in the context of a nonparametric hazard model for survival data with time-dependent covariates, McKeague and Utikal (1990b) study Beran's (1981) estimator of the conditional cumulative hazard function, which has a functional form similar to our estimator. Their estimator can be viewed as a local likelihood estimator with the local constant assumption on the duration time scale τ, which can not be easily used to estimate the semi-Markov kernel in our context.

We can then estimate the transition rate function $\alpha_{hj}(\tau; t_0)$ by

$$\begin{aligned} \hat{\alpha}_{hj}(\tau; t_0) &= \int K_b(\tau - u) \hat{A}_{hj}(du; t_0), \\ &= \sum_{i,m} \int_0^\infty \frac{K_w(t_0 - T_{m-1,i}) K_b(\tau - u) dN_i^{hj}(u; m)}{\sum_{m,i} K_w(t_0 - T_{m-1,i}) Y_i^h(u; m)}, \end{aligned} \tag{4.3.2}$$

where b is another bandwidth.

For the sole purpose of estimating the transition rate function $\alpha_{hj}(\tau;t)$, we can construct a local maximum likelihood estimator with the local constant assumption on the both time scales. Specifically, we define the local likelihood at (t_0,τ_0) to be

$$l(\alpha;t_0,\tau_0)=\sum_i\sum_{h,j}\Big[\int_0^\infty K_{b,w}(\tau_0-B_i(t),t_0-U_i(t))\log\left[\alpha_{hj}\left(B_i(t);U_i(t)\right)\right]d\tilde{N}_i^{hj}(t)$$
$$-\int_0^\infty K_{b,w}(\tau_0-B_i(t),t_0-U_i(t))\tilde{Y}_i^h(t)\alpha_{hj}\left(B_i(t);U_i(t)\right)dt\Big],$$

where $K_{b,w}(.,.)$ is a bivariate smoothing kernel. Assuming $\alpha_{hj}(\tau;t)$ to be locally constant around (t_0,τ_0), we have

$$l(\alpha;t_0,\tau_0)=\sum_i\sum_{h,j}\Big[\log\left[\alpha_{hj}(\tau_0;t_0)\right]\int_0^\infty K_{b,w}(\tau_0-B_i(t),t_0-U_i(t))d\tilde{N}_i^{hj}(t)$$
$$-\alpha_{hj}(\tau_0;t_0)\int_0^\infty K_{b,w}(\tau_0-B_i(t),t_0-U_i(t))\tilde{Y}_i^h(t)dt\Big].$$

It is easy to show that $l(\alpha;t_0,\tau_0)$ is maximized at

$$\tilde{\alpha}_{hj}(\tau_0;t_0)=\frac{\sum_i\int_0^\infty K_{b,w}(\tau_0-B_i(t),t_0-U_i(t))d\tilde{N}_i^{hj}(t)}{\sum_i\int_0^\infty K_{b,w}(\tau_0-B_i(t),t_0-U_i(t))\tilde{Y}_i^h(t)dt}. \tag{4.3.3}$$

This estimator has also been studied by Nielsen and Linton (1995), in the context of a nonparametric hazard model for survival data with time-dependent covariates.

4.3.2 Asymptotic Properties

Let

$$F_h(t,u;m)=P\{T_{m-1,i}\le t,Y_i^h(u;m)=1\} \tag{4.3.4}$$

and $f_h(t,u;m)$ be the corresponding density with respect to Lebesgue measure. Denote $f_h(t,u)=\sum_m f_h(t,u;m)$, which we assume to be finite.

We assume that the kernel function $K(\cdot)$ is continuous with support $[-1,1]$, and is symmetric about 0. Define the kernel moments

$$\mathcal{K}_1=\int_{-1}^1 q^2K(q)dq,\quad\text{and } \mathcal{K}_2=\int_{-1}^1\left(K(q)\right)^2dq. \tag{4.3.5}$$

In what follows, we work with interior points so that we do not need to worry about boundary effects.

Lemma 4.3.2. *Suppose $\alpha_{hj}(u;t)$ is twice, and $f_h(t,u;m)$ and $f_h(t,u)$ is once continuously differentiable in t. Then for each fixed t, as $w \to 0$ and $nw \to \infty$,*

$$\frac{1}{n}\sum_{i,m} K_w(t - T_{m-1,i})Y_i^h(u;m) \xrightarrow{P} f_h(t,u), \tag{4.3.6}$$

$$\frac{w}{n}\sum_{i,m} \left(K_w(t - T_{m-1,i})\right)^2 Y_i^h(u;m)\alpha_{hj}(u;T_{m-1,i}) \xrightarrow{P} \mathcal{K}_2\alpha_{hj}(u;t)f_h(t,u), \tag{4.3.7}$$

and

$$\begin{aligned}&\frac{1}{nw^2}\sum_{i,m} K_w(t - T_{m-1,i})Y_i^h(u;m)\left[\alpha_{hj}(u;T_{m-1,i}) - \alpha_{hj}(u;t)\right]\\ &\xrightarrow{P} \mathcal{K}_1\left[\frac{\partial\alpha_{hj}}{\partial t}(u;t)\frac{\partial f_h}{\partial t}(t,u) + \frac{f_h(t,u)}{2}\frac{\partial^2\alpha_{hj}}{\partial t^2}(u;t)\right],\end{aligned} \tag{4.3.8}$$

uniformly in u, as $n \to \infty$.

Proof. We first show (4.3.6). Note that

$$\begin{aligned}E\left(\frac{1}{n}\sum_{i,m} K_w(t - T_{m-1,i})Y_i^h(u;m)\right) &= \sum_m \int K_w(t-s)f_h(s,u;m)ds\\ &= \sum_m \int K(q)f_h(t - wq, u;m)dq,\end{aligned}$$

by a change of variable $q = (t-s)/w$. By continuity of $f_h(t,u;m)$ and Lebesgue's dominated convergence theorem,

$$\sum_m \int K(q)f_h(t-wq,u;m)dq \longrightarrow \sum_m f_h(t,u;m) = f_h(t,u).$$

Then (4.3.6) follows by the uniform law of large numbers (Pollard, 1990, Theorem 8.2, page 39).

Similarly,

$$\begin{aligned}&E\left(\frac{w}{n}\sum_{i,m}\left(K_w(t - T_{m-1,i})\right)^2 Y_i^h(u;m)\alpha_{hj}(u;T_{m-1,i})\right)\\ &= \sum_m w\int \left(K_w(t-s)\right)^2\alpha_{hj}(u;s)f_h(s,u;m)ds\\ &= \sum_m \int \left(K(q)\right)^2\alpha_{hj}(u;t-wq)f_h(t-wq,u;m)dq\\ &\longrightarrow \mathcal{K}_2\alpha_{hj}(u;t)f_h(t,u),\end{aligned}$$

by continuity of $\alpha_{hj}(u;t)$ and $f_h(t,u;m)$, and Lebesgue's dominated convergence theorem. Then (4.3.7) follows by the uniform law of large numbers (Pollard, 1990, Theorem 8.2, page 39).

It remains to show (4.3.8). After a change of variable $q=(t-s)/w$,

$$\begin{aligned}
&E\left(\frac{1}{n}\sum_{i,m} K_w(t-T_{m-1,i})Y_i^h(u;m)\left[\alpha_{hj}(u;T_{m-1,i})-\alpha_{hj}(u;t)\right]\right)\\
&=\sum_m \int K_w(t-s)\left[\alpha_{hj}(u;s)-\alpha_{hj}(u;t)\right] f_h(s,u;m)ds\\
&=\sum_m \int K(q)\left[\alpha_{hj}(u;t-wq)-\alpha_{hj}(u;t)\right] f_h(t-wq,u;m)ds\\
&=\int K(q)\left[\alpha_{hj}(u;t-wq)-\alpha_{hj}(u;t)\right] f_h(t-wq,u)ds
\end{aligned}$$

By Taylor's theorem,

$$\alpha_{hj}(u;t-wq)-\alpha_{hj}(u;t)=-wq\frac{\partial\alpha_{hj}}{\partial t}(u;t)+\frac{w^2q^2}{2}\frac{\partial^2\alpha_{hj}}{\partial t^2}(u;t^*),$$

$$f_h(t-wq,u)=f_h(t,u)-wq\frac{\partial f_h}{\partial t}(t^{**},u),$$

where t^* and t^{**} are between t and $t-wq$. Since $\int_{-1}^{1}K(q)dq=0$ by symmetry of $K(\cdot)$ about 0, we have

$$\begin{aligned}
&\int K(q)\left[\alpha_{hj}(u;t-wq)-\alpha_{hj}(u;t)\right] f_h(t-wq,u)ds\\
&=w^2\mathcal{K}_1\left[\frac{\partial\alpha_{hj}}{\partial t}(u;t)\frac{\partial f_h}{\partial t}(t,u)+\frac{f_h(t,u)}{2}\frac{\partial^2\alpha_{hj}}{\partial t^2}(u;t)\right]\left[1+o(1)\right],
\end{aligned}$$

by continuity and Lebesgue's dominated convergence theorem. Now by the uniform law of large numbers (Pollard, 1990, Theorem 8.2, page 39),

$$\begin{aligned}
&\frac{1}{nw^2}\sum_{i,m} K_w(t-T_{m-1,i})Y_i^h(u;m)\left[\alpha_{hj}(u;T_{m-1,i})-\alpha_{hj}(u;t)\right]\\
&\xrightarrow{P}\mathcal{K}_1\left[\frac{\partial\alpha_{hj}}{\partial t}(u;t)\frac{\partial f_h}{\partial t}(t,u)+\frac{f_h(t,u)}{2}\frac{\partial^2\alpha_{hj}}{\partial t^2}(u;t)\right],
\end{aligned}$$

uniformly in u.

□

Let

$$A^*_{hj}(\tau;t)=\int_0^\tau\frac{\sum_{i,m}K_w(t-T_{m-1,i})Y_i^h(u;m)\alpha_{hj}(u;T_{m-1,i})du}{\sum_{i,m}K_w(t-T_{m-1,i})Y_i^h(u;m)}. \tag{4.3.9}$$

Then

$$\begin{aligned}\hat{A}_{hj}(\tau;t) &= \int_0^\tau \frac{\bar{N}_w^{hj}(du;t)}{\bar{Y}_w^h(u;t)} \\ &= \int_0^\tau \frac{\sum_{i,m} K_w(t-T_{m-1,i})dN_i^{hj}(u;m)}{\sum_{i,m} K_w(t-T_{m-1,i})Y_i^h(u;m)} \\ &= \int_0^\tau \frac{\sum_{i,m} K_w(t-T_{m-1,i})dM_i^{hj}(u;m)}{\sum_{i,m} K_w(t-T_{m-1,i})Y_i^h(u;m)} + A_{hj}^*(\tau;t).\end{aligned}$$

So we can write

$$\hat{A}_{hj}(\tau;t) - A_{hj}(\tau;t) = \left(\hat{A}_{hj}(\tau;t) - A_{hj}^*(\tau;t)\right) + \left(A_{hj}^*(\tau;t) - A_{hj}(\tau;t)\right),$$

where

$$\hat{A}_{hj}(\tau;t) - A_{hj}^*(\tau;t) = \int_0^\tau \frac{\sum_{i,m} K_w(t-T_{m-1,i})dM_i^{hj}(u;m)}{\sum_{i,m} K_w(t-T_{m-1,i})Y_i^h(u;m)} \tag{4.3.10}$$

and

$$A_{hj}^*(\tau;t) - A_{hj}(\tau;t) = \int_0^\tau \frac{\sum_{i,m} K_w(t-T_{m-1,i})Y_i^h(u;m)\left[\alpha_{hj}(u;T_{m-1,i}) - \alpha_{hj}(u;t)\right]du}{\sum_{i,m} K_w(t-T_{m-1,i})Y_i^h(u;m)}. \tag{4.3.11}$$

Theorem 4.3.3. *Suppose $\alpha_{hj}(u;t)$ is twice, and $f_h(t,u;m)$ and $f_h(t,u)$ is once continuously differentiable in t. Then for each fixed t, as $w \to 0$ and $nw \to \infty$,*

(i) $\sqrt{nw}\left(\hat{A}_{hj}(\cdot;t) - A_{hj}^*(\cdot;t)\right)$ converges weakly to a mean zero Gaussian process $\mathcal{A}(\cdot;t)$, which has independent increments and variance function

$$\mathrm{var}\left(\mathcal{A}(\tau;t)\right) = \mathcal{K}_2 \int_0^\tau \frac{\alpha_{hj}(u;t)du}{f_h(t,u)}; \tag{4.3.12}$$

(ii)

$$w^{-2}\left(A_{hj}^*(\tau;t) - A_{hj}(\tau;t)\right) \xrightarrow{P} \int_0^\tau \mathcal{K}_1 \left[\frac{\frac{\partial \alpha_{hj}}{\partial t}(u;t)\frac{\partial f_h}{\partial t}(t,u)}{f_h(t,u)} + \frac{1}{2}\frac{\partial^2 \alpha_{hj}}{\partial t^2}(u;t)\right] du. \tag{4.3.13}$$

Proof. By (4.3.10) and Lemma 4.3.2,

$$\begin{aligned}&\mathrm{Var}\left[\sqrt{nw}\left(\hat{A}_{hj}(\cdot;t) - A_{hj}^*(\cdot;t)\right)\right] \\ &\quad = nw \int_0^\tau \frac{\sum_{i,m}(K_w(t-T_{m-1,i}))^2 Y_i^h(u;m)\alpha_{hj}(u;T_{m-1,i})du}{\left[\sum_{i,m} K_w(t-T_{m-1,i})Y_i^h(u;m)\right]^2} \\ &\quad \xrightarrow{P} \mathcal{K}_2 \int_0^\tau \frac{\alpha_{hj}(u;t)du}{f_h(t,u)}.\end{aligned}$$

Using empirical process theory similar to Chapter 3, we can show that, as a function of τ for fixed t, $\sqrt{nw}\mathcal{V}_{hj}(t,\tau)$ converges weakly to a zero mean Gaussian process with independent increment and the above variance function. (ii) follows immediately by (4.3.11) and Lemma 4.3.2. □

Remark 4.3.3. Theorem 4.3.3 can be used to construct confidence intervals and confidence bands of $A_{hj}(\tau;t)$. The optimal bandwidth has order $w \sim n^{-1/5}$.

Remark 4.3.4. From (1.2.9), we can estimate the semi-Markov kernel $Q_{hj}(\tau;t)$ by

$$\hat{Q}_{hj}(\tau;t) = \int_0^\tau \exp\left[-\sum_{k\neq h}\hat{A}_{hk}(u;t)\right]\hat{A}_{hj}(du;t).$$

The asymptotic property can be derived by the functional delta method.

We now consider the asymptotic properties of $\hat{\alpha}_{hj}(\tau;t)$, the transition rate estimator defined in (4.3.2). Write

$$\begin{aligned}\hat{\alpha}_{hj}(\tau;t) &= \int K_b(\tau-u)\hat{A}_{hj}(du;t)\\ &= \int \frac{K_b(\tau-u)\sum_{i,m}K_w(t-T_{m-1,i})dN_i^{hj}(u;m)}{\sum_{i,m}K_w(t-T_{m-1,i})Y_i^h(u;m)}.\end{aligned}$$

So

$$\hat{\alpha}_{hj}(\tau;t)-\alpha_{hj}(\tau;t) = \mathcal{V}_{hj}(t,\tau)+\mathcal{B}_{hj}(t,\tau),$$

where

$$\mathcal{V}_{hj}(t,\tau) = \int \frac{K_b(\tau-u)\sum_{i,m}K_w(t-T_{m-1,i})dM_i^{hj}(u;m)}{\sum_{i,m}K_w(t-T_{m-1,i})Y_i^h(u;m)} \quad (4.3.14)$$

and

$$\begin{aligned}\mathcal{B}_{hj}(t,\tau) &= \left(\int K_b(\tau-u)\alpha_{hj}(u;t)du-\alpha_{hj}(\tau;t)\right) && (4.3.15)\\ &+\int \frac{K_b(\tau-u)\sum_{i,m}K_w(t-T_{m-1,i})Y_i^h(u;m)[\alpha_{hj}(u;T_{m-1,i})-\alpha_{hj}(u;t)]du}{\sum_{i,m}K_w(t-T_{m-1,i})Y_i^h(u;m)} && (4.3.16)\\ & && (4.3.17)\end{aligned}$$

Theorem 4.3.4. *Suppose $\alpha_{hj}(u;t)$ is twice, and $f_h(t,u;m)$ and $f_h(t,u)$ is once continuously differentiable in t. For fixed (t,τ), if $b \sim w$ and $nw^2 \to \infty$, then*

(i)

$$\sqrt{nw^2}\mathcal{V}_{hj}(t,\tau) \xrightarrow{D} N\left(0, \frac{\mathcal{K}_2^2\alpha_{hj}(\tau;t)}{f_h(t,\tau)}\right); \tag{4.3.18}$$

(ii)

$$w^{-2}\mathcal{B}_{hj}(t,\tau) \xrightarrow{P} \mathcal{K}_1\left[\frac{\frac{\partial\alpha_{hj}}{\partial t}(\tau;t)\frac{\partial f_h}{\partial t}(t,\tau)}{f_h(t,\tau)} + \frac{1}{2}\frac{\partial^2\alpha_{hj}}{\partial t^2}(\tau;t) + \frac{1}{2}\frac{\partial^2\alpha_{hj}}{\partial \tau^2}(\tau;t)\right]. \tag{4.3.19}$$

Proof. We have

$$\begin{aligned}
&\operatorname{Var}\left(\sqrt{nw^2}\mathcal{V}_{hj}(t,\tau)\right)\\
&= nw^2\int (K_b(\tau-u))^2\,\frac{\sum_{i,m}(K_w(t-T_{m-1,i}))^2Y_i^h(u;m)\alpha_{hj}(u;T_{m-1,i})du}{\left[\sum_{i,m}K_w(t-T_{m-1,i})Y_i^h(u;m)\right]^2}\\
&\xrightarrow{P} nw^2\int (K_b(\tau-u))^2\,\frac{\mathcal{K}_2\alpha_{hj}(u;t)du}{nwf_h(t,u)}, \quad \text{(by Lemma 4.3.2)}\\
&= \frac{w}{b}\int (K(v))^2\,\frac{\mathcal{K}_2\alpha_{hj}(\tau-bv;t)dv}{f_h(t,\tau-bv)}, \quad (v=(\tau-u)/b)\\
&\xrightarrow{P} \frac{\mathcal{K}_2^2\alpha_{hj}(\tau;t)}{f_h(t,\tau)},
\end{aligned}$$

by continuity and Lebesgue's dominated convergence theorem. Using empirical process theory similarly to Chapter 3, we can show that $\sqrt{nw^2}\mathcal{V}_{hj}(t,\tau)$ converges weakly to a zero mean Gaussian random variable with the above variance.

By a change of variable and Taylor's theorem, we have

$$\begin{aligned}
\int K_b(\tau-u)\alpha_{hj}(u;t)du - \alpha_{hj}(\tau;t) &= \int K(v)\left[\alpha_{hj}(\tau-bv;t)-\alpha_{hj}(\tau;t)\right]\\
&= \int K(v)\frac{b^2v^2}{2}\frac{\partial^2\alpha_{hj}}{\partial\tau^2}(\tau^*;t)\\
&= b^2\frac{\mathcal{K}_1}{2}\frac{\partial^2\alpha_{hj}}{\partial\tau^2}(\tau;t)\left[1+o(1)\right],
\end{aligned}$$

by continuity and Lebesgue's dominated convergence theorem. By Lemma 4.3.2 and Lebesgue's dominated convergence theorem,

$$\begin{aligned}
&\int \frac{K_b(\tau-u)\sum_{i,m}K_w(t-T_{m-1,i})Y_i^h(u;m)[\alpha_{hj}(u;T_{m-1,i})-\alpha_{hj}(u;t)]du}{\sum_{i,m}K_w(t-T_{m-1,i})Y_i^h(u;m)}\\
&\xrightarrow{P} w^2\int K_b(\tau-u)\mathcal{K}_1\left[\frac{\frac{\partial\alpha_{hj}}{\partial t}(u;t)\frac{\partial f_h}{\partial t}(t,u)}{f_h(t,u)} + \frac{1}{2}\frac{\partial^2\alpha_{hj}}{\partial t^2}(u;t)\right]du\\
&\xrightarrow{P} w^2\mathcal{K}_1\left[\frac{\frac{\partial\alpha_{hj}}{\partial t}(\tau;t)\frac{\partial f_h}{\partial t}(t,\tau)}{f_h(t,\tau)} + \frac{1}{2}\frac{\partial^2\alpha_{hj}}{\partial t^2}(\tau;t)\right].
\end{aligned}$$

Thus

$$w^{-2}\mathcal{B}_{hj}(t,\tau) \xrightarrow{P} \mathcal{K}_1 \left[\frac{\frac{\partial \alpha_{hj}}{\partial t}(\tau;t)\frac{\partial f_h}{\partial t}(t,\tau)}{f_h(t,\tau)} + \frac{1}{2}\frac{\partial^2 \alpha_{hj}}{\partial t^2}(\tau;t) + \frac{1}{2}\frac{\partial^2 \alpha_{hj}}{\partial \tau^2}(\tau;t) \right]$$

□

Remark 4.3.5. Theorem 4.3.4 can be used to construct confidence intervals and confidence bands of $\alpha_{hj}(\tau;t)$. The optimal bandwidth has order $w \sim n^{-1/6}$.

4.4 Nonparametric Multiplicative Model

4.4.1 Model

In this section, we consider a structured nonparametric model for the transition intensity functions, a nonparametric multiplicative model:

$$\alpha_{hj}(\tau;t) = \psi_{hj}(t+\tau)\gamma_{hj}(\tau).$$

Under the nonparametric multiplicative model, the log-likelihood reduces to

$$l(\psi,\gamma) = \sum_i \sum_{h,j} \Big[\int_0^\infty \log \psi_{hj}(t) d\tilde{N}_i^{hj}(t) + \int_0^\infty \log \gamma_{hj}(B_i(t)) d\tilde{N}_i^{hj}(t) - \int_0^\infty \tilde{Y}_i^h(t)\psi_{hj}(t)\gamma_{hj}(B_i(t))dt \Big].$$

Note that $\psi_{hj}(.)$ and $\gamma_{hj}(.)$ are not simultaneously identifiable. One may multiply one component by an arbitrary constant and divide the other component by the same component to obtain the same $\alpha_{hj}(\cdot;\cdot)$. To avoid this nonidentifiability and without loss of generality, we impose the normalization that $\gamma_{hj}(0) = 1$.

If $\gamma_{hj}(.)$ is known, we can estimate the cumulative version of $\psi_{hj}(.)$, $\Psi_{hj}(t) = \int_0^t \psi_{hj}(u)du$, by a Nelson-Aalen type estimator $\hat{\Psi}_{hj}(t)$, and then estimate $\psi_{hj}(.)$ by kernel smoothing of $\hat{\Psi}_{hj}(t)$. If $\psi_{hj}(.)$ is known, we can estimate the cumulative version of $\gamma_{hj}(.)$, $\Gamma_{hj}(\tau) = \int_0^\tau \gamma_{hj}(v)dv$, by a Nelson-Aalen type estimator after changing the time scale of counting processes from the study time to the duration time, and then estimate $\gamma_{hj}(.)$ by kernel smoothing of its estimated cumulative version. A detailed estimation procedure is presented in the following.

4.4.2 Algorithm

We propose an algorithm for estimating $\alpha_{hj}(\tau;t)$. Starting from an estimate $\hat{\gamma}_{hj}(.)$ of $\gamma_{hj}(.)$, it iteratively estimates $\psi_{hj}(.)$ and then $\gamma_{hj}(.)$ until convergence. Specifically,

Step 1. Estimate $\psi_{hj}(.)$ based on

$$l_1(\psi) = \sum_{h,j}\sum_i \left[\int_0^\infty \log\psi_{hj}(t)d\tilde{N}_i^{hj}(t) - \int_0^\infty \tilde{Y}_i^h(t)\psi_{hj}(t)\hat{\gamma}_{hj}(B_i(t))dt\right],$$

which gives

$$\hat{\psi}_{hj}(t) = \sum_i \int_0^\infty \frac{K_w(t-u)d\tilde{N}_i^{hj}(u)}{\sum_i \tilde{Y}_i^h(u)\hat{\gamma}_{hj}(B_i(u))}.$$

Step 2. Estimate $\gamma_{hj}(.)$ based on

$$l_2(\gamma) = \sum_{h,j}\sum_i \left[\int_0^\infty \log\gamma_{hj}(B_i(t))d\tilde{N}_i^{hj}(t) - \int_0^\infty \tilde{Y}_i^h(t)\hat{\psi}_{hj}(t)\gamma_{hj}(B_i(t))dt\right],$$

which leads to

$$\hat{\gamma}_{hj}(\tau) = \sum_i \int_0^\infty \frac{K_b(\tau-u)dN_i^{hj}(u)}{\sum_{m,i} Y_i^h(u;m)\hat{\psi}_{hj}(T_{m-1,i}+u)}.$$

Step 3. Iterative between Step 1 and Step 2 until both $\psi_{hj}(.)$ and $\gamma_{hj}(.)$ converge.

Remark 4.4.1. The algorithm we propose is a backfitting algorithm. Convergence of the algorithm warrants further investigation. We use the nonparametric estimate obtained from Section 4.3 as the initial value, and have not encountered convergence difficulty in our applications.

The estimation in Step 1 can be implemented using kernel smoothing of

$$\hat{\Psi}_{hj}(t) = \sum_i \int_0^t \frac{d\tilde{N}_i^{hj}(u)}{\sum_i \tilde{Y}_i^h(u)\hat{\gamma}_{hj}(B_i(u))},$$

a Nelson-Aalen type estimator of $\Psi_{hj}(t) = \int_0^t \psi_{hj}(u)du$. Step 2 can be implemented as follows. Change the time scale of the counting processes in $l_2(\gamma)$ from study time to duration time. After some algebra, we have

$$l_2(\gamma) = \sum_{h,j}\sum_i \left[\int_0^\infty \log\gamma_{hj}(\tau)dN_i^{hj}(\tau) - \int_0^\infty \sum_m Y_i^h(\tau;m)\hat{\psi}_{hj}(T_{m-1,i}+\tau)\gamma_{hj}(\tau)d\tau\right].$$

We can then estimate $\Gamma_{hj}(\tau) = \int_0^\tau \gamma_{hj}(v)dv$ by a Nelson-Aalen type estimator

$$\hat{\Gamma}_{hj}(\tau) = \sum_i \int_0^\tau \frac{dN_i^{hj}(u)}{\sum_{m,i} Y_i^h(u;m)\hat{\psi}_{hj}(T_{m-1,i}+u)}.$$

Finally, $\hat{\gamma}_{hj}(.)$ is obtained by kernel smoothing of $\hat{\Gamma}_{hj}(.)$.

4.5 Semiparametric Estimation

In this section, we consider semiparametric estimation with NHSM processes by specifying the effect of the study time or the duration time in a parametric form. Specifically, we consider the following two semiparametric models:

$$\alpha_{hj}(\tau; t) = \alpha_{0hj}(\tau) \exp(\theta' Z_{hj}(t)), \tag{4.5.1}$$

and

$$\alpha_{hj}(\tau; t) = \alpha_{0hj}(t + \tau) \exp(\theta' Z_{hj}(\tau)), \tag{4.5.2}$$

where $Z_{hj}(\cdot)$ is a deterministic vector function. Note that Model (4.5.2) has a multiplicative intensity form, and has been well studied (Andersen and Gill, 1982; Andersen *et al.*, 1993). Model (4.5.1) is a special modulated semi-Markov model we studied in Chapter 3. In what follows, we focus on estimation with Model (4.5.1), which use the duration time as the basic time scale in the baseline transition rate function.

For simplicity of notation and without loss of generality, we work on the univariate function $Z_{hj}(t) \equiv t$. Thus Model (4.5.1) simplifies to

$$\alpha_{hj}(\tau; t) = \alpha_{0hj}(\tau) \exp(\theta t).$$

Applying the methods in Chapter 3, we can get estimators $\hat{\theta}$ of θ, and $\hat{A}_{0hj}(\cdot)$ of $A_{0hj}(\cdot)$, where

$$A_{0hj}(\tau) = \int_0^\tau \alpha_{0hj}(u) du$$

is the cumulative baseline transition rate function. We can then estimate $\alpha_{0hj}(\cdot)$ by kernel smoothing:

$$\hat{\alpha}_{0hj}(\tau) = \int K_b(\tau - u) d\hat{A}_{0hj}(u).$$

Thus an estimator for the transition rate function $\alpha_{hj}(\tau; t)$ is

$$\hat{\alpha}_{hj}(\tau; t) = \hat{\alpha}_{0hj}(\tau) \exp(\hat{\theta} t).$$

From (1.2.9), we can then estimate the semi-Markov kernel $Q_{hj}(\tau; t)$ by

$$\hat{Q}_{hj}(\tau; t) = \int_0^\tau \exp\left[-\sum_{k \neq h} \hat{A}_{hk}(u; t) \right] \hat{A}_{hj}(du; t),$$

where

$$\hat{A}_{hj}(u; t) = \hat{A}_{0hj}(u) \exp(\hat{\theta} t).$$

4.6 Simulation

We simulated $N = 300$ realizations of a three-state process with state 3 absorbing. Each path starts from state 1 or 2 equally likely and being observed until censoring time $C = 5$. The true transition rate functions are

$$\alpha_{12}(\tau;t) = 1.6\tau e^{-t/\xi_0}, \alpha_{13}(\tau;t) = 2\tau\left(1 - 0.8e^{-t/\xi_0}\right),$$
$$\alpha_{21}(\tau;t) = 1.2\tau e^{-t/\xi_0}, \alpha_{23}(\tau;t) = 2\tau\left(1 - 0.6e^{-t/\xi_0}\right).$$

Here the transition probabilities are given by $P_{12}(t) = 0.8e^{-t/\xi_0}$ and $P_{21}(t) = 0.6e^{-t/\xi_0}$. The conditional sojourn time distribution $F_{hj}(\tau;t)$ is Weibull with shape parameter 2 and scale parameter 1, for all $h \neq j$ and t.

We estimated the cumulative transition rates, semi-Markov Kernel, and transition rates. The semiparametric, piecewise constant, multiplicative nonparametric, and nonparametric estimation procedures were implemented and compared. Model (4.5.1) with $Z_{hj}(t) \equiv t$ was used in the semiparametric estimation. Four equally spaced partitions on the study time scale were used in the piecewise constant approach. The sample means of the estimates based on 100 replicates were calculated. We consider two different simulation settings:

Setting 4.1. $\xi_0 = \infty$, which leads to a homogeneous semi-Markov model.

Setting 4.2. $\xi_0 = 3$, which gives a nonhomogeneous semi-Markov model.

In the first setting, all the model assumptions required by the above four approaches are satisfied. Thus we can compare the efficiency of the different approaches. In the second setting, the semiparametric and the nonparametric multiplicative model assumptions are violated. Thus we can study robustness of the approaches against model misspecification. We used the Epanechnikov kernel function given in Table 4.1. The bandwidths w for the study time scale and b for the duration time scale were chosen as 0.4 in both settings.

The results of Setting 4.1 are shown in Figures 4.1 to 4.8. Figures 4.1 to 4.4 present the 3-dimensional surface plots of the sample means of the estimated transition rate functions in the two times. Figures 4.5 and 4.6 show the associated profiles of the sample means and standard deviations of the estimated transition rate functions. The profile functions of the estimated semi-Markov kernels are given in Figures 4.7 and 4.8. All the approaches gave approximately unbiased estimates. Approaches with more structured models, such as the semiparametric model, have smaller standard deviations and thus lead to more efficient inferences.

The results of Setting 4.2 are shown in Figures 4.9 to 4.16. In this setting, the estimation procedures based on the semiparametric model produced obviously biased estimates due to the model misspecification. The piecewise constant approach and the nonparametric estimation procedure performed well under the setting as expected. The estimation with the multiplicative nonparametric model, which is not correctly specified, showed certain robustness of the approach to the model misspecification.

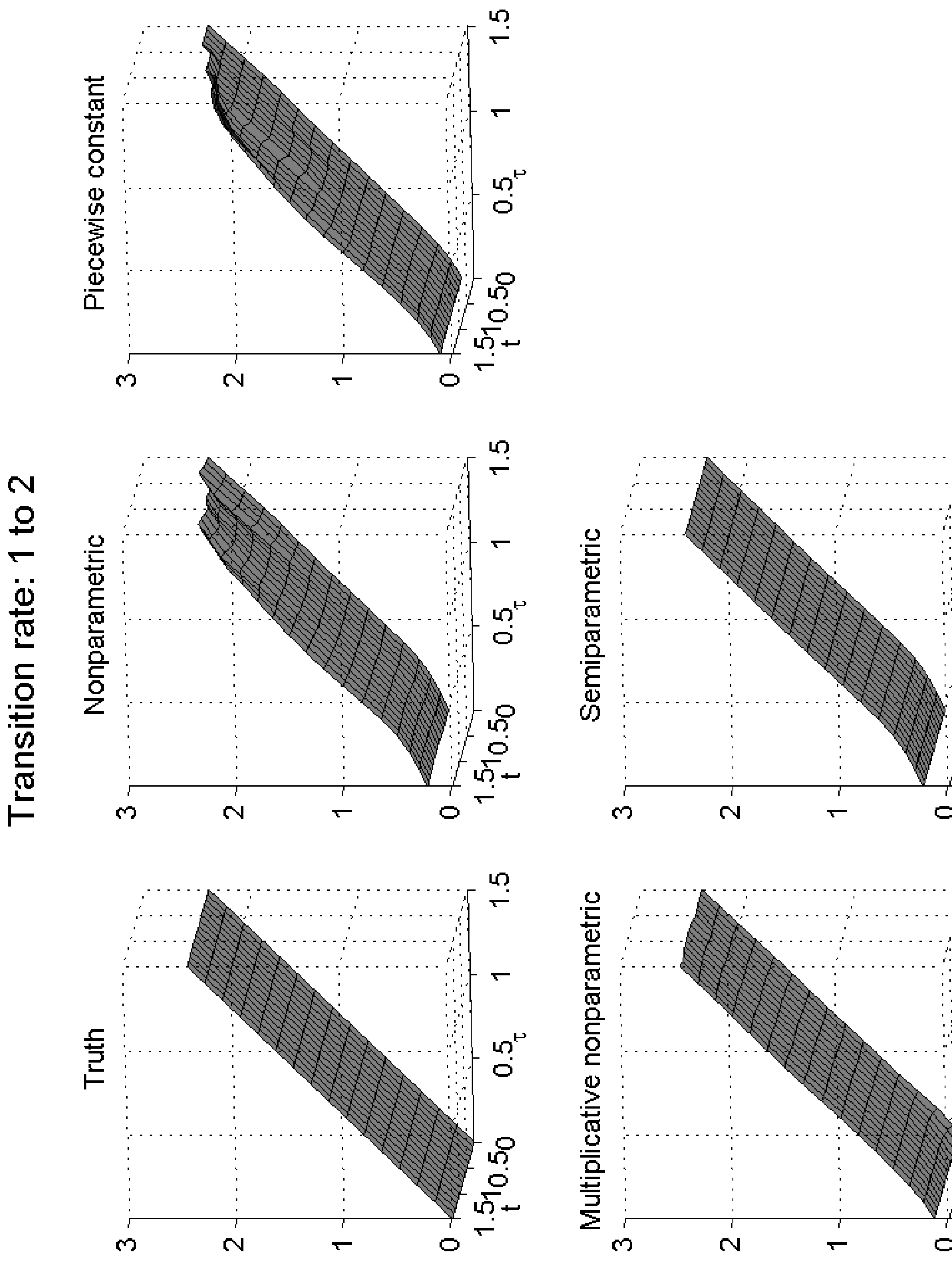

Figure 4.1: Truth and sample means of estimated transition rate functions from state 1 to state 2 in simulation setting 4.1

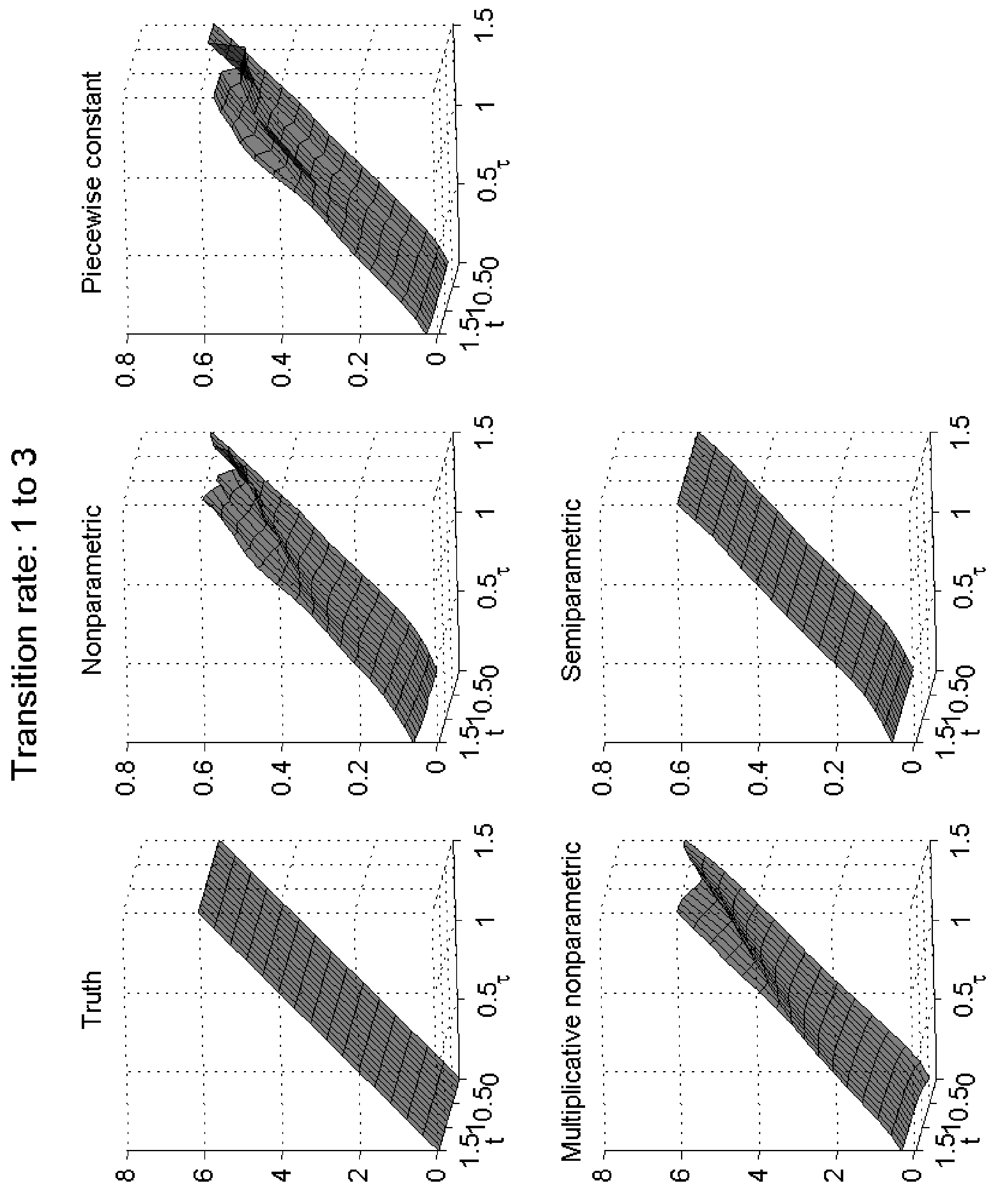

Figure 4.2: Truth and sample means of estimated transition rate functions from state 1 to state 3 in simulation setting 4.1

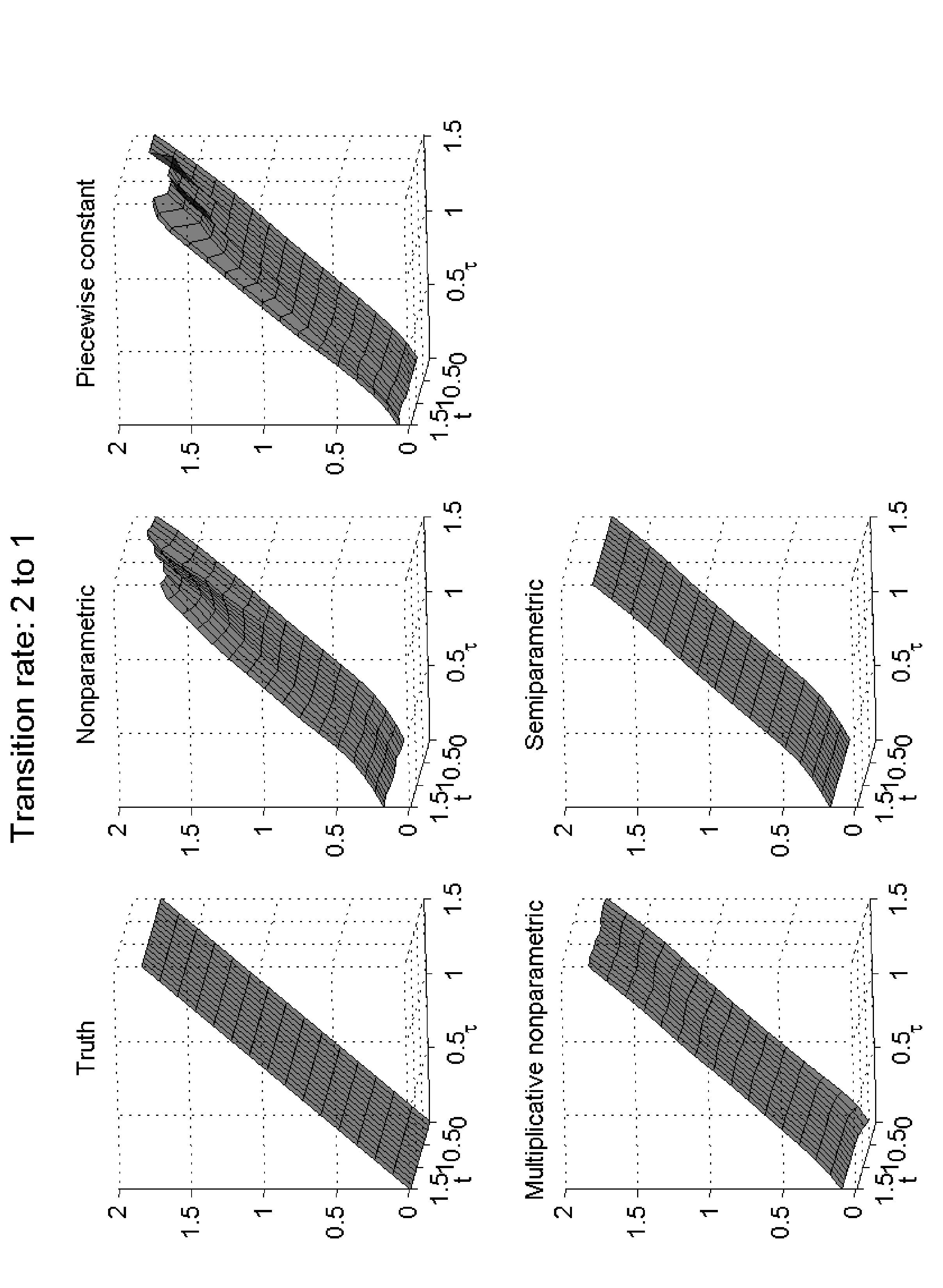

Figure 4.3: Truth and sample means of estimated transition rate functions from state 2 to state 1 in simulation setting 4.1

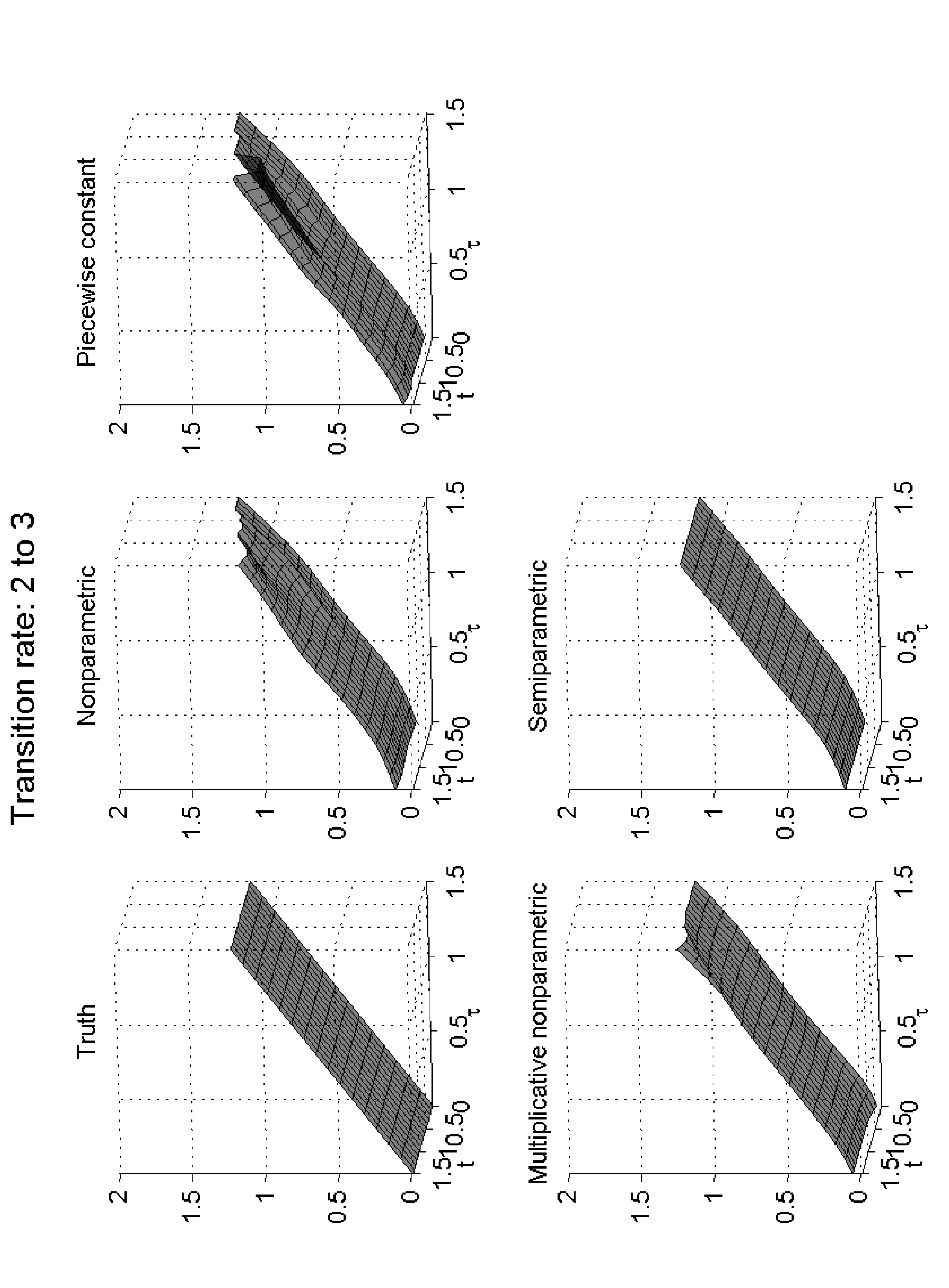

Figure 4.4: Truth and sample means of estimated transition rate functions from state 2 to state 3 in simulation setting 4.1

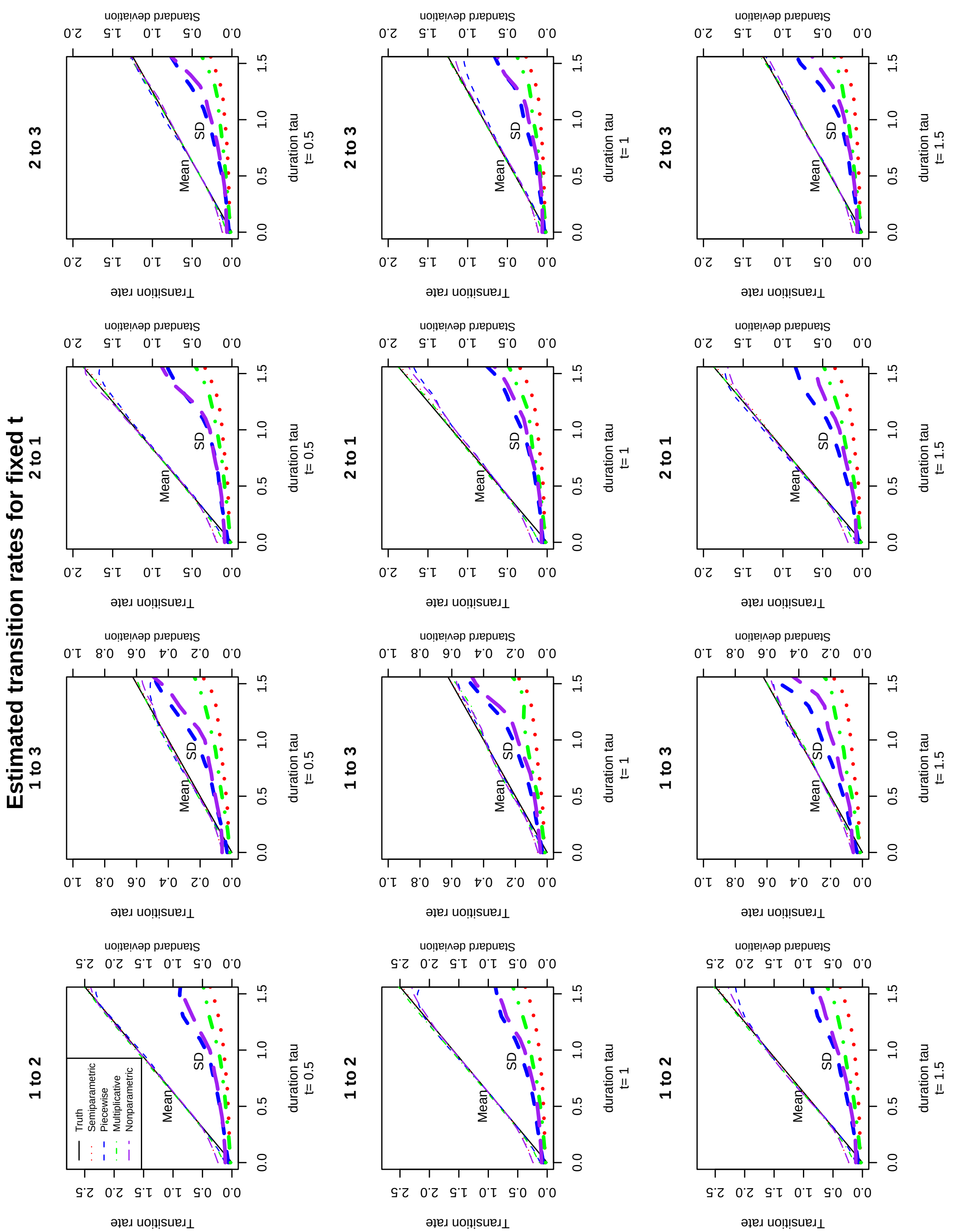

Figure 4.5: Truth and sample mean of estimated transition rate functions for fixed t in simulation setting 4.1 (Thinner lines: sample means; Thicker lines: sample standard deviations)

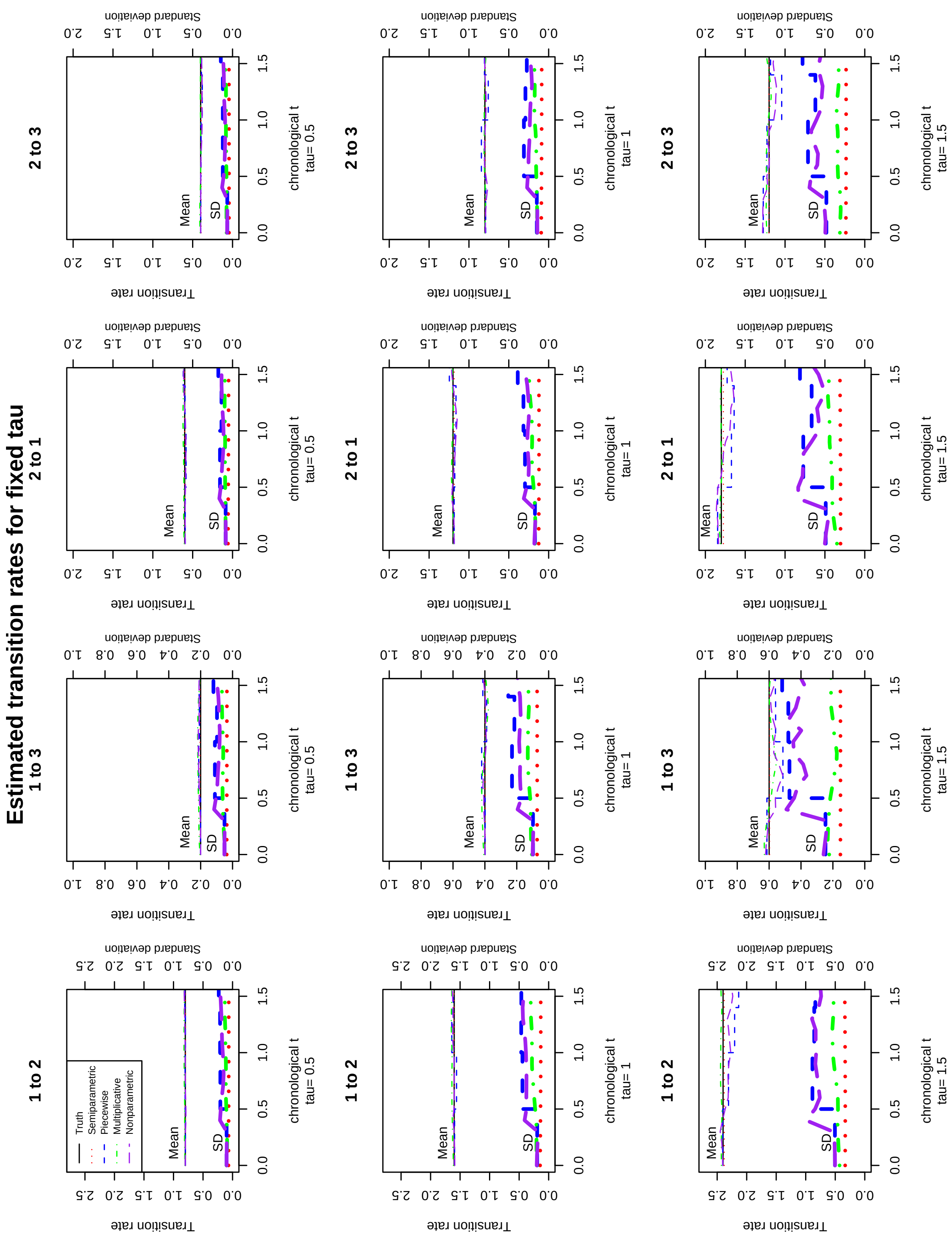

Figure 4.6: Truth and sample mean of estimated transition rate functions for fixed τ in simulation setting 4.1 (Thinner lines: sample means; Thicker lines: sample standard deviations)

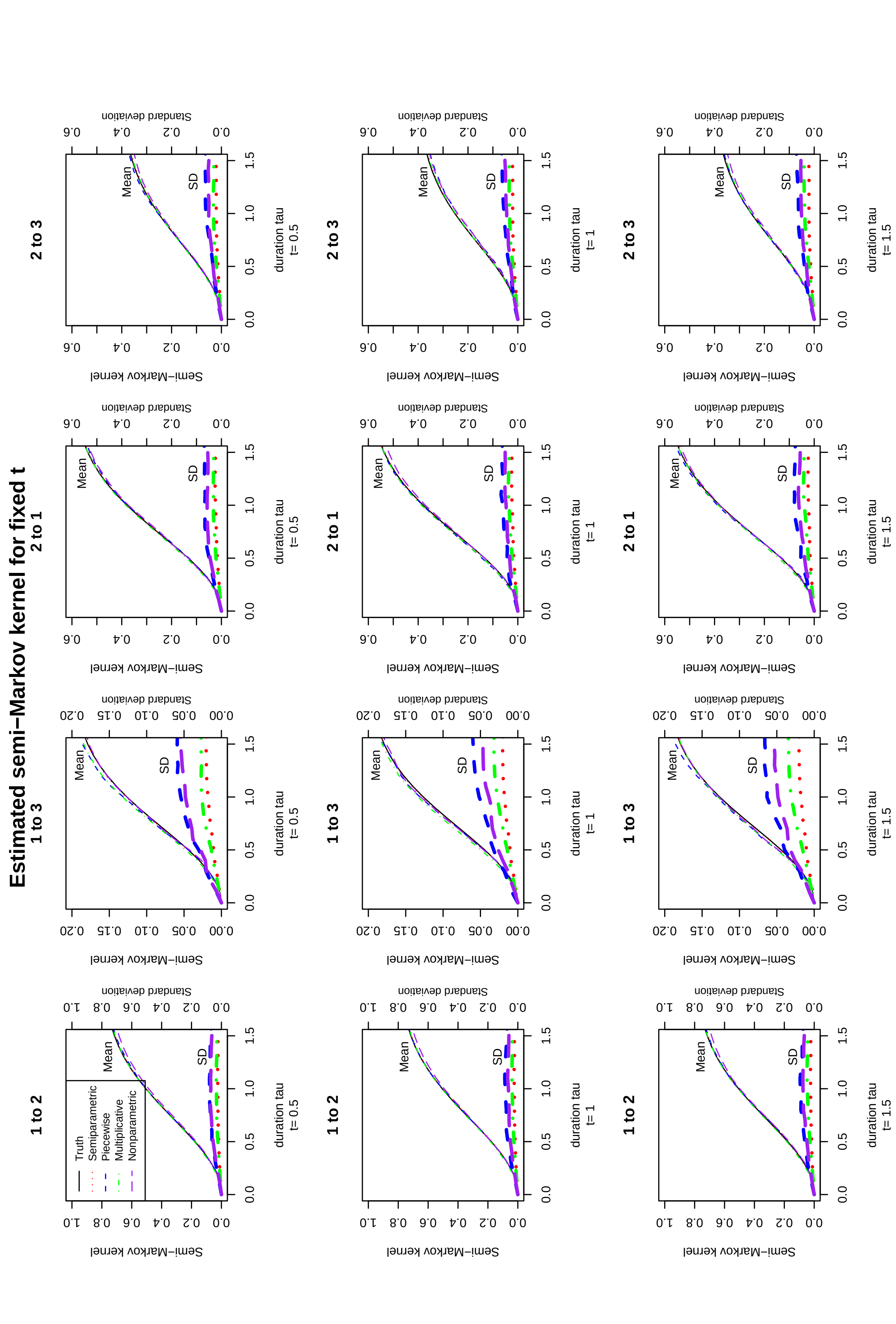

Figure 4.7: Truth and sample mean of estimated semi-Markov kernel for fixed t in simulation setting 4.1 (Thinner lines: sample means; Thicker lines: sample standard deviations)

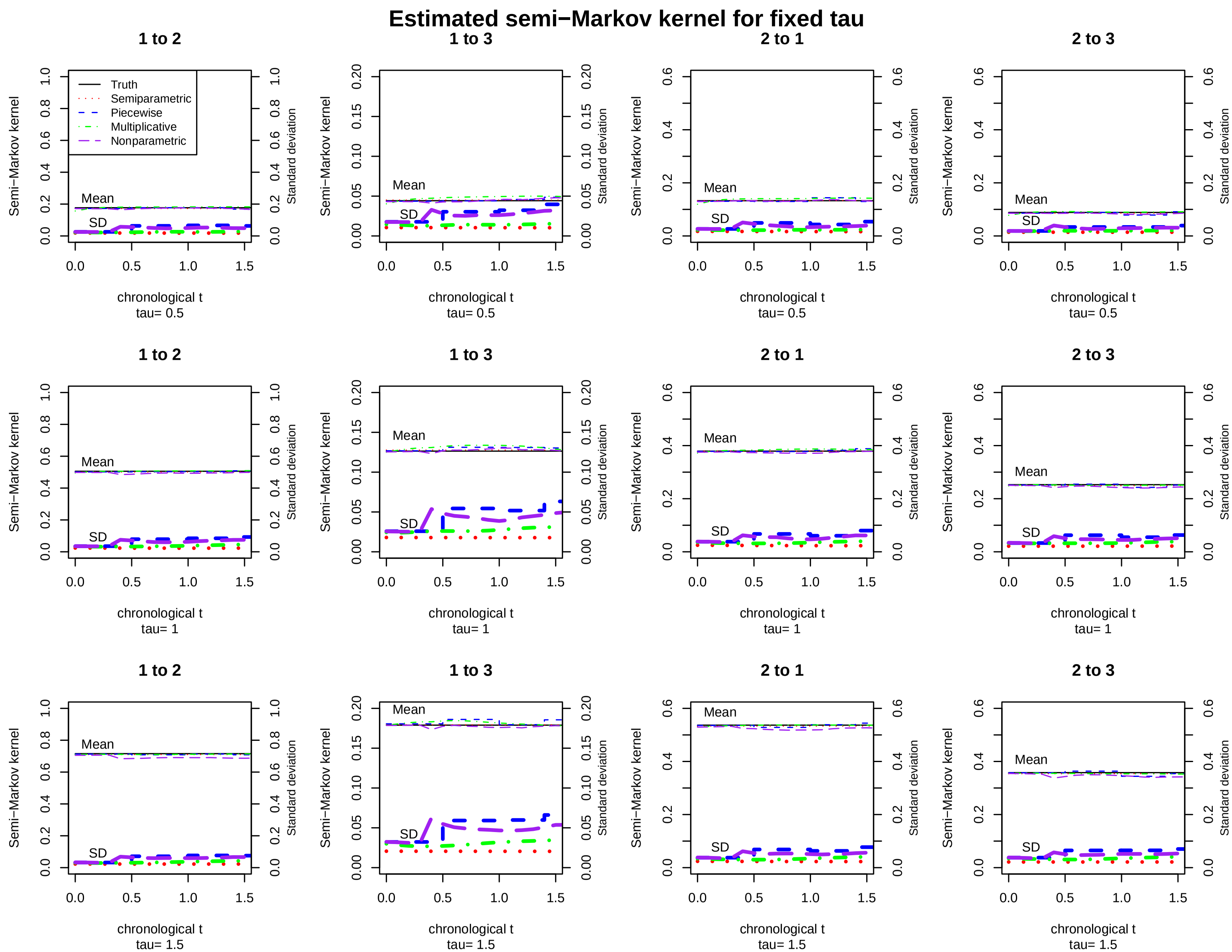

Figure 4.8: Truth and sample mean of estimated semi-Markov kernel for fixed τ in simulation setting 4.1 (Thinner lines: sample means; Thicker lines: sample standard deviations)

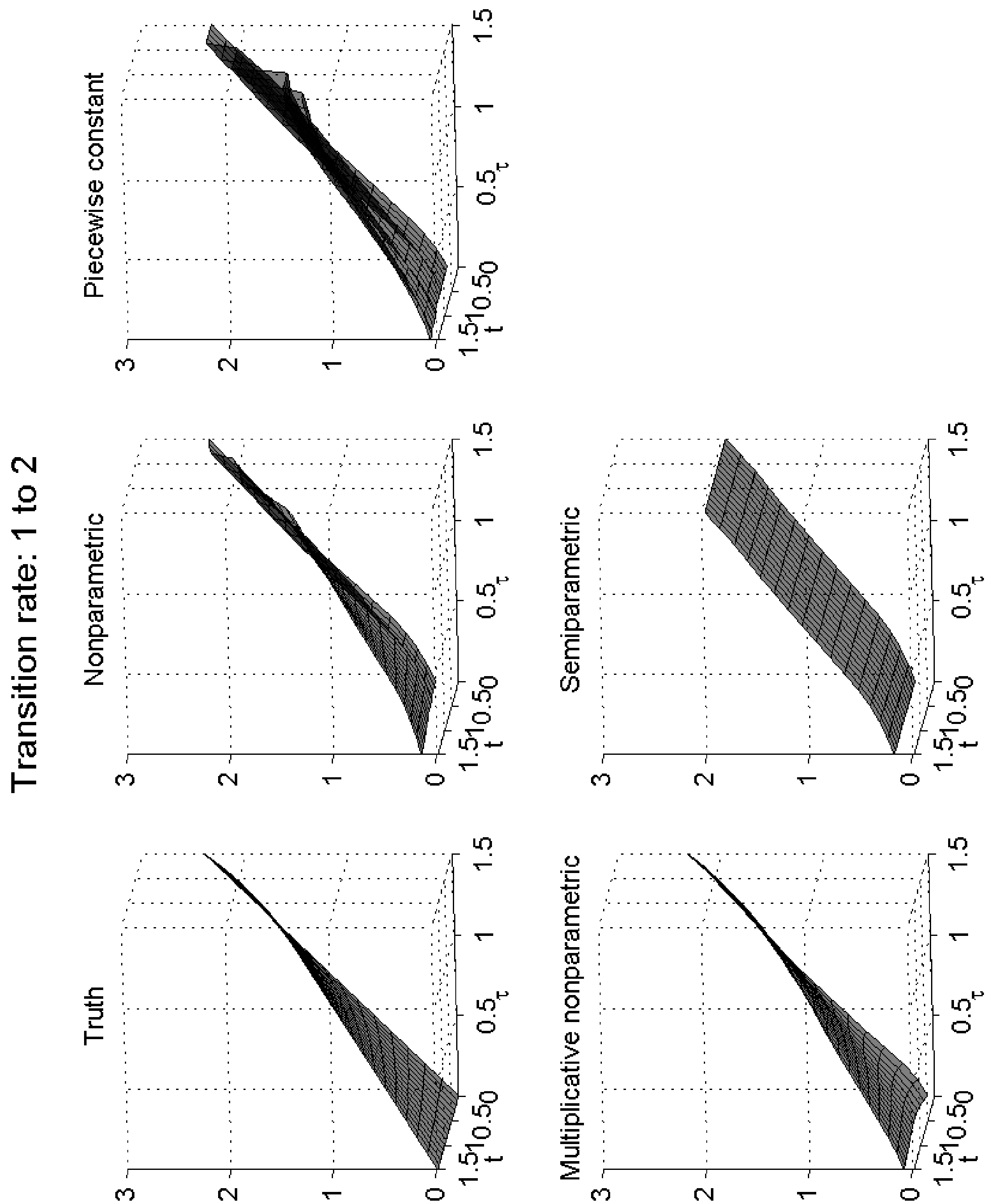

Figure 4.9: Truth and sample means of estimated transition rate functions from state 1 to state 2 in simulation setting 4.2

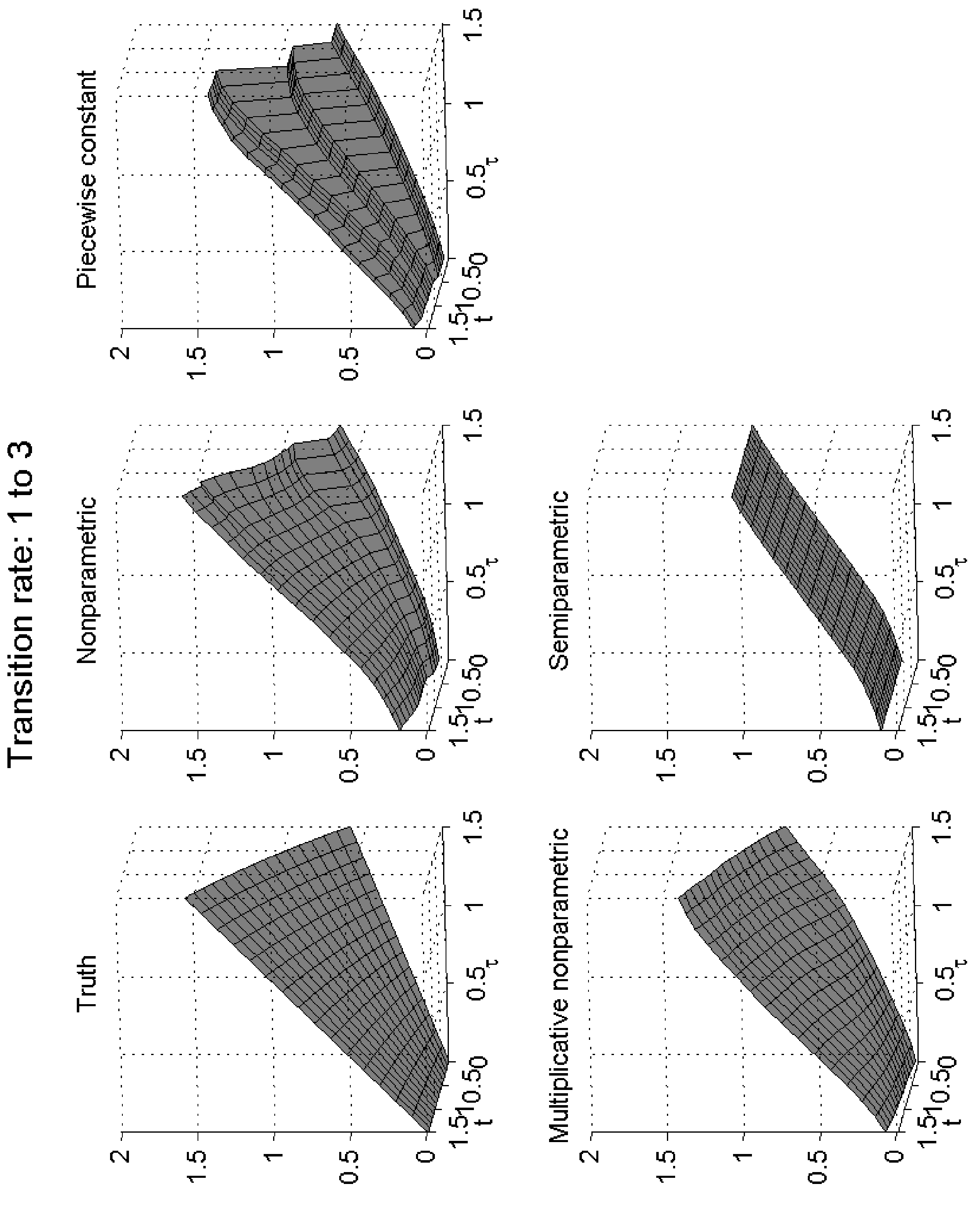

Figure 4.10: Truth and sample means of estimated transition rate functions from state 1 to state 3 in simulation setting 4.2

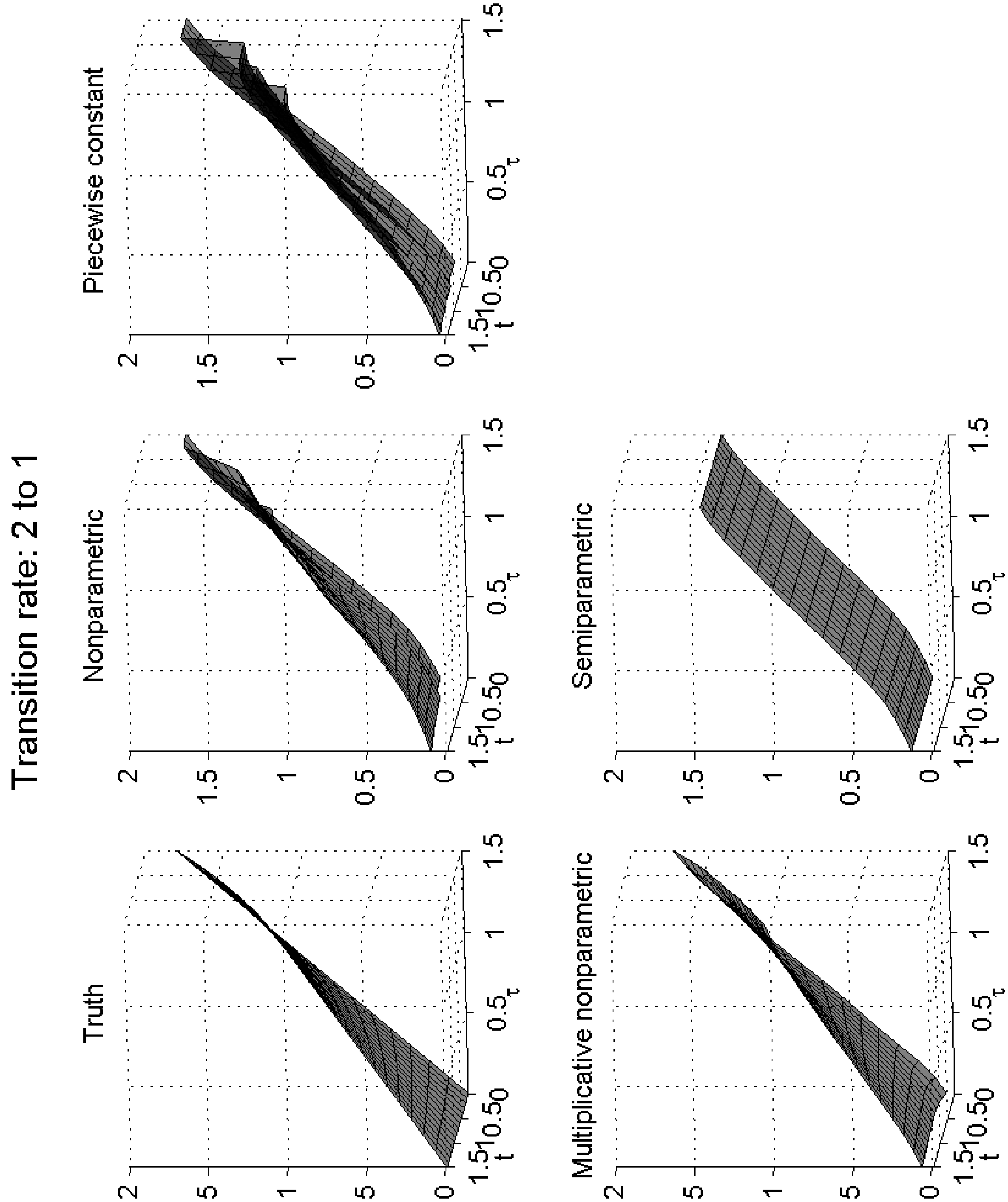

Figure 4.11: Truth and sample means of estimated transition rate functions from state 2 to state 1 in simulation setting 4.2

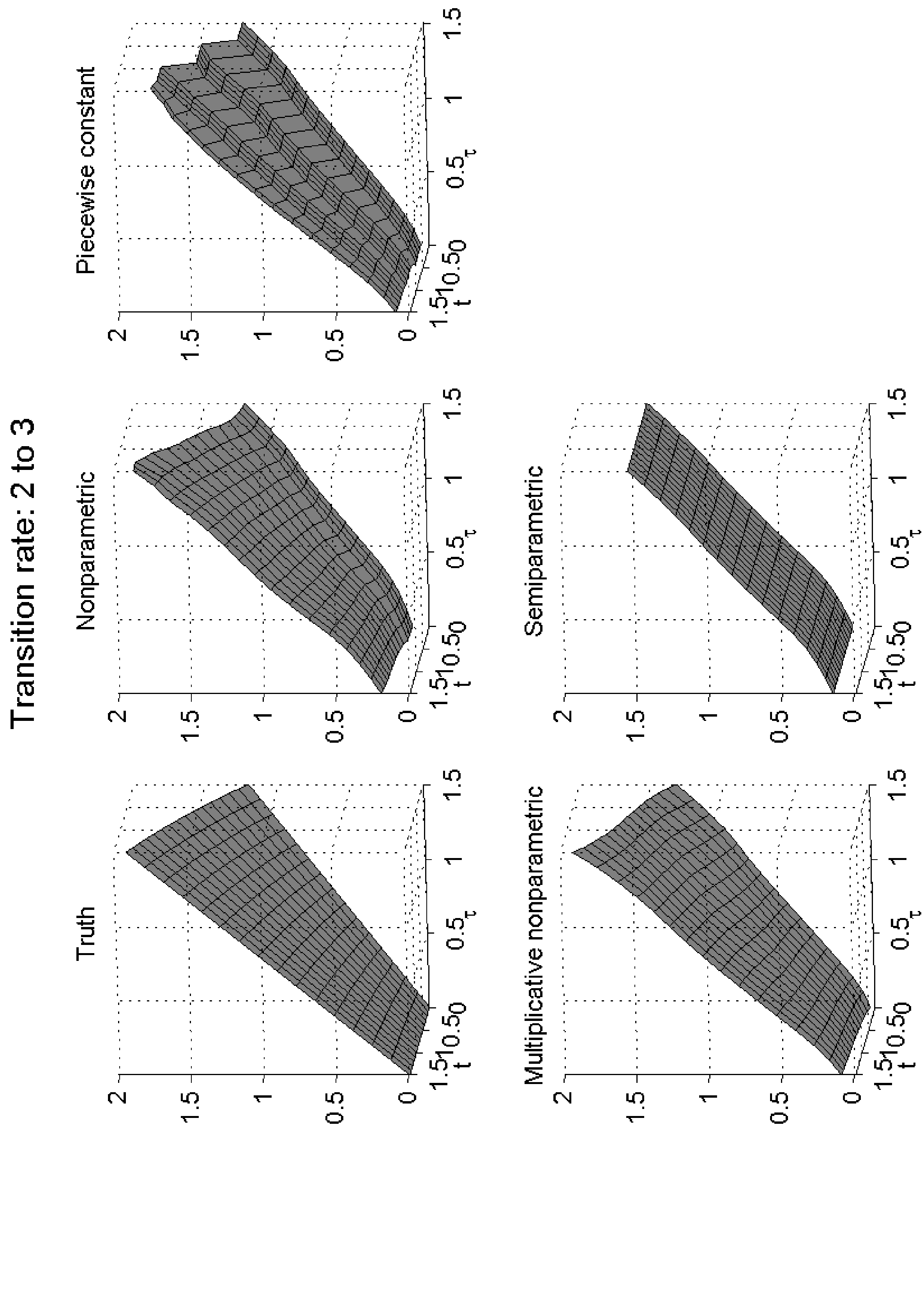

Figure 4.12: Truth and sample means of estimated transition rate functions from state 2 to state 3 in simulation setting 4.2

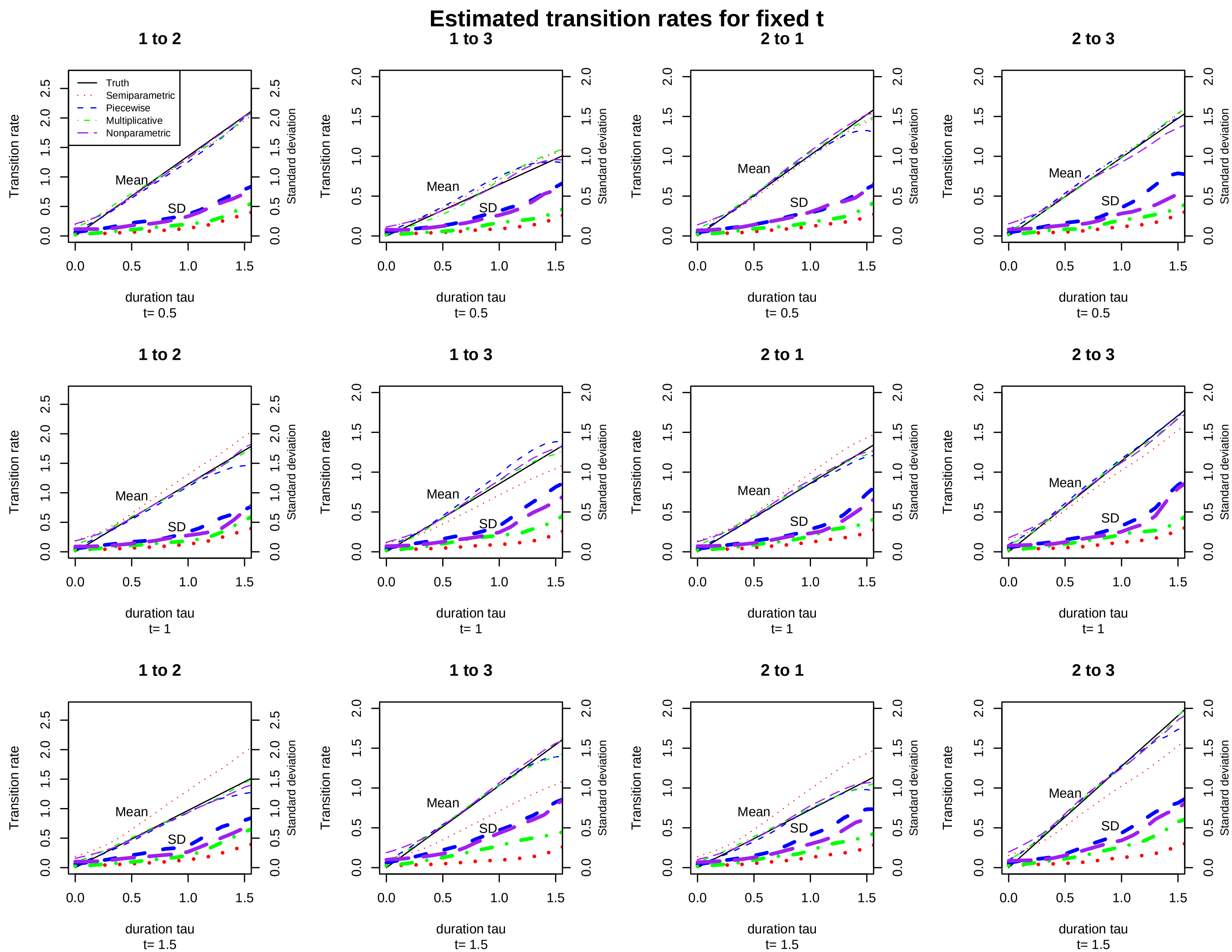

Figure 4.13: Truth and sample mean of estimated transition rate functions for fixed t in simulation setting 4.2 (Thinner lines: sample means; Thicker lines: sample standard deviations)

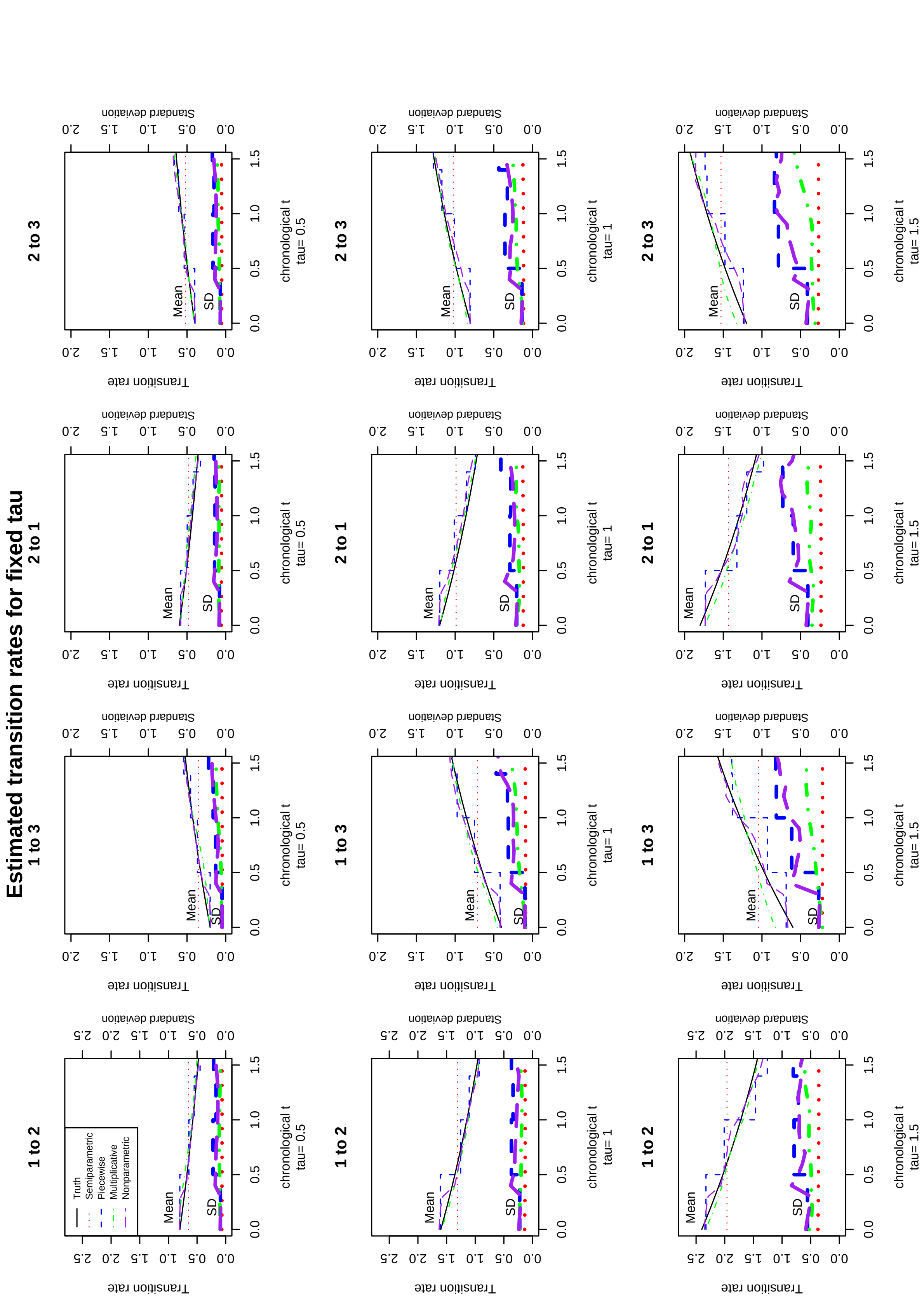

Figure 4.14: Truth and sample mean of estimated transition rate functions for fixed τ in simulation setting 4.2 (Thinner lines: sample means; Thicker lines: sample standard deviations)

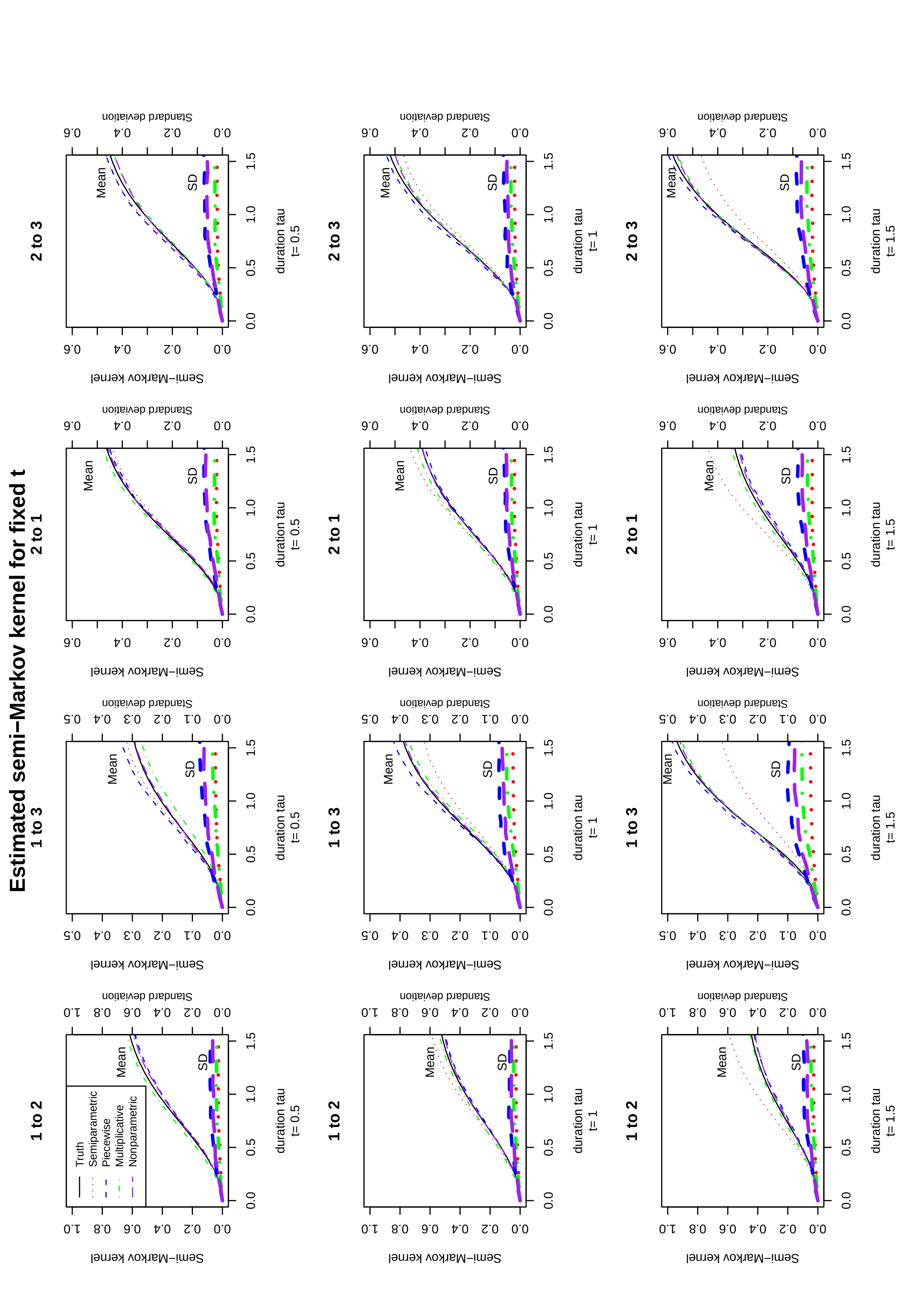

Figure 4.15: Truth and sample mean of estimated semi-Markov kernel for fixed t in simulation setting 4.2 (Thinner lines: sample means; Thicker lines: sample standard deviations)

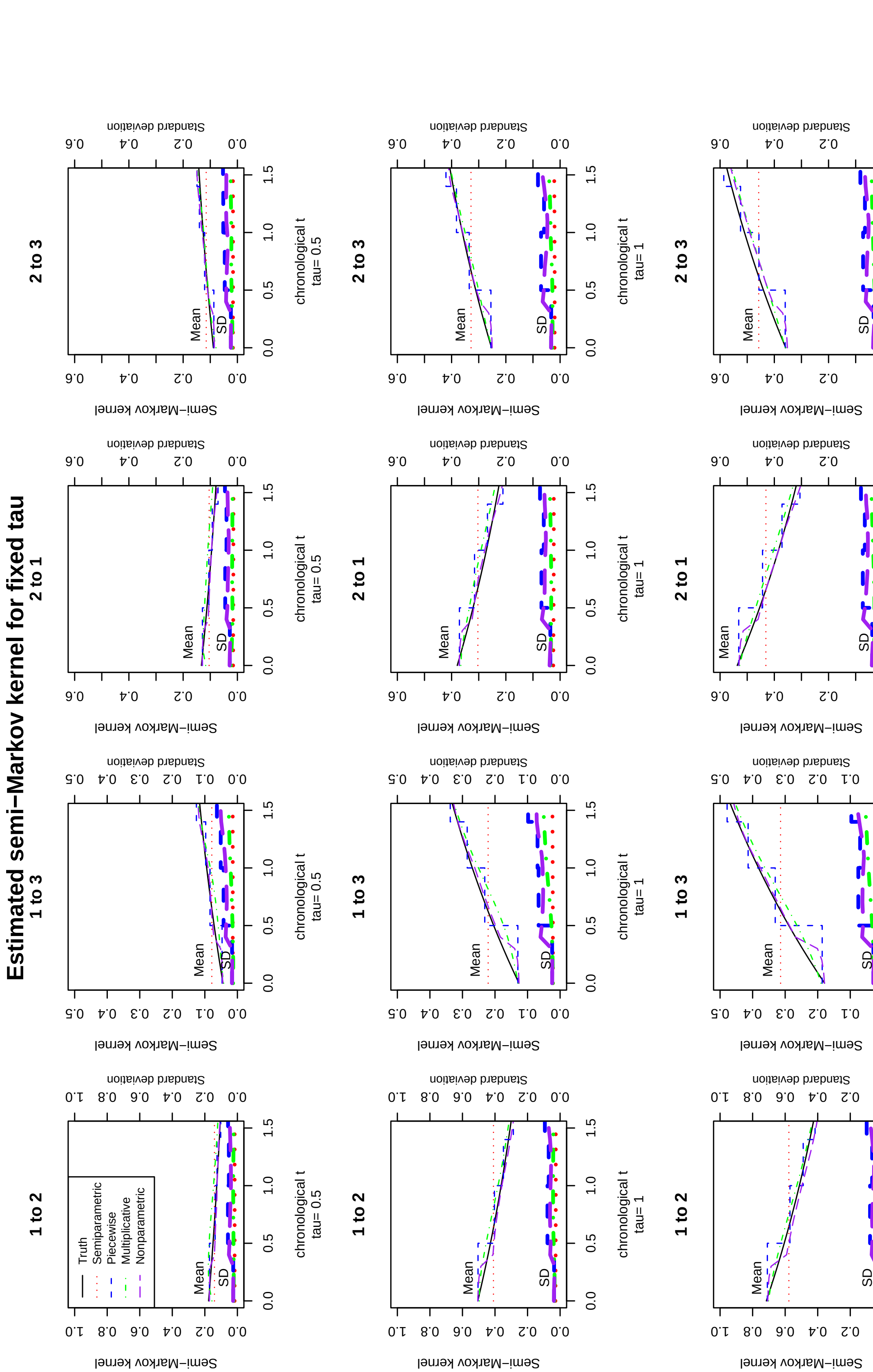

Figure 4.16: Truth and sample mean of estimated semi-Markov kernel for fixed τ in simulation setting 4.2 (Thinner lines: sample means; Thicker lines: sample standard deviations)

4.7 Extensions

In this chapter, we consider estimation procedures with NHSM processes under different model assumptions. The models are nested, which can be utilized to conduct model checking. For instance, if the multiplicative nonparametric model fit the data adequately, the estimated cumulative transition rate functions under the multiplicative nonparametric model should be close to the one based on the fully nonparametric approach. If the multiplicative nonparametric model is appropriate, the difference should be fluctuate around zero without any pattern.

In Section 4.4, we consider a particular nonparametric structured model: the nonparametric multiplicative model. Another nonparametric structured model for future research is the nonparametric additive model, in which the transition intensities depend on the study time and the duration time additively, and the functional forms are left unspecified. The iterative algorithm developed in Section 4.4 can be adapted. Theoretical justification for convergence of the algorithms is left for future research.

Kernel smoothing methods are used in both the nonparametric and structured nonparametric estimation procedures for the NHSM model. Thus bandwidth selection is an issue. The asymptotic distribution of the estimators may be too complicated to be used in selecting the bandwidth by the plug-in method. Bandwidth selection has been well studied in the context of hazard rate estimation with survival data. Patil (1993) proposes the least squares cross-validation method. González-Manteiga *et al.* (1996) introduce a bootstrap approach. These methods can be potentially adapted to the current setting.

Chapter 5

Semi-Markov Process with Informative Censoring

5.1 Introduction

Event history data are often incompletely observed. Up to the last chapter, we have worked on the observation scheme that the event process is continuously observed subject to right censoring at time C, which is assumed to be independent of the process itself. Although the independent censoring assumption may be plausible in some situations such as with administrative censoring, it is often questionable in many other situations where censoring is due to dropout or competing risks.

With survival data, the simplest structure of event history data, the well-known Kaplan-Meier (1958) estimator is inconsistent for the survival function when the censoring is dependent on the survival time. In fact, the survival function is not identifiable from the observable data (Tsiatis, 1975). Without additional assumptions about the dependence, it is only possible to determine the bounds for the survival function (Peterson, 1976). However, such bounds are often too wide to be of practical use. Two alternative approaches have been proposed in the literature, both of which require assumptions that can not be verified with the observed data. The first approach assumes that the survival time and the censoring time are independent, conditional on some available covariates (Robins, 1987; Robins and Rotnitzky, 1992; Robins, 1993; Satten et al., 2001), or some latent variables (Link, 1989; Oakes, 1989; Huang and Wolfe, 2002). The second approach imposes assumptions about the dependence between survival time and censoring time (Fisher and Kanareck, 1974;

Williams and Lagakos, 1977; Slud and Rubinstein, 1983; Klein and Moeschberger, 1988; Carrière, 1995; Zheng and Klein, 1995; Rivest and Wells, 2001; Braekers and Veraverbeke, 2005). Based on some some prior knowledge or subjective information about the range of possible strengths of the association between the survival time and the censoring time, one can then conduct sensitivity analysis and provide plausible bounds on the survival function (Zheng and Klein, 1995; Huang and Zhang, 2008).

In this chapter, we consider a particular type of informative right censoring involved in semi-Markov process observation. Motivated by the competing risks formulation of HSM processes (Lagakos *et al.*, 1978), we model the possible informative censoring mechanism as another competing risk. We assume that the resulting process becomes a new semi-Markov process if censoring is viewed as a new absorbing state in addition to the original process. Thus the large literature on dependent competing risks can be adapted in the setting. In particular, we adapt a copula-based approach proposed by Zheng and Klein (1995). The advantage of a copula approach is that the marginal distributions need not to be specified, and can be estimated nonparametrically.

The rest of this chapter is organized as follows. In Section 5.2 and Section 5.3, we present a model for informative censoring and the associated estimation procedure, respectively. We examine the approach with a simulation study in Section 5.4. Section 5.5 concludes this chapter with some remarks.

5.2 Modeling Informative Censoring

5.2.1 Assumptions

Consider a multi-state process, formulated by the two-dimensional process $(\mathbf{J}, \mathbf{T}) = \{(J_m, T_m) : m = 0, 1, \ldots\}$, where the process $\{J_m : m = 0, 1, 2, \ldots\}$, taking values in the set $\mathcal{E} = \{1, 2, \ldots, r\}$, gives the sequence of states visited by the system, and the sequence $\{T_m : m = 0, 1, \ldots\}$ is the set of the corresponding transition times.

Suppose the original event process is observed up to a censoring time D. We view the censoring as a new absorbing state, denoted by $r + 1$. Then we end up with a new multi-state process which can be presented by a new two-dimensional process $(\mathbf{J}^*, \mathbf{T}^*) = \{(J^*_m, T^*_m) : m = 0, 1, \ldots\}$, where the process $\{J^*_m, m = 0, 1, 2, \ldots\}$, taking values in the set $\mathcal{E}^* = \{1, 2, \ldots, r, r + 1\}$, gives the sequence of states visited by the new multi-state process, and $\{T^*_m : m = 0, 1, \ldots\}$ are the corresponding transition times.

To see the relationship between $(\mathbf{J}^*, \mathbf{T}^*)$ and $(\mathbf{J}, \mathbf{T})$, let M be the random variable such that $T_M < D \leq T_{M+1}$. In the case that the original multi-state process has absorbing states and the original process enters an absorbing state before censored by D, such M does not exist, and we define $M = \infty$. Otherwise, we can easily see that $(J_m^*, T_m^*) = (J_m, T_m)$ for $m = 0, \ldots, M$, $J_{M+1}^* = r + 1$, and $T_{M+1}^* = D$.

We first make the following assumptions about the informative censoring mechanism:

Assumption 5.A1. Given $D \geq T_m$, the joint distribution of $(D - T_m, J_{m+1}, X_{m+1})$ depends on the past history $\{J_0, T_0, \ldots, J_m, T_m\}$ only through (J_m, T_m), i.e.,

$$\begin{aligned}&\left[D - T_m, J_{m+1}, X_{m+1} \middle| D \geq T_m, J_m, T_m, \ldots, J_0, T_0\right] \\ &\sim \left[D - T_m, J_{m+1}, X_{m+1} \middle| D \geq T_m, J_m, T_m\right],\end{aligned} \tag{5.2.1}$$

which is free of m.

Assumption 5.A2. In addition to Assumption 5.A1, the conditional distribution in (5.2.1) is also free of T_m.

Proposition 5.2.1. *Under Assumption 5.A1 (5.A2), the resulting process* $(\boldsymbol{J}^*, \boldsymbol{T}^*)$ *forms a NHMR (HMR) process.*

Proof. For any $h \leq r$, we have

$$\begin{aligned}&P\left\{J_{m+1}^* = r+1, X_{m+1}^* \leq \tau \middle| J_m^* = h, T_m^* = t, J_{m-1}^*, T_{m-1}^*, \cdots\right\} \\ =&P\left\{\begin{array}{l} D - T_m \leq X_{m+1}, D - T_m \leq \tau \big| \\ \qquad\qquad D \geq T_m, J_m = h, T_m = t, J_{m-1}, T_{m-1}, \cdots \end{array}\right\} \\ =&P\left\{D - T_m \leq X_{m+1}, D - T_m \leq \tau \middle| D \geq T_m, J_m = h, T_m = t\right\} \\ =&P\left\{J_{m+1}^* = r+1, X_{m+1}^* \leq \tau \middle| J_m^* = h, T_m^* = t\right\},\end{aligned}$$

where the first and the third equalities follow from the relationship between $(\mathbf{J}^*, \mathbf{T}^*)$ and $(\mathbf{J}, \mathbf{T})$, and the second equality holds by Assumption 5.A1.

Similarly, for any $h, j \in \mathcal{E}$,

$$\begin{aligned}&P\{J_{m+1}^* = j, X_{m+1}^* \leq \tau \big| J_m^* = h, T_m^* = t, J_{m-1}^*, T_{m-1}^*, \cdots\} \\ =&P\left\{\begin{array}{l} D - T_m > X_{m+1}, J_{m+1} = j, X_{m+1} \leq \tau \big| \\ \qquad\qquad D \geq T_m, J_m = h, T_m = t, J_{m-1}, T_{m-1}, \cdots\} \end{array}\right\} \\ =&P\{D - T_m > X_{m+1}, J_{m+1} = j, X_{m+1} \leq \tau \big| D \geq T_m, J_m = h, T_m = t\} \\ =&P\{J_{m+1}^* = j, X_{m+1}^* \leq \tau \big| J_m^* = h, T_m^* = t\}.\end{aligned}$$

Thus the resulting process $(\mathbf{J}^*, \mathbf{T}^*)$ is a NHMR process under Assumption 5.A1. Similarly $(\mathbf{J}^*, \mathbf{T}^*)$ is a HMR process under Assumption 5.A2.

□

Remark 5.2.1. The simple survival process can be represented by $\{J_0 = 0, T_0 = 0, J_1 = 1, T_1 = T\}$, where T is the survival time, 0 represents the state 'alive', and 1 for 'death'. If we view the informative censoring as a new absorbing state 2, then Assumptions 5.A1 and 5.A2 are both trivially satisfied. The resulting process is a dependent competing risks process (and a semi-Markov process, of course).

Remark 5.2.2. Assumption 5.A1 implies that the censoring mechanism has a certain renewal property with the original semi-Markov process: given $D \geq T_m$, the remaining time to censoring $D - T_m$ depends on the past history $(J_0, T_0), \ldots, (J_m, T_m)$ only through (J_m, T_m), the state and time of the mth transition of the original multi-state process. Furthermore, Assumption 5.A2 implies that given $D \geq T_m$, the remaining time to censoring $D - T_m$ depends on the past history $(J_0, T_0), \ldots, (J_m, T_m)$ only through J_m.

Remark 5.2.3. Assumption 5.A1 also implies that

$$\left[J_{m+1}, X_{m+1} \middle| D \geq T_m, J_m, T_m, \ldots, J_0, T_0\right] \sim \left[J_{m+1}, X_{m+1} \middle| D \geq T_m, J_m, T_m\right],$$

free of m. It reduces to the semi-Markov kernel of the original semi-Markov process if "$D \geq T_m$" is removed. We thus make the assumption as follows.

Assumption 5.B. Assume that for all m, (J_{m+1}, X_{m+1}) and $I\{D \geq T_m\}$ are independent conditional on $\{J_m, T_m, \ldots, J_0, T_0\}$.

Remark 5.2.4. Assumption 5.B is weaker than the assumption that (J_{m+1}, X_{m+1}) and D are independent conditional on $\{J_m, T_m, \ldots, J_0, T_0\}$. For instance, Assumption 5.B is automatically satisfied for the survival process subject to informative censoring at time D. But (J_{m+1}, X_{m+1}) and D are independent conditional on $\{J_m, T_m, \ldots, J_0, T_0\}$ implies that T and D are independent.

For the survival process subject to informative censoring at time D, the intensities of the resulting 3-state process are simply the cause-specific hazards of the competing risks:

$$h_{01}(t) = \lim_{\triangle t \downarrow 0} \frac{P\{T \in [t, t + \triangle t) | T \geq t, D \geq t\}}{\triangle t},$$

$$h_{02}(t) = \lim_{\triangle t \downarrow 0} \frac{P\{D \in [t, t + \triangle t) | T \geq t, D \geq t\}}{\triangle t}.$$

The cause-specific hazards are estimable. However, if we are interested in the original survival process, i.e., the survival distribution of T, we need to remove the 'censoring cause' from the process to estimate

$$h_T(t) = \lim_{\triangle t \downarrow 0} \frac{P\{T \in [t, t + \triangle t) | T \geq t\}}{\triangle t}.$$

The nonidentifiability issue then arises as discussed in Section 5.1. An assumption concerning the dependence between the survival time T and the censoring time D is needed.

Thus to make inference about the original semi-Markov process, which is more general than the survival process, we need to further model the relationship between (J_{m+1}, X_{m+1}) and the remaining time to censoring $D - T_m$, given $D \geq T_m$ and (J_m, T_m), the state and time of the mth transition of the original multi-state process. If there are only two states in the original multi-state process, i.e., $\mathcal{E} = \{1, 2\}$, then J_{m+1} is known given J_m. In this case, we only need to specify the dependence of X_{m+1} and $D - T_m$ given (J_m, T_m). In what follows, we focus on this simple case and model the dependence by copulas. Extensions to deal with the case that $\mathcal{E}$ has more than two states is discussed in Section 5.5.

5.2.2 Copula Models

The copula models provide a flexible way to specify the dependence structure between two random variables. A good introduction to copula functions is given by Nelsen (1999). A two-dimensional copula is a bivariate distribution function $H(y_1, y_2)$ defined on the square $[0, 1] \times [0, 1]$ with uniform one-dimensional marginal distributions. The simplest copula function is the independence copula $H(y_1, y_2) = y_1 y_2$. In this chapter, we focus on the class of the Archimedean copulas, which have the form

$$H(y_1, y_2) = \phi^{-1} \left[\phi(S_1(y_1)) + \phi(S_2(y_2))\right] \tag{5.2.2}$$

where $S_1(\cdot)$ and $S_2(\cdot)$ are two survival functions, and $\phi(\cdot)$ is a twice differentiable, decreasing convex function defined on (0,1] satisfying $\phi(0) = \infty$ and $\phi(1) = 0$. As a measure of association, the Kendall's τ for the Archimedean copulas can be conveniently computed by

$$\tau = 1 + 4 \int_0^1 \frac{\phi(u)}{\phi'(u)} du.$$

As an example, the independence copula is a member of the Archimedean copulas. It has $\phi(u) = \log(1/u)$ and Kendall's $\tau = 0$. More Archimedean copulas are described in Section 5.4.

As indicated in the previous section, we focus on the case that $\mathcal{E} = \{1, 2\}$. The model we consider is

$$\begin{aligned} &P\{X_{m+1} > u, D - T_m > v, J_{m+1} = 3 - h | D \geq T_m, J_m = h, T_m = t\} \\ &= \phi_{h,t}^{-1}[\phi_{h,t}(S_{h,3-h}(u; t)) + \phi_{h,t}(S_{h3}(v; t))], \end{aligned} \tag{5.2.3}$$

where $\phi_{h,t}(\cdot)$ is a known Archimedean copula function for $h = 1, 2$ and for each t,

$$S_{h,3-h}(u; t) = P\{X_{m+1} > u, J_{m+1} = 3 - h | D \geq T_m, J_m = h, T_m = t\},$$

and

$$S_{h3}(v; t) = P\{D - T_m > v | D \geq T_m, J_m = h, T_m = t\}.$$

According to Assumption 5.B,

$$S_{h,3-h}(u; t) = P\{X_{m+1} > u, J_{m+1} = 3 - h | J_m = h, T_m = t\},$$

which equals $1 - Q_{h,3-h}(u; t)$, where $Q_{h,3-h}(u; t)$ is the semi-Markov kernel of the original semi-Markov process. If the corresponding processes are HSM processes, we can simply drop 't' in Model (5.2.3).

5.3 Estimation Procedure

We consider estimation for the semi-Markov kernel of the original two-state semi-Markov process based on observations subject to the informative right censoring under Assumptions 5.A1 (5.A2) and 5.B, and Model (5.2.3). In addition to the informative censoring time D, we also allow the observation subject to another censoring time C, which is assumed independent of the original multi-state process and D. We consider estimation procedures based on n i.i.d. resulting processes subject to the two types of censoring.

We first consider estimation procedure for the new NHSM process. Viewing censoring due to D as a new absorbing state 3, the resulting three-state process is a NHSM process, which is observed subject to noninformative right censoring time C. The semi-Markov kernel of the new NHSM process is given by

$$Q_{12}^*(\tau; t) = P\{X_{m+1} \leq \tau, D - T_m \geq X_{m+1} | D \geq T_m, J_m = 1, T_m = t\},$$

$$Q_{13}^*(\tau;t) = P\{D - T_m \le \tau, D - T_m < X_{m+1} | D \ge T_m, J_m = 1, T_m = t\},$$

$$Q_{21}^*(\tau;t) = P\{X_{m+1} \le \tau, D - T_m \ge X_{m+1} | D \ge T_m, J_m = 2, T_m = t\},$$

and

$$Q_{23}^*(\tau;t) = P\{D - T_m \le \tau, D - T_m < X_{m+1} | D \ge T_m, J_m = 2, T_m = t\}.$$

These kernel functions can be estimated by the methods developed in Chapter 4. Denote the estimators by $\hat{Q}_{12}^*(\tau;t)$, $\hat{Q}_{13}^*(\tau;t)$, $\hat{Q}_{21}^*(\tau;t)$, and $\hat{Q}_{23}^*(\tau;t)$, respectively.

We are interested in the semi-Markov kernel of the original NHSM process, $Q_{12}(\cdot;t)$ and $Q_{21}(\cdot;t)$. For a fixed $t = t_0$, let $u = v$ in Model (5.2.3),

$$\phi_{1,t_0}(S_1(u;t_0)) = \phi_{1,t_0}(S_{12}(u;t_0)) + \phi_{1,t_0}(S_{13}(u;t_0)), \tag{5.3.1}$$

where $S_1(u;t_0) = 1 - Q_{12}^*(u;t_0) - Q_{13}^*(u;t_0)$, which can be estimated by $\hat{S}_1(u;t_0) = 1 - \hat{Q}_{12}^*(u;t_0) - \hat{Q}_{13}^*(u;t_0)$. We assume that the semi-Markov kernel and the censoring time distribution are continuous so that no tied observations occur. In this case, $\hat{Q}_{12}^*(\cdot;t_0)$ and $\hat{Q}_{13}^*(\cdot;t_0)$ do not jump together.

Our estimator for the survival function $S_{12}(\cdot;t_0)$ is a right continuous decreasing step function $\hat{S}_{12}(\cdot;t_0)$ with $\hat{S}_{12}(0;t_0) = 1$ and only changes when $\hat{Q}_{12}^*(\cdot;t_0)$ does. Define

$$\hat{Q}_{12}^*(\Delta u;t_0) = \hat{Q}_{12}^*(u-;t_0) - \hat{Q}_{12}^*(u;t_0)$$

as the jump size of $\hat{Q}_{12}^*(\cdot;t_0)$ at time u. Since $\hat{Q}_{12}^*(\cdot;t_0)$ and $\hat{Q}_{13}^*(\cdot;t_0)$ do not jump together, according to (5.3.1), for each u such that $\hat{Q}_{12}^*(\Delta u;t_0) > 0$,

$$\phi_{1,t_0}(\hat{S}_{12}(u-;t_0)) - \phi_{1,t_0}(\hat{S}_{12}(u;t_0)) = \phi_{1,t_0}(\hat{S}_1(u-;t_0)) - \phi_{1,t_0}(\hat{S}_1(u;t_0)). \tag{5.3.2}$$

Summing (5.3.2) over all u's less equal than τ for which $\hat{Q}_{12}^*(\Delta u;t_0) > 0$, we obtain a closed form expression

$$\hat{S}_{12}(\tau;t_0) = \phi_{1,t_0}^{-1}\left[-\sum_{u\le\tau,\hat{Q}_{12}^*(\Delta u;t_0)>0}\Big(\phi_{1,t_0}(\hat{S}_1(u-;t_0)) - \phi_{1,t_0}(\hat{S}_1(u;t_0))\Big)\right].$$

The semi-Markov kernel $Q_{12}(\tau;t_0)$ can then be estimated by $\hat{Q}_{12}(\tau;t_0) = 1 - \hat{S}_{12}(\tau;t_0)$. Similarly we can get a closed form estimator for the semi-Markov kernel $Q_{21}(\tau;t_0)$.

Estimation procedure with a HSM process is a special case, in which we drop t and t_0 everywhere in the above procedure.

5.4 Simulation

We simulated 300 subjects entering the study with staggered entry times generated from $unif(0,3)$. The administrative end of study censoring time was a constant $C = 4$. The exponential distribution with mean $1.5\exp(-t/\xi_0)$ are used as the distributions $1-S_{12}(\cdot;t)$ and $1-S_{23}(\cdot;t)$, and the exponential distribution with mean $\exp(-t/\xi_0)$ as the distributions $1-S_{13}(\cdot;t)$ and $1-S_{21}(\cdot;t)$. We considered two different simulation settings:

Setting 5.1. $\xi_0 = \infty$, the original two-state process is a HSM process,

Setting 5.2. $\xi_0 = 5$, the original two-state process is a NHSM process.

The true copula used in the simulation for Model (5.2.3) belongs to Clayton's family (Clayton, 1978):

$$H(y_1, y_2; \alpha) = [y_1^{-\alpha} + y_2^{-\alpha} - 1]^{-1/\alpha}, \quad \alpha > 0.$$

It is a class of the Archimedean copulas with $\phi(x) = (x^{-\alpha} - 1)/\alpha$ and $\phi^{-1}(x) = (1 + \alpha x)^{-1/\alpha}$. It has Kendall's $\tau = \alpha/(\alpha + 2)$. It reduces to the independence copula when $\alpha \to 0$. In this simulation, we took $\alpha = 1$ so that Kendall's $\tau = 1/3$ for all t. Note that this copula specification also results from a proportional frailty model, where X_{m+1} and $D - T_m$, provided $D > T_m$ and $(J_m, T_m) = (1, t)$, are assumed to be independent conditional on a gamma distributed latent variable with mean 1 and variance $1/\alpha$.

To examine the performance of the estimation procedures, we estimated the sojourn time distributions of the original two-state process based on the true copula function. We also evaluated the estimators based on the correct copula function but with wrongly specified association parameter $\alpha = 3$ which gives Kendall's $\tau = 0.6$. In addition, as a comparison, we estimated the sojourn time distribution by ignoring the dependent censoring.

To study the robustness of our estimator, we evaluated the estimators based on the misspecified copula functions with Kendall's τ's equal to the true value, i.e., 1/3. We used two copula families. The first family is the Gumbel-Hougaard copulas (Gumbel, 1961; Hougaard, 1986):

$$H(y_1, y_2; \alpha) = \exp\left[-\left\{(-\log y_1)^{1/\alpha} + (-\log y_2)^{1/\alpha}\right\}^{\alpha}\right], \quad \alpha \in (0,1),$$

which is a class of the Archimedean copulas with $\phi(x) = (-\log x)^{1/\alpha}$ and $\phi^{-1}(x) = e^{-x^{\alpha}}$. It has Kendall's $\tau = 1-\alpha$, and the independence copula corresponds to $\alpha \to 1$. The second family is Frank's copulas (Frank, 1979):

$$H(y_1, y_2; \alpha) = \log_{\alpha}\left\{1 + \frac{(\alpha^{y_1}-1)(\alpha^{y_2}-1)}{\alpha - 1}\right\}, \quad \alpha > 0,$$

which is another class of the Archimedean copulas with $\phi(x) = -\log\left(\frac{1-\alpha^x}{1-\alpha}\right)$ and $\phi^{-1}(x) = \log_{\alpha}\{1-(1-\alpha)e^{-x}\}$. It has Kendall's τ given by

$$1 + \frac{4}{\log\alpha}\left(\frac{1}{\log\alpha}\int_0^{-\log\alpha}\frac{t}{e^t - 1}dt + 1\right).$$

An important property of this family is that the association can be either positive (when $\alpha < 1$) or negative (when $\alpha > 1$). The independence copula corresponds to $\alpha \to 1$.

The sample mean and sample standard deviations of the estimates based on 1000 replicates were calculated. The results of Settings 5.1 and 5.2 are summarized in Figures 5.1 and 5.3, respectively. Note that the estimated semi-Markov kernel based on the true copula function is approximately unbiased. The estimates based on the wrong copula function but the correct Kendall's τ are less biased than the estimates with the correct copula function but the wrong Kendall's τ. The estimates of the semi-Markov kernel from state 1 to state 2 are more sensitive to the correct specification of the association, Kendall's τ, than that from state 2 to state 1.

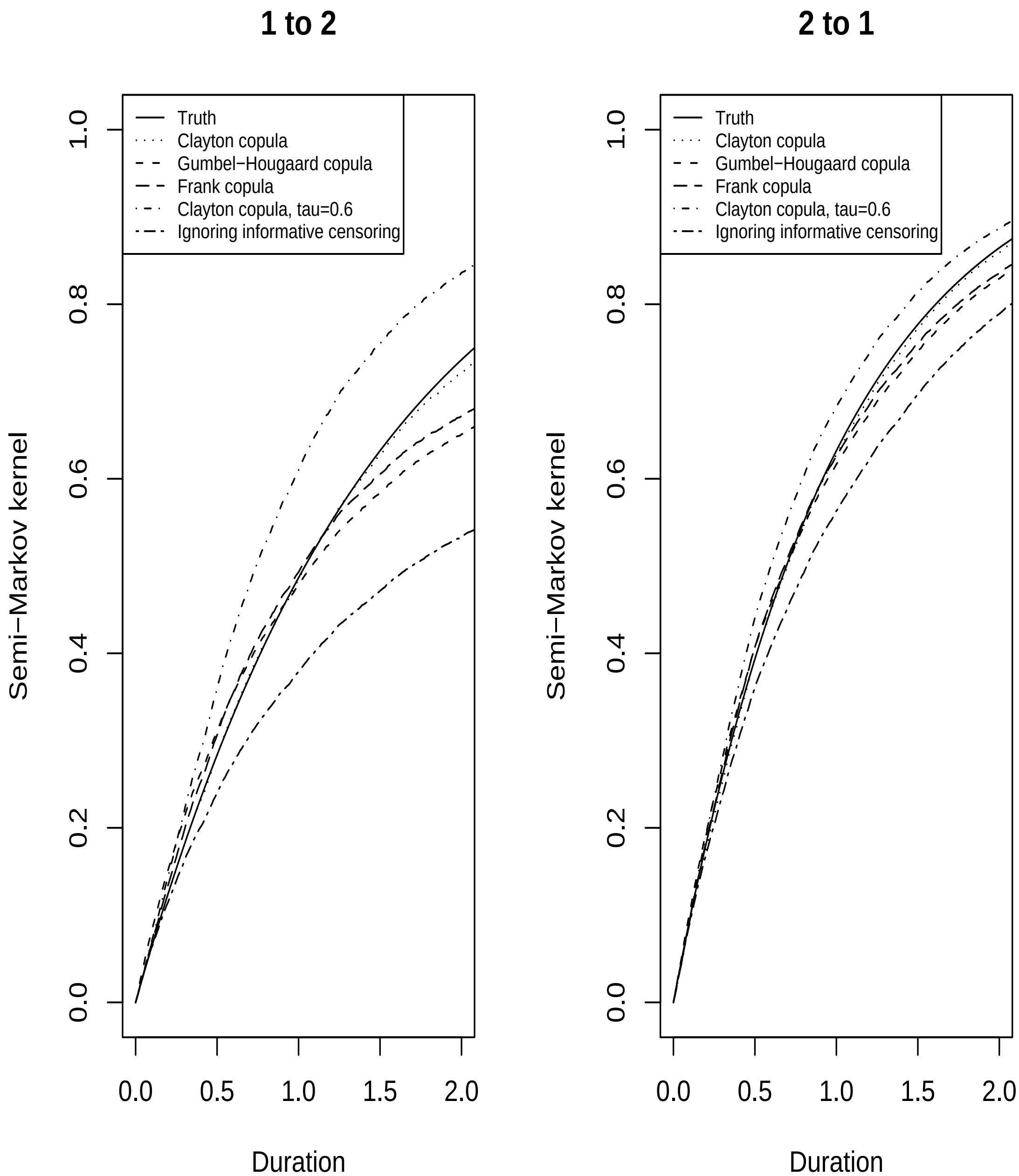

Figure 5.1: Truth and sample mean of estimated semi-Markov kernels in simulation setting 5.1 (Solid: truth; Dotted: Clayton copula and Kendall's $\tau = 1/3$; Short Dashed: Gumbel-Hougaard copula and Kendall's $\tau = 1/3$; Long dashed: Frank copula and Kendall's $\tau = 1/3$; Short dotted dash: Clayton copula and Kendall's $\tau = 0.6$; Long dotted dash: estimates ignoring informative censoring)

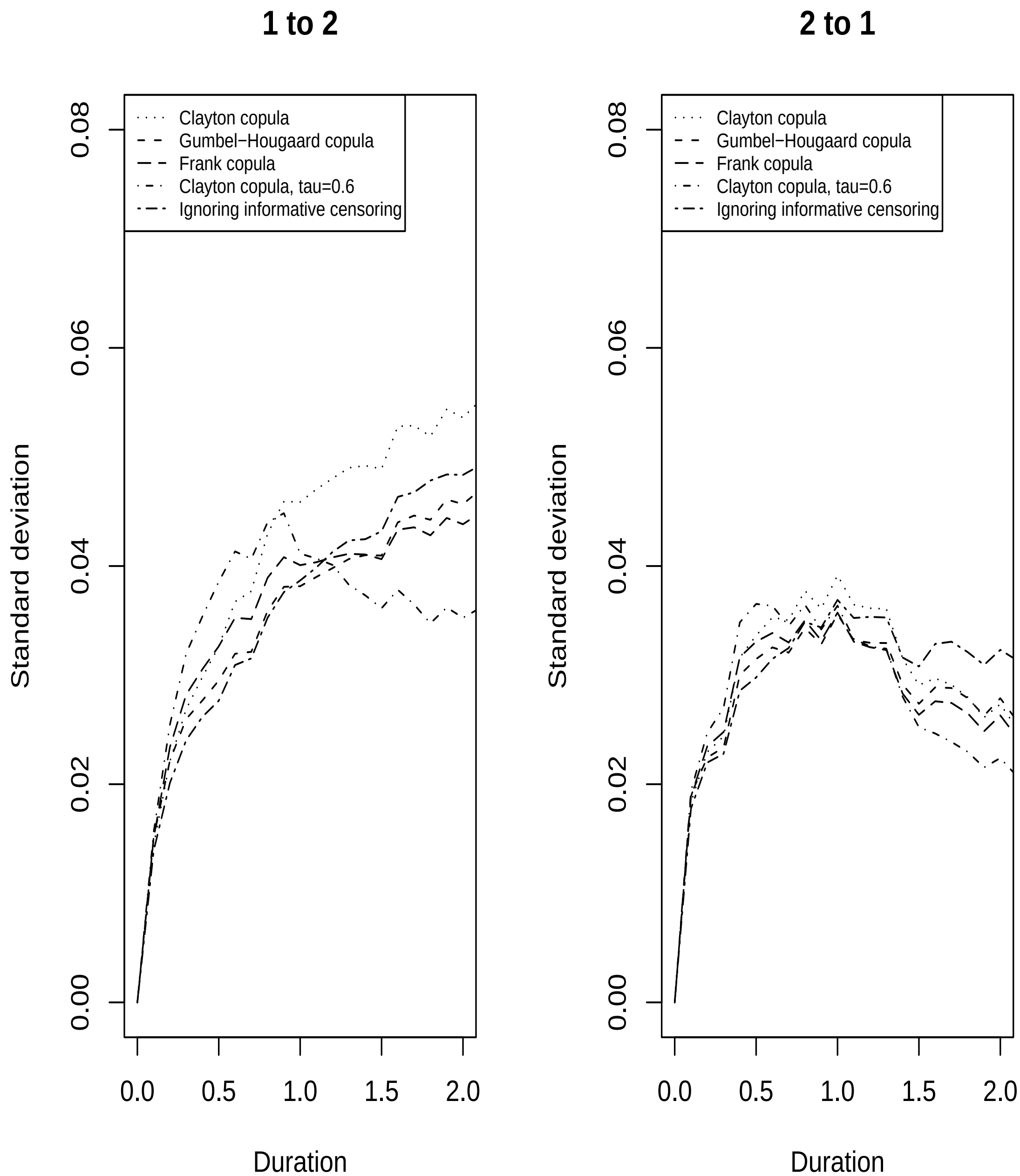

Figure 5.2: Sample standard deviation of estimated semi-Markov kernels in simulation setting 5.1 (Dotted: Clayton copula and Kendall's $\tau = 1/3$; Short Dashed: Gumbel-Hougaard copula and Kendall's $\tau = 1/3$; Long dashed: Frank copula and Kendall's $\tau = 1/3$; Short dotted dash: Clayton copula and Kendall's $\tau = 0.6$; Long dotted dash: estimates ignoring informative censoring)

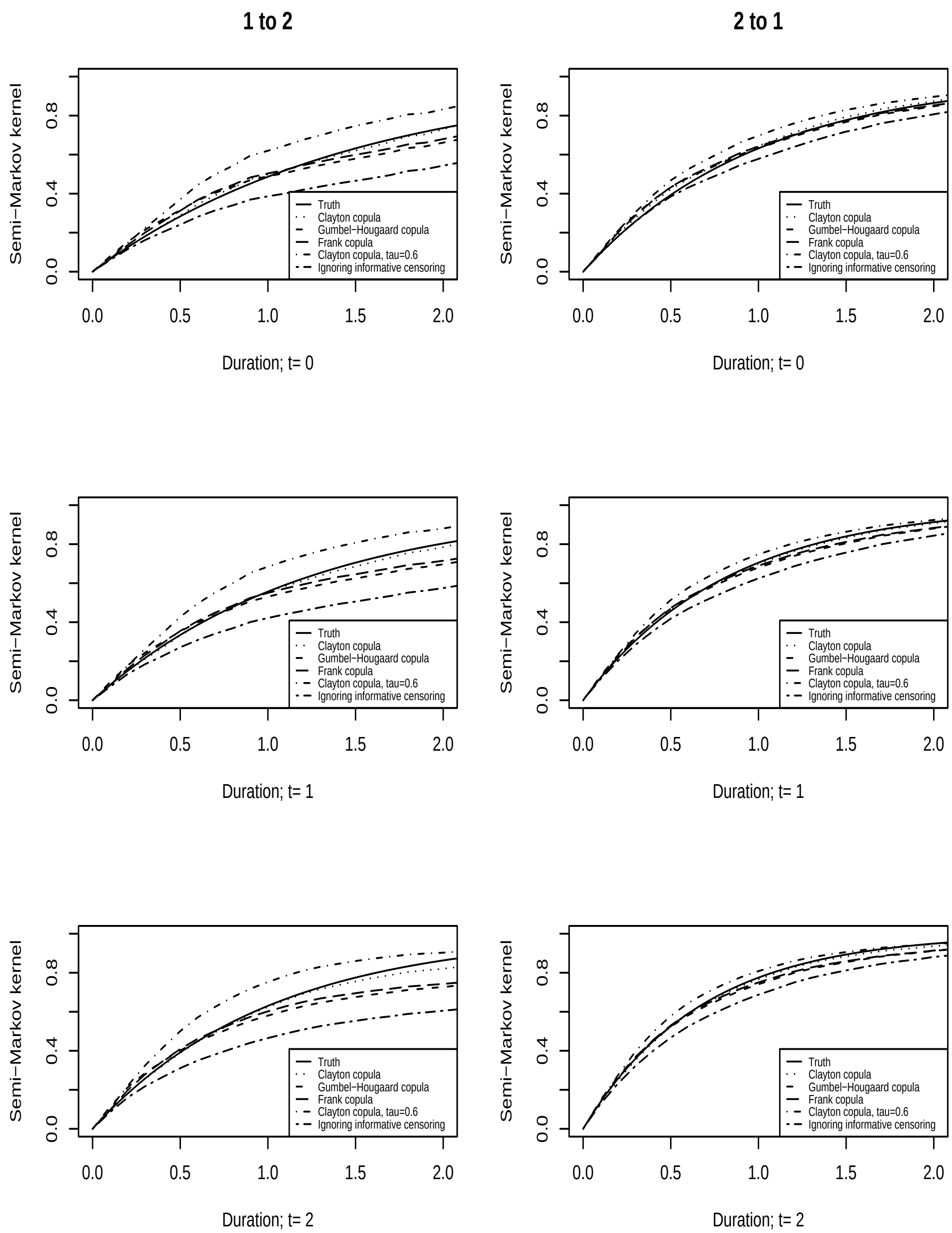

Figure 5.3: Truth and sample mean of estimated semi-Markov kernels in simulation setting 5.2 (Solid: truth; Dotted: Clayton copula and Kendall's $\tau = 1/3$; Short Dashed: Gumbel-Hougaard copula and Kendall's $\tau = 1/3$; Long dashed: Frank copula and Kendall's $\tau = 1/3$; Short dotted dash: Clayton copula and Kendall's $\tau = 0.6$; Long dotted dash: estimates ignoring informative censoring)

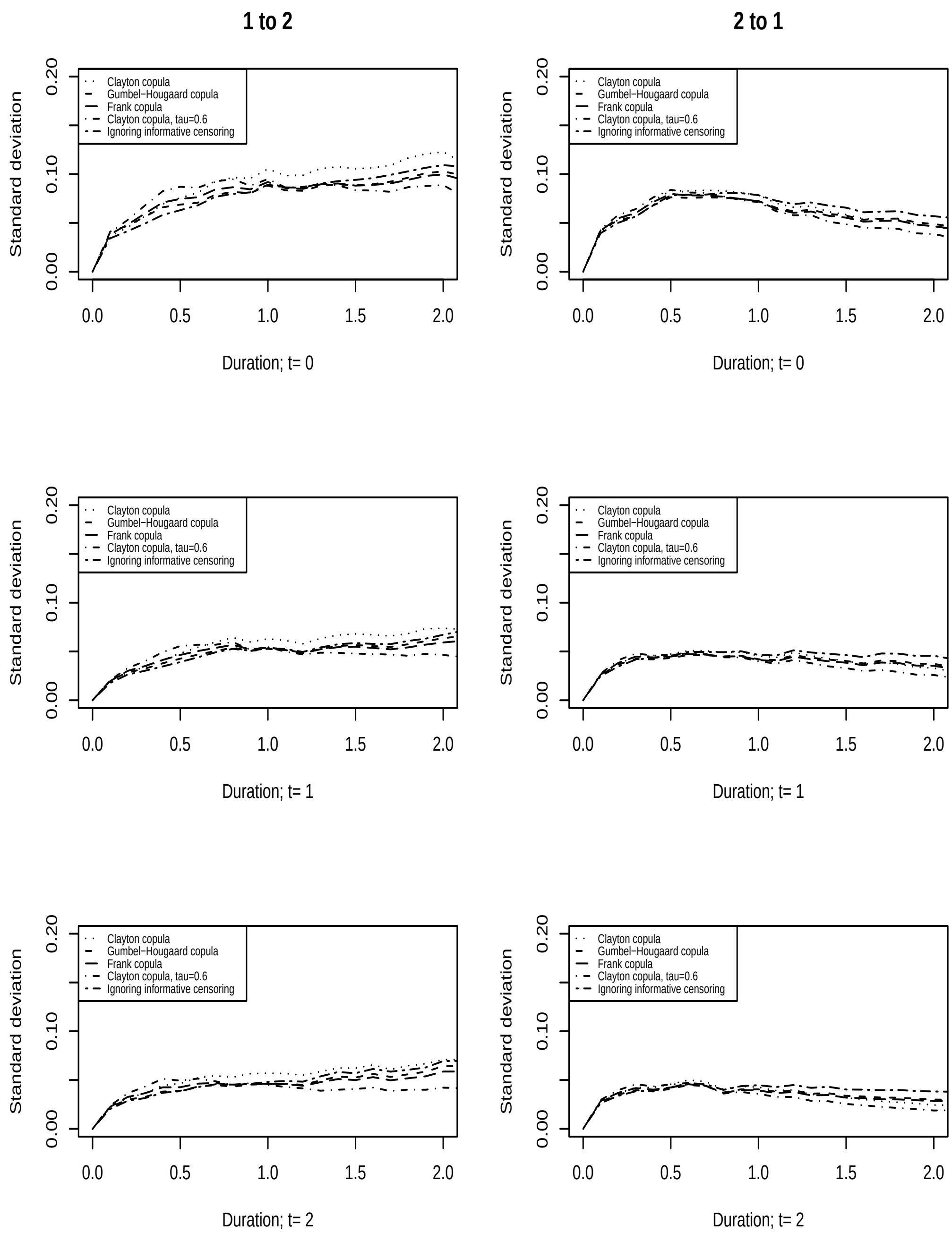

Figure 5.4: Sample standard deviation of estimated semi-Markov kernels in simulation setting 5.2 (Dotted: Clayton copula and Kendall's $\tau = 1/3$; Short Dashed: Gumbel-Hougaard copula and Kendall's $\tau = 1/3$; Long dashed: Frank copula and Kendall's $\tau = 1/3$; Short dotted dash: Clayton copula and Kendall's $\tau = 0.6$; Long dotted dash: estimates ignoring informative censoring)

5.5 Remarks

In this chapter, we proposed a model and corresponding estimation procedure for a type of informative right censored multi-state processes, specifically, semi-Markov processes. We assume that the censoring scheme has certain “renewal” property with respect to the semi-Markov processes. Under this assumption, the resulting process can be broken into pieces of competing risks according to every transition. The literature on dependent competing risks can then be adapted. We adapted the copula-graphic approach proposed by Zheng and Klein (1995). Simulation studies suggest that the inference procedure works well when the copula is correctly specified. When the prior knowledge about the copula is not available, our approach can be used in a sensitivity analysis for the assumption of noninformative censoring.

We focused on semi-Markov processes with two states. It is of interest to extend the approach to the situations with more states. There is some extra difficulty to specify the dependence between censoring and the semi-Markov processes, because the next state to be visited by the original process can not be determined by the current state occupied.

In the presence of continuous covariates, Heckman and Honore (1989) show that the nonidentifiability problem of competing risks can be solved under a quite general model structure (includes both proportional hazards and accelerated failure time models), given the covariates satisfy a quite strong assumption (eg, the range is the entire real line). Fermanian (2003) develops a nonparametric kernel estimator under the model of Heckman and Honore (1989). This method may potentially be adapted in the setting we consider.

Chapter 6

Applications

This chapter presents the applications of the proposed methods to the two real data sets described in Chapter 1.

6.1 Human Sleep Data

6.1.1 Description

Zung et al. (1965) model human sleep patterns with a Markov chain. Yang and Hursch (1973) test the Markov chain model with more data and find it likely inadequate in describing sleep stage sequences. They show that a semi-Markov model, which takes into account the duration in the current state, represents the underlying process better. By dividing the whole night into hourly intervals, they also find the time heterogeneity of the human sleep process.

We analyzed the human sleep data introduced in Example 1.2 applying the methods developed in this thesis. Two realizations of the sleep processes are shown in Figure 6.2. The frequency of all the observed transitions from the study, and median and mean of the sojourn times are shown in Table 6.1. Note that the number of direct Awake to REM transitions is small. More than half of the sojourn times in the state Awake are less equal than 30 seconds.

Table 6.1: Frequency of the observed transitions, and median and mean of the sojourn times

Transition	Number	Median[1]	Mean[1]
Awake to Non-REM	1660	0.5	2.1
Awake to REM	55	0.5	0.6
Awake to Censoring	30	6.5	11.9
Non-REM to Awake	1368	3.5	10.6
Non-REM to REM	756	2.5	10.3
Non-REM to Censoring	27	13.0	13.1
REM to Awake	377	5.5	7.3
REM to Non-REM	421	4.5	6.8
REM to Non-Censoring	13	3.5	5.8

[1] in minutes

6.1.2 Analyses of Human Sleep Data

6.1.2.1 Analysis with Homogeneous Semi-Markov Model

We first modeled the sleep process with the HSM model. Note from Table 6.1 that the proportion of censoring is low from every state so that the transition probabilities can be well estimated. The estimated transition probabilities of the embedding Markov chain are summarized in Table 6.2 and Figure 6.1. All approaches, including the robust approach, give similar estimates and confidence intervals for the transition probabilities because of the negligible censoring. This indicates that the efficiency of the robust approach is comparable with other approaches in the application. The probability of the direct transition from Awake to REM is small (0.032 with 95% CI 0.021 to 0.042). The probability of the direct transition from Non-REM to REM is estimated as 0.356 with 95% CI 0.313 to 0.399. About half of REM sleep transits to Awake directly (0.473 with 95% CI 0.421 to 0.524).

The confidence bands of the semi-Markov kernel is presented in Figures 6.3. The plot of estimated $Q_{12}(\cdot)$ in Figure 6.3 indicates that, starting from Awake, about 90% of subjects stay at non-REM sleep within 30 minutes. From the plots of estimated $Q_{21}(\cdot)$ and $Q_{23}(\cdot)$ in Figure 6.3, we see that about 50% of subjects starting from non-REM sleep turn to Awake in 30 minutes, about 30% of them move to REM sleep, and about 20% of them remain in non-REM sleep in 30 minutes. Figure 6.4 shows the confidence bands for the sojourn time distributions based on the robust approach.

Table 6.2: Estimates and 95% confidence intervals of the transition probabilities for the human sleep data (1: Awake, 2: Non-REM, 3: REM)

		Plug-in	Normalized	Phelan	Robust approach
P_{12}	Estmate	0.968	0.968	0.968	0.968–0.968
	CI	(0.958, 0.979)	(0.958, 0.979)	(0.957, 0.979)	(0.958, 0.979)
P_{13}	Estmate	0.032	0.032	0.032	0.032–0.032
	CI	(0.021, 0.042)	(0.021, 0.042)	(0.021, 0.043)	(0.021, 0.042)
P_{21}	Estmate	0.644	0.644	0.644	0.644–0.644
	CI	(0.601, 0.687)	(0.601, 0.687)	(0.601, 0.687)	(0.601, 0.687)
P_{23}	Estmate	0.356	0.356	0.356	0.356–0.356
	CI	(0.313, 0.399)	(0.313, 0.399)	(0.313, 0.399)	(0.313, 0.399)
P_{31}	Estmate	0.473	0.473	0.472	0.473–0.473
	CI	(0.421, 0.524)	(0.421, 0.524)	(0.421, 0.524)	(0.421, 0.524)
P_{32}	Estmate	0.527	0.527	0.528	0.527–0.527
	CI	(0.476, 0.579)	(0.476, 0.579)	(0.476, 0.579)	(0.476, 0.579)

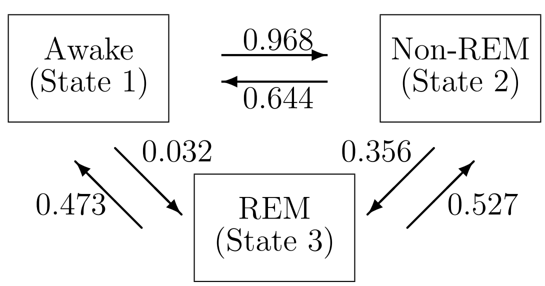

Figure 6.1: Estimated transition probabilities for human sleep process

The plot of estimated $F_{12}(\cdot)$ indicates that 90% of the sojourn times from Awake to non-REM sleep are less than 15 minutes.

6.1.2.2 Analysis with Modulated Semi-Markov Model

We then modeled the sleep process as a three-state modulated semi-Markov process. We incorporated the study time since the onset of sleep as a time-dependent covariate, denoted by $Z(t)$, and allowed this covariate to have different regression effect on different transitions. The specific model we used is given by

$$\alpha_{hj}(t|\mathcal{F}_t, Z_{hj}(t)) = \alpha_{0hj}\left(B(t)\right) e^{\theta_{hj} Z(t)}, \tag{6.1.1}$$

for $h \neq j \in \{1, 2, 3\}$. We used hour as the metric of time.

The estimated regression parameter θ_{hj}'s are summarized in Table 6.3. Note that none of the 95% confidence intervals of the regression parameters contains 0, which indicates that all the transitions are nonhomogeneous in the study time scale. Specifically, the transition rates from Awake to Non-REM sleep and REM to Non-REM are decreasing during the night, while the transition rates from Awake to REM, Non-REM to Awake, Non-REM to REM, and REM to Awake are increasing during the night.

Table 6.3: Point estimate, standard error and confidence interval of the regression parameters for the human sleep data

Transition	Estimate	SE	CI
Awake to Non-REM	-0.023	0.010	(-0.043, -0.002)
Awake to REM	0.164	0.062	(0.042, 0.286)
Non-REM to Awake	0.038	0.012	(0.014, 0.062)
Non-REM to REM	0.154	0.017	(0.121, 0.187)
REM to Awake	0.096	0.028	(0.042, 0.150)
REM to Non-REM	-0.057	0.025	(-0.107, -0.008)

6.1.2.3 Analysis with Nonhomogeneous Semi-Markov Model

In the analysis with the modulated semi-Markov model, we have found evidence that the sleep process is not homogeneous in study time. Thus the NHSM model is likely a better model for the process.

We modeled the sleep process as a NHSM process with three states: Awake, Non-REM and REM. We considered four different specifications: the fully nonparametric, the piecewise constant, the nonparametric multiplicative, and the semiparametric models. Eight equally spaced partitions on the study time scale were used in the piecewise constant approach. The semiparametric model is just (6.1.1) considered in Section 6.1.2.2. The bandwidths were $w = 1.5$ hours in the study time scale, and $b = 0.5$ hour in the duration time scale.

Figures 6.5 to 6.10 present the 3-dimensional plots of the estimated transition rate functions in the two time scales. It appears that the transition rate functions vary with cyclic patterns in the study time since the onset of sleep, which indicates the time non-homogeneity in the sleep process. The semiparametric estimates, however, can not capture the cyclic patterns of the transition rates in the study time scale due to the model specification. For instance, Figures 6.5 and 6.6 show that the estimates based on the first three approaches suggest that the transition rates from Awake to Non-REM and REM be cyclic with a peak at about 3 hours after onset of sleep, but the semiparametric estimate fails to capture the peak. In Figures 6.7 and 6.9, the transition rates from Non-REM and REM to Awake are higher at the end of the night (about 7 hours after onset of sleep). This finding is consistent with Kneib and Hennerfeind (2008).

The transition rate functions also depend on the duration in the current state. From Figure 6.5, we see that the transition rate function from Awake to Non-REM is increasing to the peak after about 15 minutes in the state Awake, and then starts to decrease. From Figure 6.9 and 6.10, we find the transition rates from REM to Awake and Non-REM are maximized after about half an hour in the present state. Figure 6.8 shows that the peak for transition from Non-REM to REM lies between 1 and 1.5 hours in the present state.

Figures 6.11 and 6.12 present the estimated semi-Markov kernels. From Figure 6.11, we find that starting from Awake, the sleep process will very likely transit to Non-REM other than REM directly. It takes longer to wake up from Non-REM at the beginning of the night than at the end of the night.

6.1.2.4 Model Checking

To check whether the study time and the duration time have multiplicative effects on the transition intensities, we plot the positive and negative parts of the difference between the estimated cumulative transition rate functions associated with the nonparametric model and the nonparametric multiplicative model in Figures 6.13 and 6.14. Note that for the transition from Awake to Non-REM, we can see the pattern that the positive residuals are concentrated at small and large study times, while the negative residuals have intermediate study times. This suggests that the multiplicative model assumption is questionable in this application.

6.1.3 Concluding Remarks

We applied the methods developed in this thesis to the human sleep data. The analysis outcomes suggest that the intensities of transitions among the sleep patterns vary in both the study and duration time scales. Thus NHSM model is likely a plausible model for the process. The semiparametric specification, which assumes a Cox regression form for the study time, can not capture the cyclic pattern of the transition rate functions in the study time scale. The preliminary model checking shows that the two time scales may not affect the transition rate functions multiplicatively.

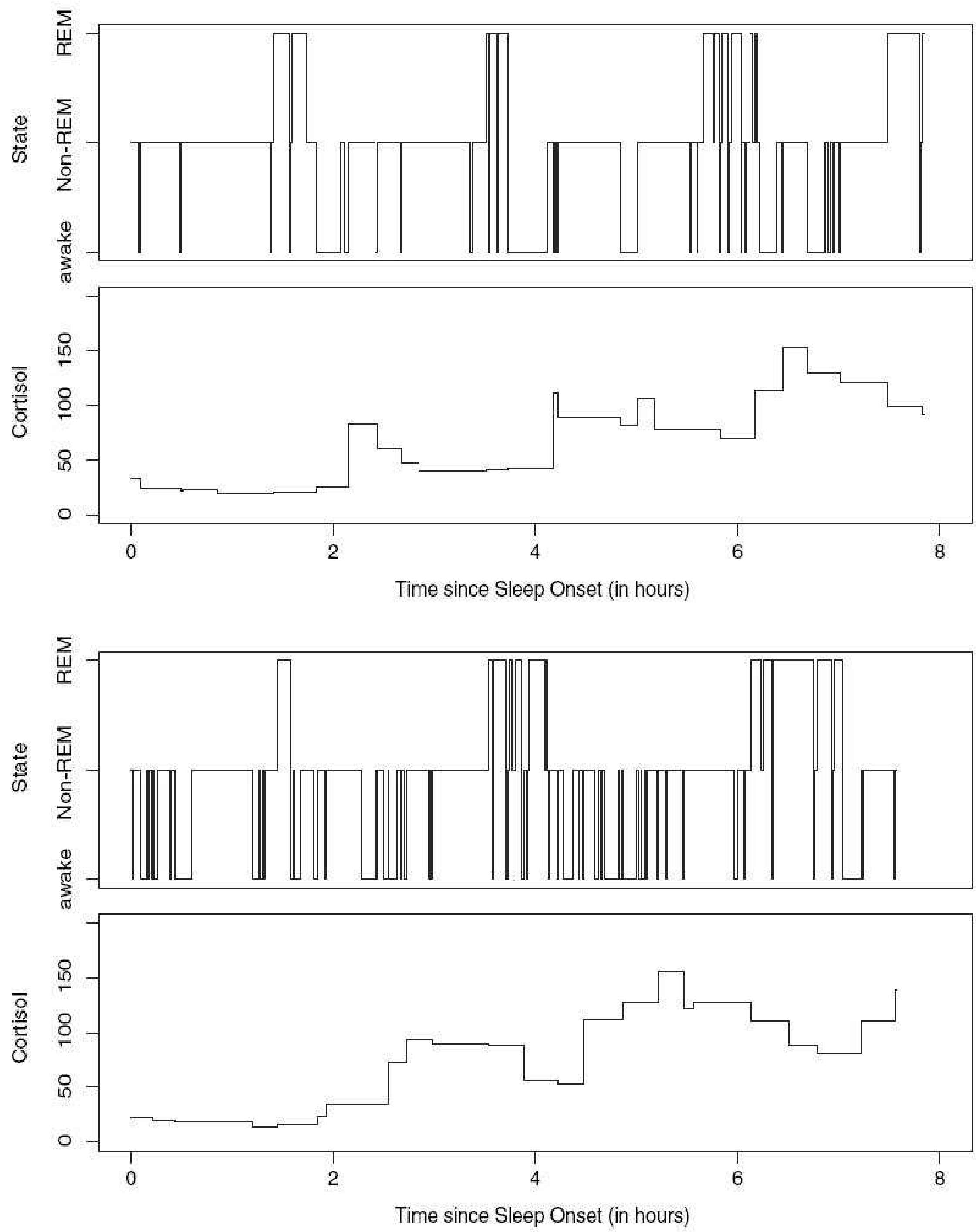

Figure 6.2: Realizations of two individual sleep processes and corresponding nocturnal cortisol secretion, cited from Kneib and Hennerfeind (2008)

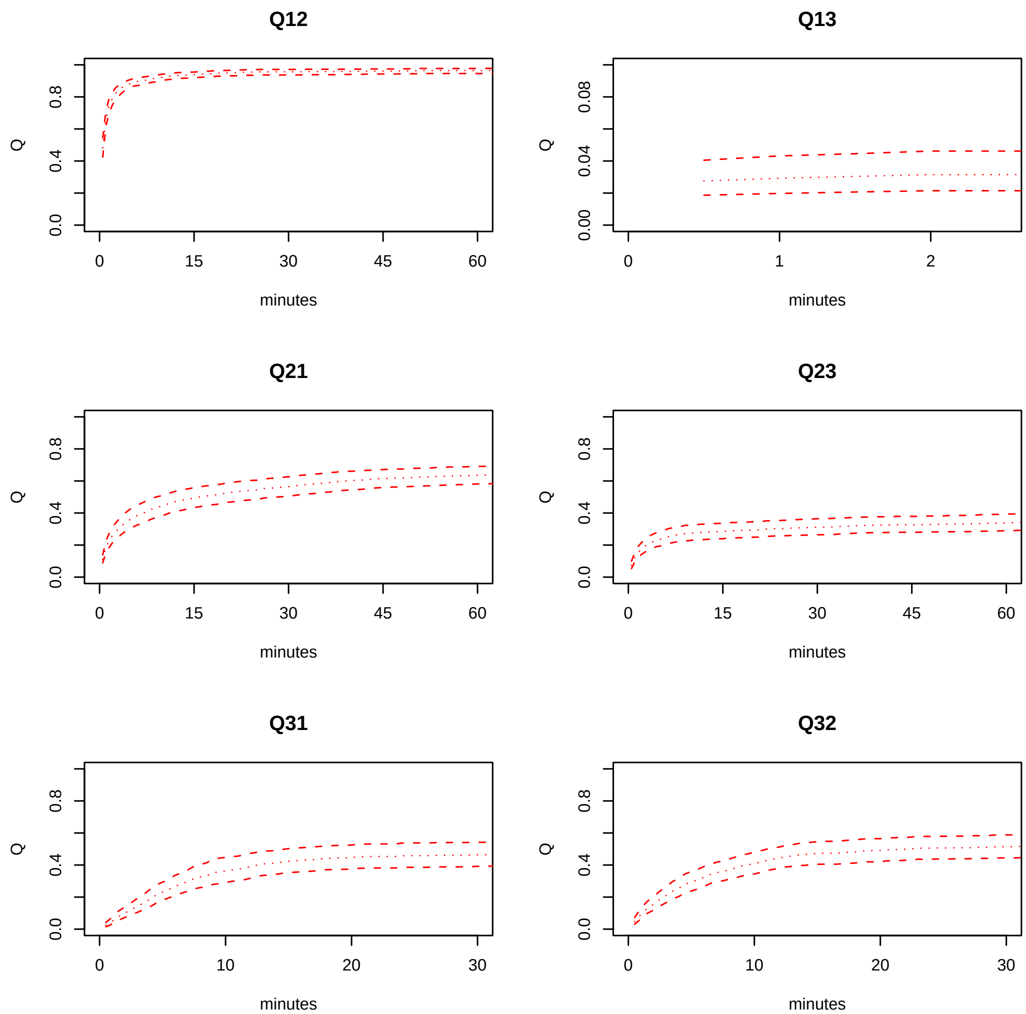

Figure 6.3: Estimated semi-Markov kernel for human sleep data (Dotted: point estimate; Dashed: 95% confidence bands)

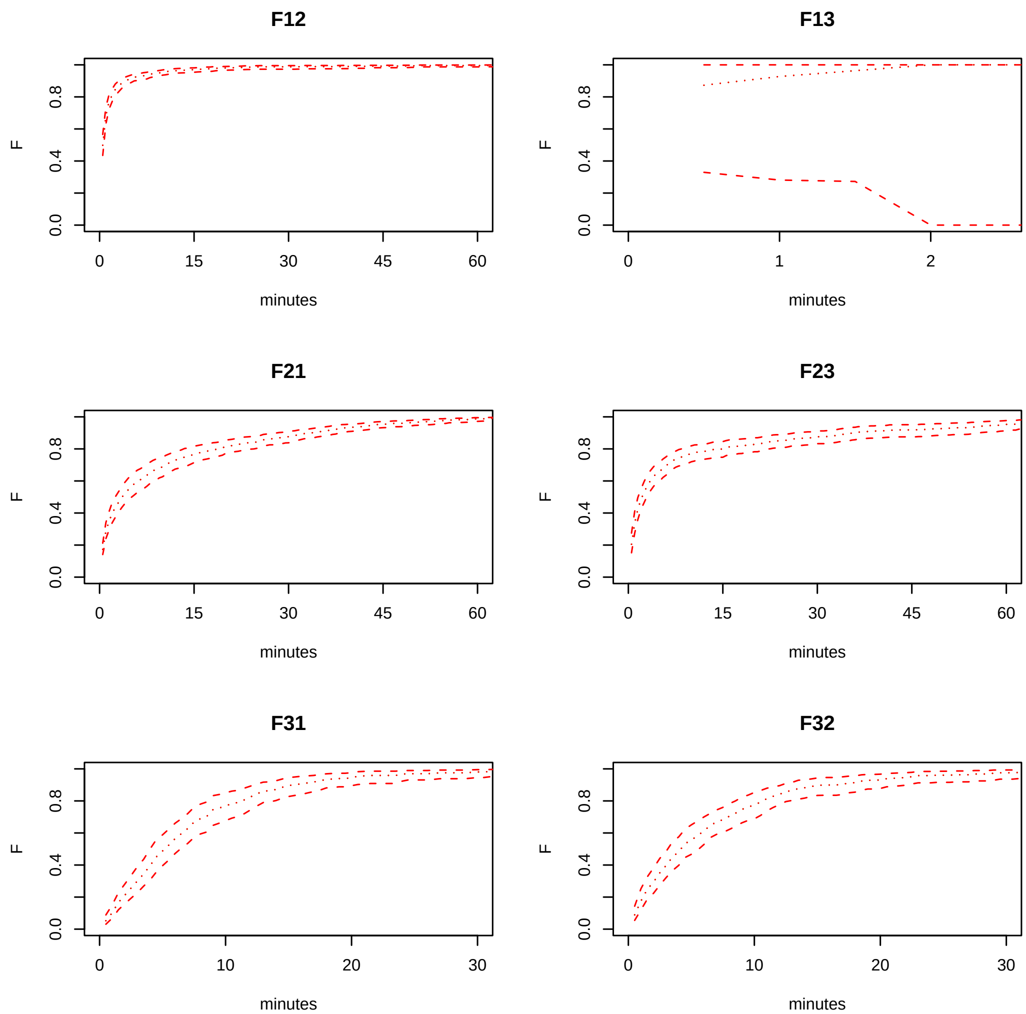

Figure 6.4: Estimated distribution of sojourn times for human sleep data (Dotted: bound estimate; Dashed: 95% confidence bands)

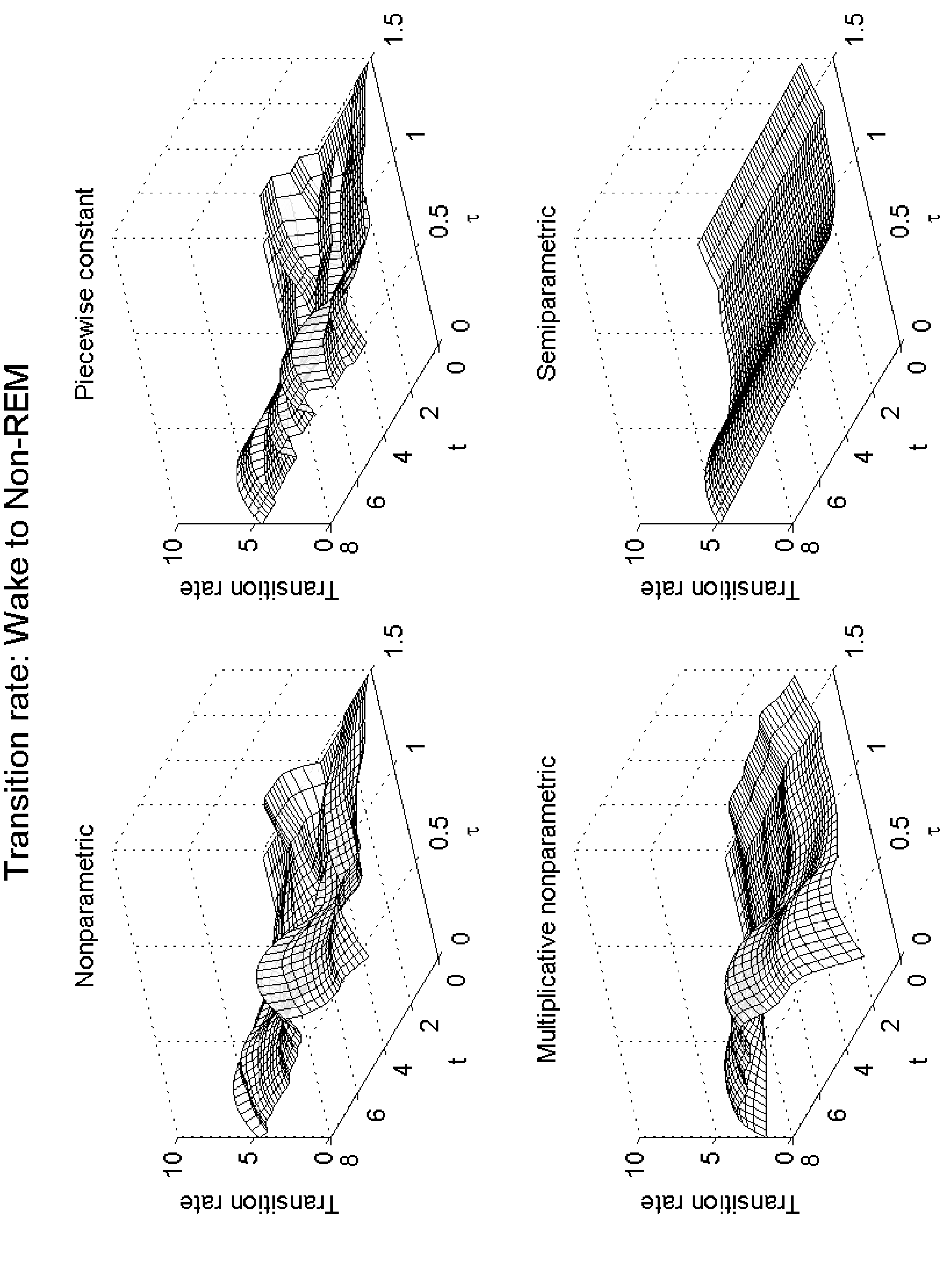

Figure 6.5: Estimated transition rate functions from Awake to Non-REM

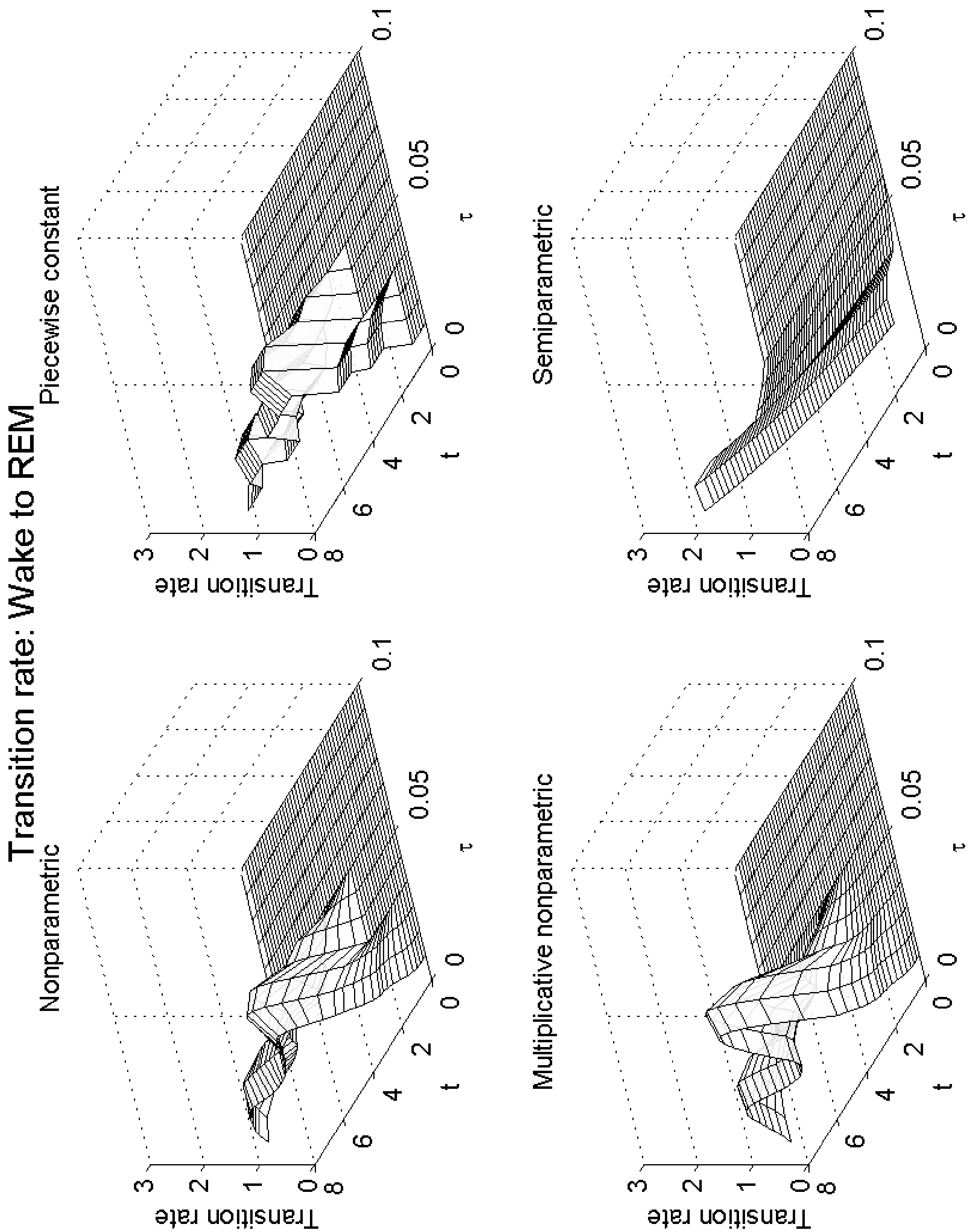

Figure 6.6: Estimated transition rate functions from Awake to REM

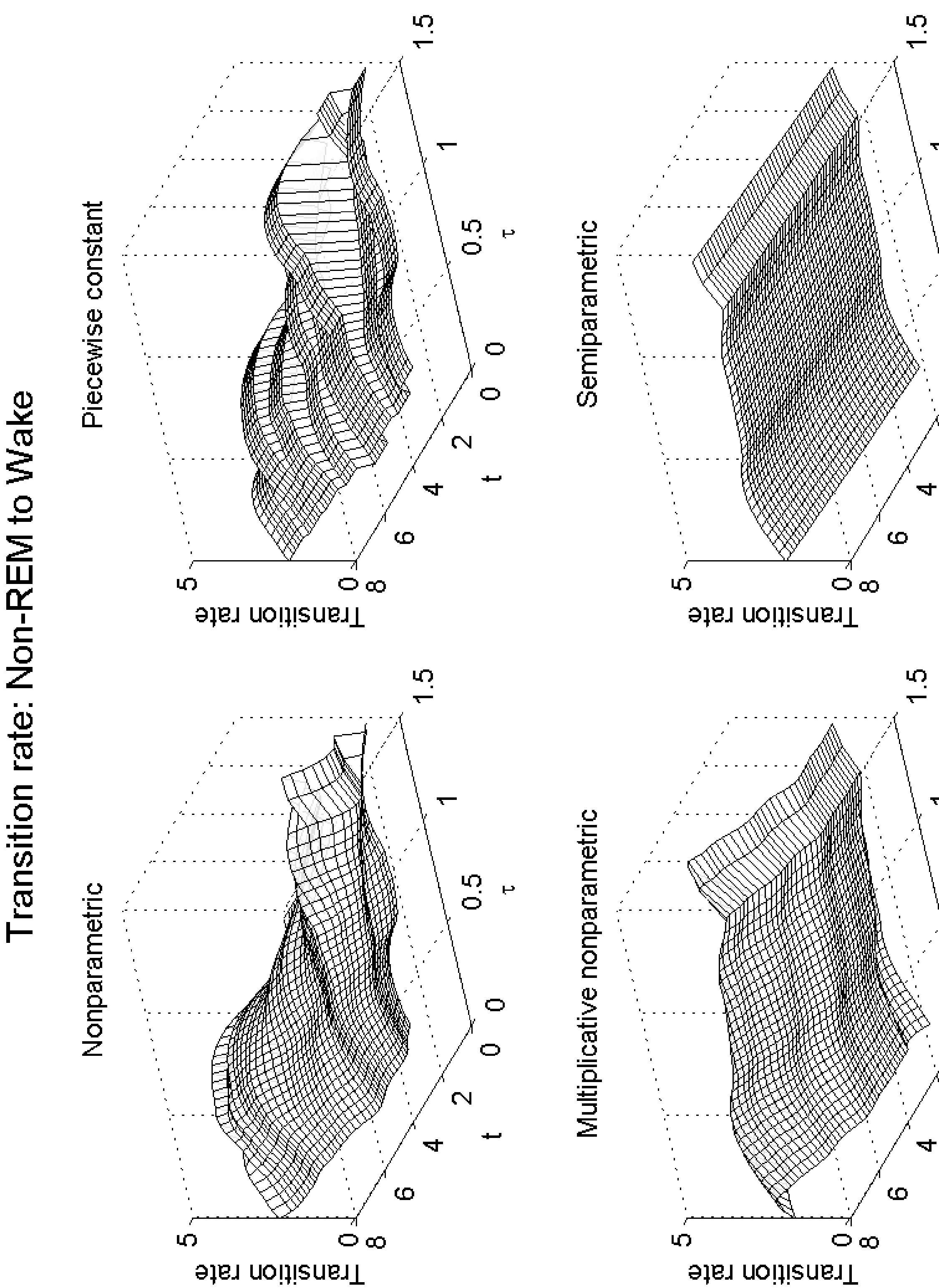

Figure 6.7: Estimated transition rate functions from Non-REM to Awake

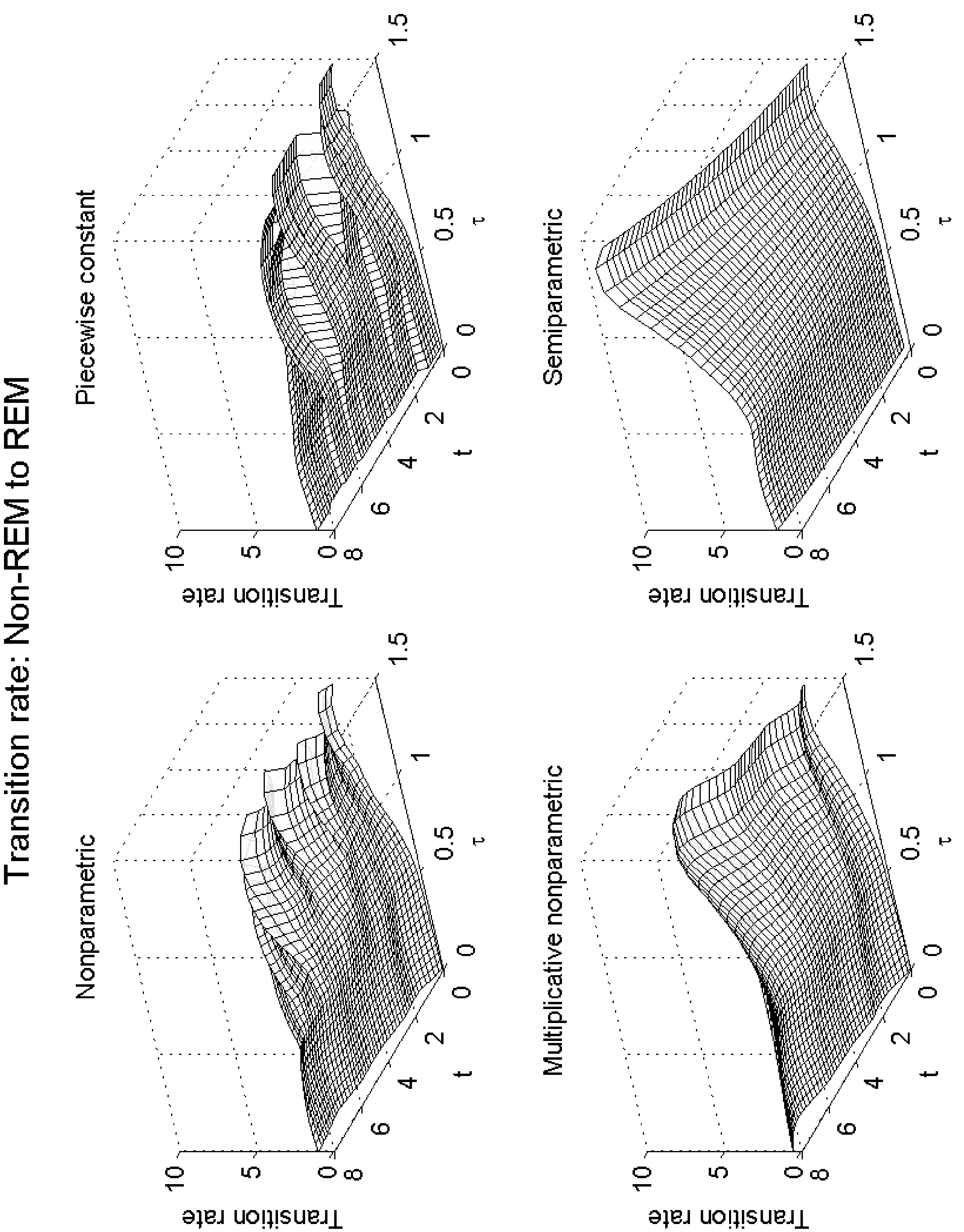

Figure 6.8: Estimated transition rate functions from Non-REM to REM

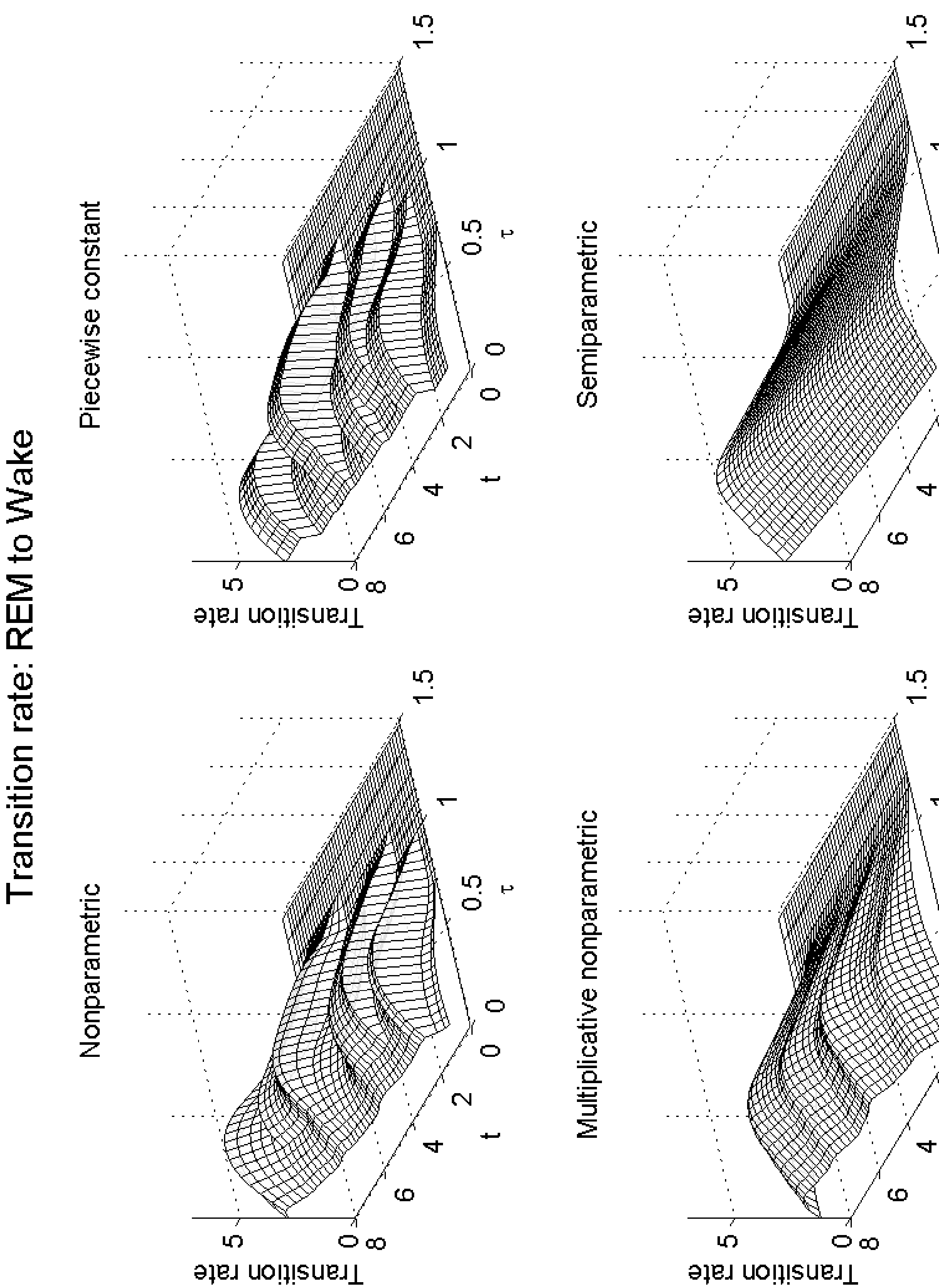

Figure 6.9: Estimated transition rate functions from REM to Awake

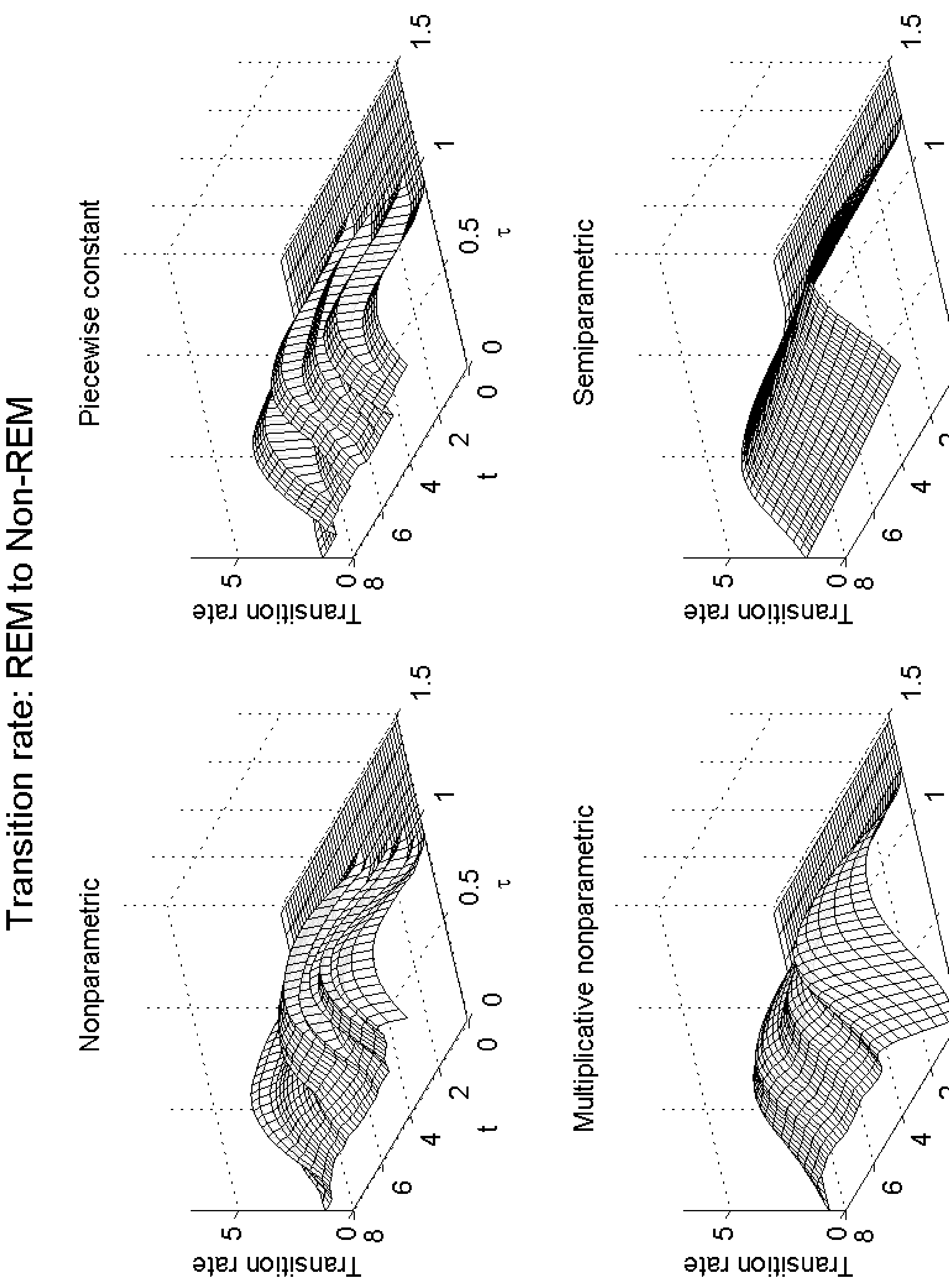

Figure 6.10: Estimated transition rate functions from REM to Non-REM

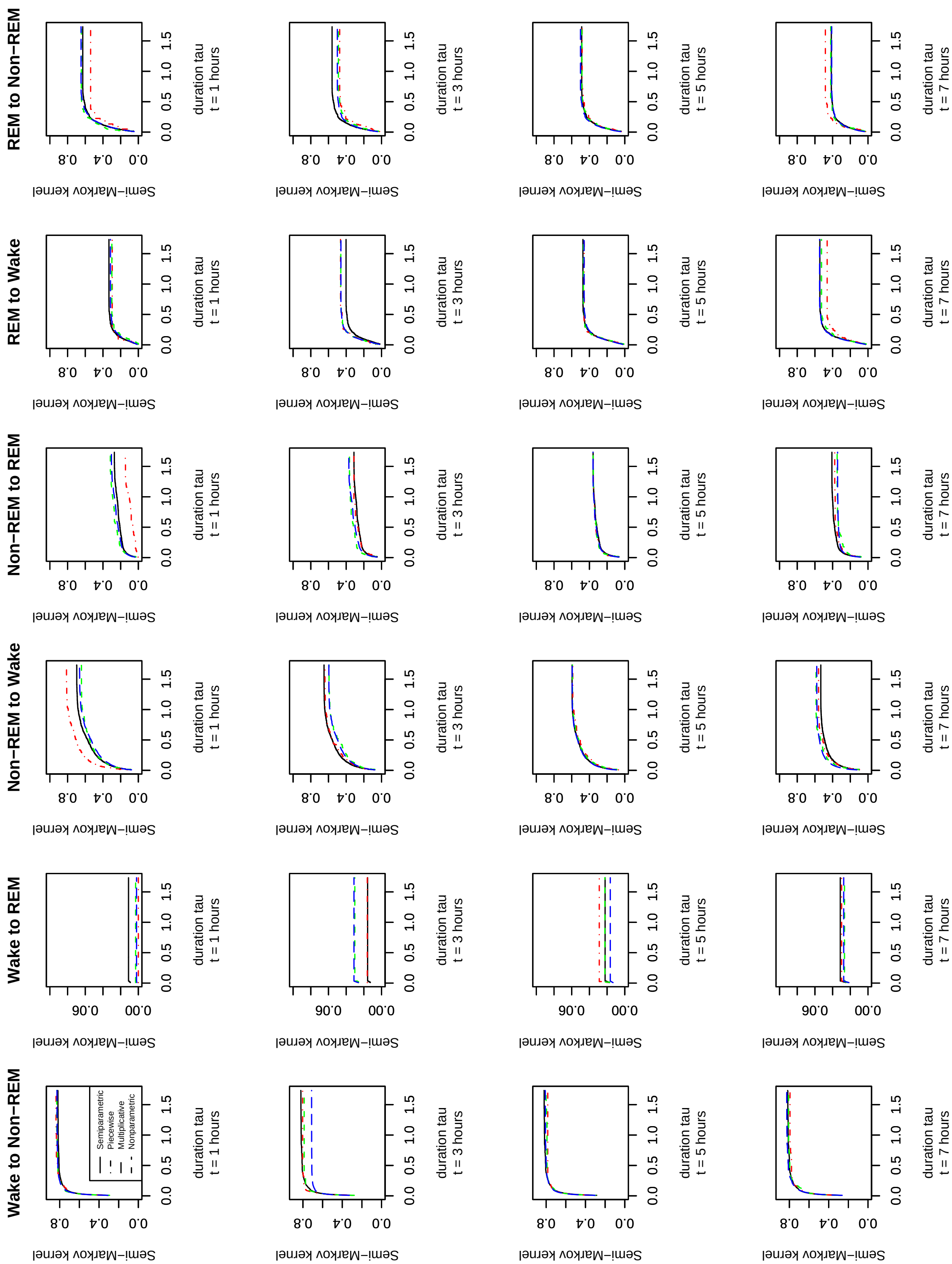

Figure 6.11: Estimated semi-Markov kernel for fixed t

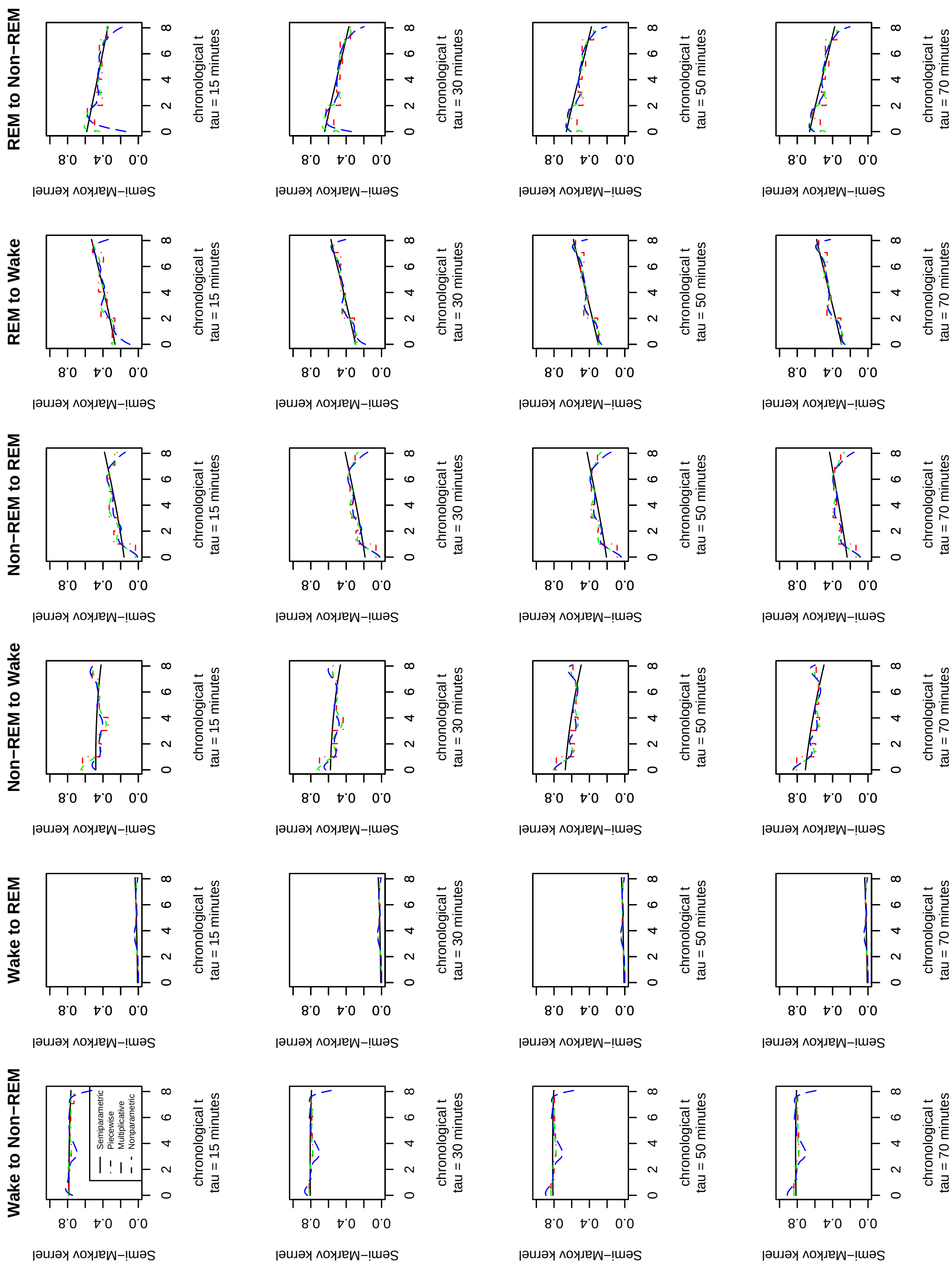

Figure 6.12: Estimated semi-Markov kernel for fixed τ

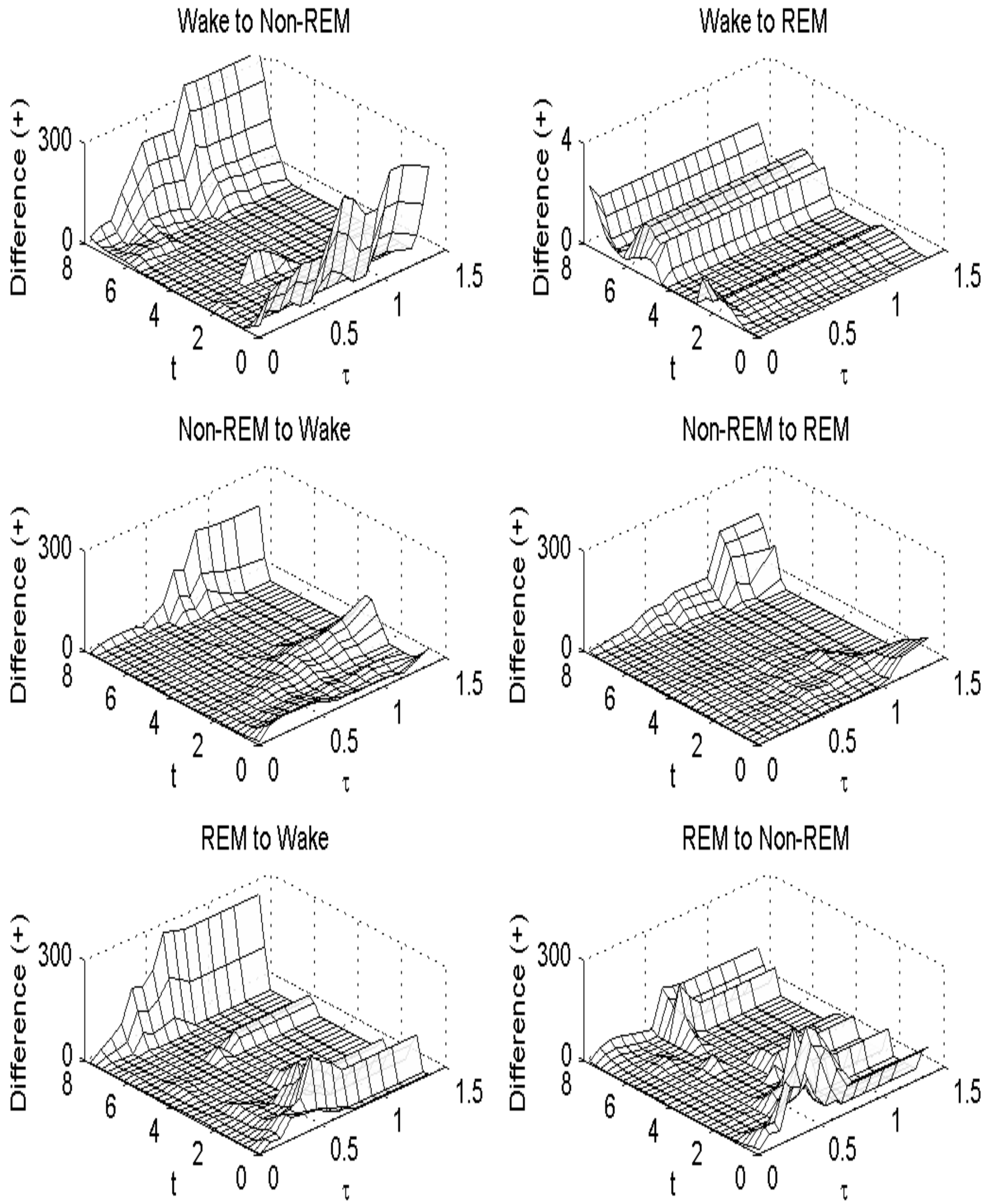

Figure 6.13: The positive parts of the difference between the estimated cumulative transition rate functions associated with the nonparametric model and the nonparametric multiplicative model

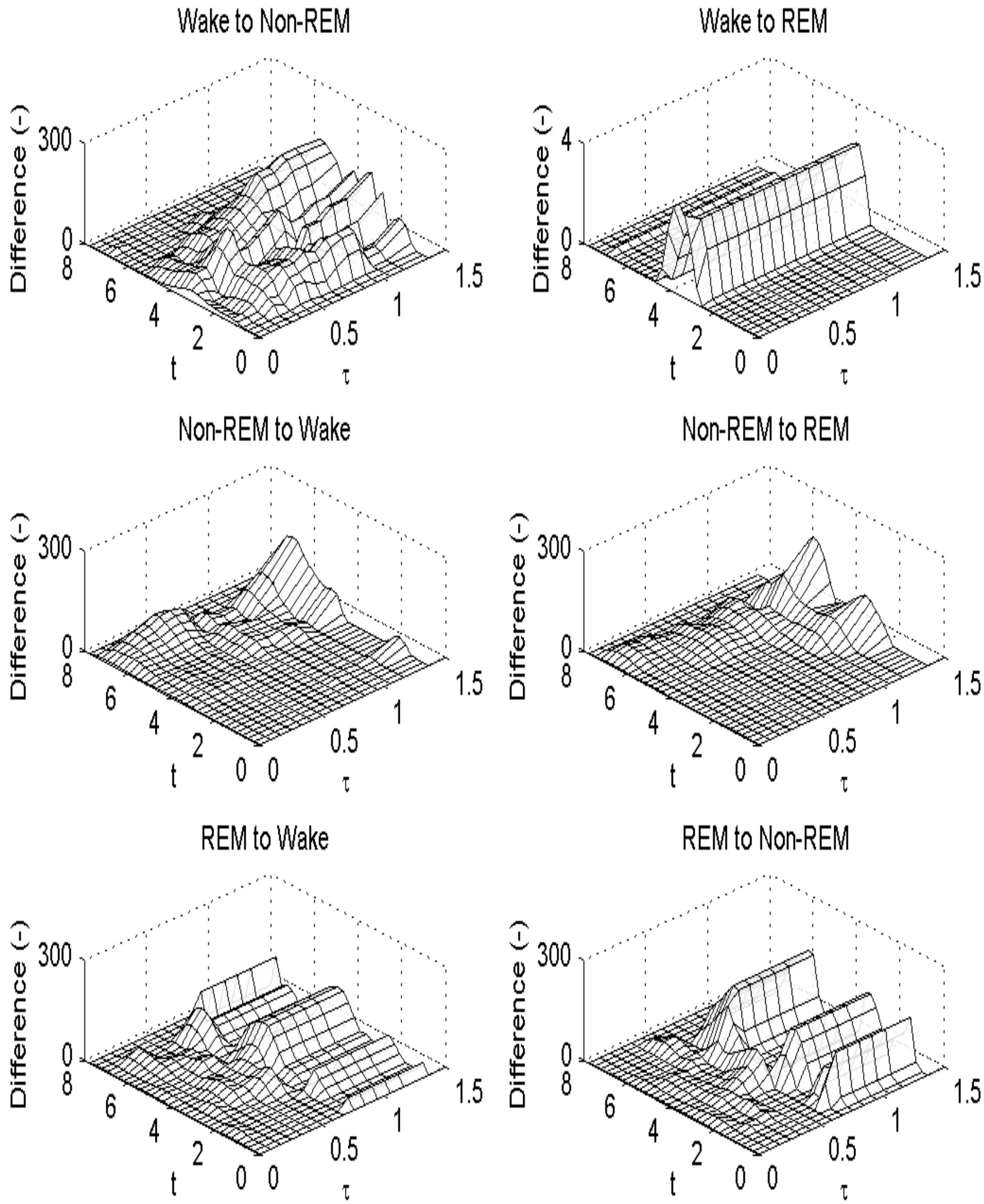

Figure 6.14: The negative parts of the difference between the estimated cumulative transition rate functions associated with the nonparametric model and the nonparametric multiplicative model

6.2 Hospitalization Data

6.2.1 Description

We applied the proposed methods to the hospitalization data introduced in Example 1.1. We formulated the hospitalization process as a multi-state process with 3 states (see Figure 1.2): 1 for "out of hospital", 2 for "in hospital", and 3 for "dead". Each of the multi-state process starts at 5 years after the diagnosis date from state 1, i.e., "out of hospital". It then transits between state 2 and 1 (i.e., being admitted to and discharged from hospital) before entering the absorbing state 3 (i.e., dead), or being censored on Dec 31, 2000. The frequency of hospital admissions accross different diagnosis year windows is presented in Table 6.4. Among the 1374 subjects, 810 subjects were censored without a hospital admission. The frequency of all the observed transitions from the study, and median and mean of the sojourn times are shown in Table 6.5. In total there were 60 deaths observed, 29 of which were in hospital. All the censorings occurred in state 1.

Table 6.4: Frequency of the number of hospital admissions

	Number of hospital admissions					
Diagnosis Year	0	1	2	3–5	≥ 6	Total
1981–1989	313	170	62	102	95	742
1990–1995	497	79	37	12	7	632
1981–1995	810	249	99	114	102	1374

Table 6.5: Frequency of the observed transitions, and median and mean of the sojourn times (1: out of hospital, 2: in hospital, 3: dead)

Transition	Number	Median[1]	Mean[1]
$1 \to 2$	2148	126.0	461.2
$1 \to 3$	31	72.0	221.7
$1 \to$ censoring	1314	1436.5	1751.7
$2 \to 1$	2119	1.0	4.6
$2 \to 3$	29	8.0	16.8
$2 \to$ censoring	0	-	-

[1] in days

6.2.2 Analyses of Hospitalization Data

6.2.2.1 Analysis with Homogeneous Semi-Markov Model

We fitted the HSM model with the data first. The estimated transition probabilities of the embedding Markov chain are presented in Table 6.6 and Figure 6.15. All approaches give similar estimates and confidence intervals for P_{21} and P_{23} because that no transition was censored at state 2, i.e., "in hospital". Note that the probability of death at hospital is small (0.014 with 95% CI 0.009 to 0.018).

Table 6.6: Estimates and 95% confidence intervals of the transition probabilities for the hospitalization data (1: out of hospital, 2: in hospital, 3: dead)

		Plug-in	Normalized	Phelan	Robust approach
P_{12}	Estmate	0.775	0.988	0.986	0.775–0.991
	CI	(0.738, 0.813)	(0.984, 0.992)	(0.981, 0.991)	(0.738, 0.994)
P_{13}	Estmate	0.009	0.012	0.014	0.009–0.225
	CI	(0.006, 0.013)	(0.008, 0.016)	(0.009, 0.019)	(0.006, 0.262)
P_{21}	Estmate	0.986	0.986	0.986	0.986–0.986
	CI	(0.982, 0.991)	(0.982, 0.991)	(0.982, 0.991)	(0.982, 0.991)
P_{23}	Estmate	0.014	0.014	0.014	0.014–0.014
	CI	(0.009, 0.018)	(0.009, 0.018)	(0.009, 0.018)	(0.009, 0.018)

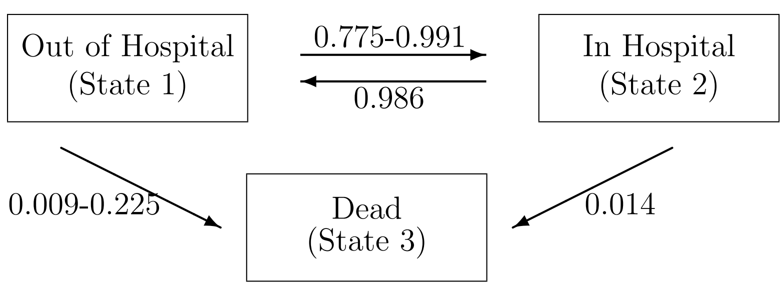

Figure 6.15: Estimated transition probabilities for the hospitalization process

However, the estimates for P_{12} and P_{13} are quite different with different approaches. From Tables 6.4, 810 subjects were censored directly from state 1, i.e., "out of hospital", without a hospital admission. This may due to the long tail of the sojourn time distribution in state 1. Thus the condition (2.3.10) required for the consistency of the plug-in estimator is questionable. On the other hand, Table 6.5 shows that the mean and median of the sojourn times in state 1 are quite different with respect to the next state to be visited. Figure 6.16 presents the estimates and confidence bands for the sojourn time distributions starting from state 1 using the normalized estimator. It shows that the two distributions $F_{12}(\cdot)$ and $F_{13}(\cdot)$ are quite different. This indicates that the inferences on the transition probabilities based on the normalized and Phelan's estimators can be biased. Thus safe conclusion concerning P_{12} and P_{13} should be drawn based on the robust approach.

Confidence bands for the semi-Markov kernel is presented in Figures 6.17. The transformed Hall-Wellner (HW) bands are wider than the transformed equal precision (EP) bands at the beginning, and then becomes narrower. It appears that about 90% of patients in hospital survive and are discharged from hospital within 15 days (see the plot of estimated $Q_{21}(\cdot)$), and about 1% of patients admitted to hospital die at the hospital within a month (see the plot of $Q_{23}(\cdot)$). From the plots of estimated $Q_{12}(\cdot)$ and $Q_{13}(\cdot)$, we see that about 50% of patients survive beyond 2.5 years without hospitalization, and less than 1% of patients die out of hospital within 2 years of discharge without further hospitalization.

Figure 6.18 shows the confidence bands for the sojourn time distributions based on the robust approach. The confidence band for $F_{12}(\cdot)$ is wide compared with $F_{21}(\cdot)$ as a result of the uncertainty in estimating P_{12}. The curve of estimated $F_{21}(\cdot)$ in Figure 6.18 shows that about 95% of the hospitalization duration (i.e., the sojourn time from in hospital to out of hospital) is less than 15 days.

6.2.2.2 Analysis with Modulated Semi-Markov Model

We then applied the modulated semi-Markov model to the hospitalization data. We incorporated the study time since the entrance of the study as a time-dependent covariate $Z(t)$, and allow this covariate to have different regression effect on different transitions. The specific model we used is

$$\alpha_{hj}(t|\mathcal{F}_t, Z_{hj}(t)) = \alpha_{0hj}\left(B(t)\right) e^{\theta_{hj} Z(t)}, \tag{6.2.1}$$

for $h \neq j \in \{1, 2, 3\}$ and $j \neq 3$. We transform the metric of time to year.

The estimates of the regression parameters are presented in Table 6.7. Note that neither of the 95% confidence intervals of the regression parameters between "in hospital" and "out of hospital" contains 0, which indicates the time nonhomogeneity of the hospitalization process. Specifically, the transition rate from "out of hospital" to "in hospital" is decreasing while the transition rate from "in hospital" to "out of hospital" is increasing with study time scale. The standard errors of the regression parameters of "out of hospital" or "in hospital" to death are large compared to the point estimates due to the small number of deaths occurred.

Table 6.7: Point estimate, standard error and confidence interval of the regression parameters for the hospitalization data (1: out of hospital, 2: in hospital, 3: dead)

Transition	Estimate	SE	CI
$1 \to 2$	-0.033	0.006	(-0.044, -0.021)
$1 \to 3$	0.036	0.049	(-0.060, 0.132)
$2 \to 1$	0.022	0.006	(0.010, 0.034)
$2 \to 3$	0.048	0.055	(-0.060, 0.156)

6.2.2.3 Analysis with Nonhomogeneous Semi-Markov Model

In the analysis with the modulated semi-Markov model, we have found evidence that the hospitalization process is not homogeneous in the study time. We modeled the hospitalization data with a NHSM process. We considered four different specifications: the fully nonparametric, the piecewise constant, the nonparametric multiplicative, and the semiparametric models. The piecewise constant approach used ten equally spaced partitions on the study time scale. The semiparametric model is just (6.2.1) considered in Section 6.2.2.2. We used the bandwidth $w = 5$ years in the study time scale. In the duration time scale, we used bandwidth $b_1 = 2$ years when the patient is out of hospital, and $b_2 = 4$ months when the patient is in hospital.

Figures 6.19 to 6.22 show the 3-dimensional plots of the estimated transition rate functions among different states, based on the four different estimation procedures. Note that all the four approaches give qualitatively the same pattern. As shown in Figure 6.19, the estimated transition rate function from "out of hospital" to "in hospital" as a function of duration τ increases until about $\tau = 1$ year and then decreases. It also has a decreasing trend in study time t. This indicates that this transition rate function depends on both time scales. Figures 6.23 and 6.24 show the estimated semi-Markov kernel by projecting it onto the two time scales. The decreasing trend of $Q_{12}(\tau; t)$ as a function of t with fixed τ indicates that less hospitalization resources are needed over time.

6.2.3 Sensitivity Analysis for Informative Censoring

We focused our interest in the two-state process alternating between "out of hospital" and "in hospital". We applied the method developed in Chapter 5 to conduct sensitivity analysis for checking about the possible informative censoring due to death. We worked with a subpopulation with 120 patients for illustration purpose. This subpopulation consists of 60 patients who are observed to death during the study, and other 60 patients who have similar cancer diagnosis dates.

Our estimation is based on Assumptions 5.A1 (5.A2) and 5.B, and Model (5.2.3). We estimated the semi-Markov kernel of the two-state process based on the assumed copulas from Frank's family

$$H(y_1, y_2; \alpha) = \log_\alpha \left\{ 1 + \frac{(\alpha^{y_1} - 1)(\alpha^{y_2} - 1)}{\alpha - 1} \right\}, \quad \alpha > 0,$$

with a plausible range of Kendall's τ in $[-0.8, 0.8]$. We also estimate the semi-Markov kernel by ignoring the possible dependent censoring due to death. The results are summarized in Figures 6.25 and 6.26, under HSM and NHSM model assumptions, respectively. Note that the estimated bounds of the semi-Markov kernel based on the plausible range of Kendall's τ are wider with the "out of hospital" to "in hospital" transition than with the "in hospital" to "out of hospital" transition. Thus the transition from "out of hospital" to "in hospital" is more sensitive to this type of dependent censoring than the transition from "in hospital" to "out of hospital".

6.2.4 Concluding Remarks

In this section, we analyzed the hospitalization data by the methods developed in this thesis. Under the HSM model, our estimates of the transition probabilities of the embedded Markov chain are quite different from the existing approaches. We find that the transition intensities of the hospitalization process vary in both the study and duration time scales. Thus the NHSM model appears more appropriate to the process. The four different specifications: the fully nonparametric, the piecewise constant, the nonparametric multiplicative, and the semiparametric models, lead to similar estimates of the transition intensities.

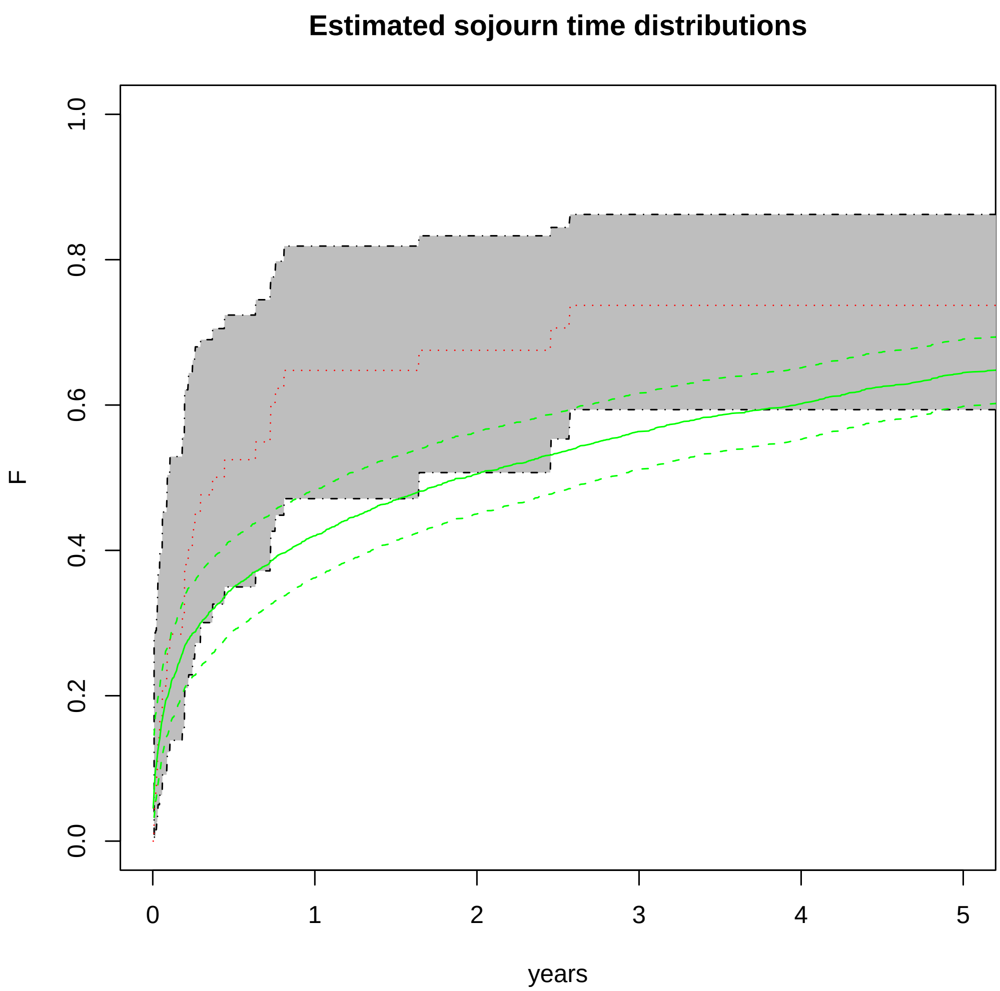

Figure 6.16: Estimated sojourn time distributions for hospitalization data based on normalized estimate (Solid: point estimate for F_{12}; Dashed: 95% confidence bands for F_{12}; Dotted: point estimate for F_{13}; Shaded: 95% confidence bands for F_{13})

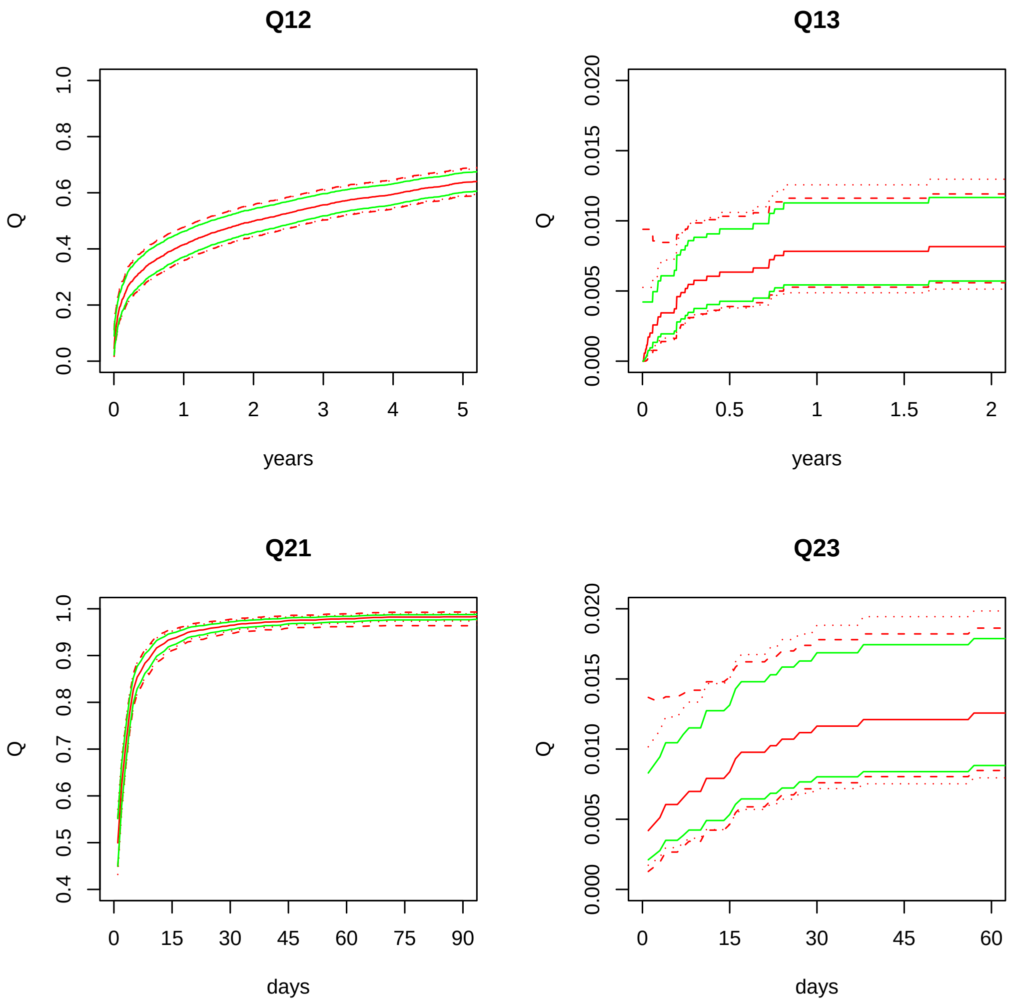

Figure 6.17: Estimated semi-Markov kernel for hospitalization data (Middle solid: point estimate; Outer solid: 95% pointwise confidence intervals; Dotted: 95% transformed EP band; Dashed: 95% transformed HW band)

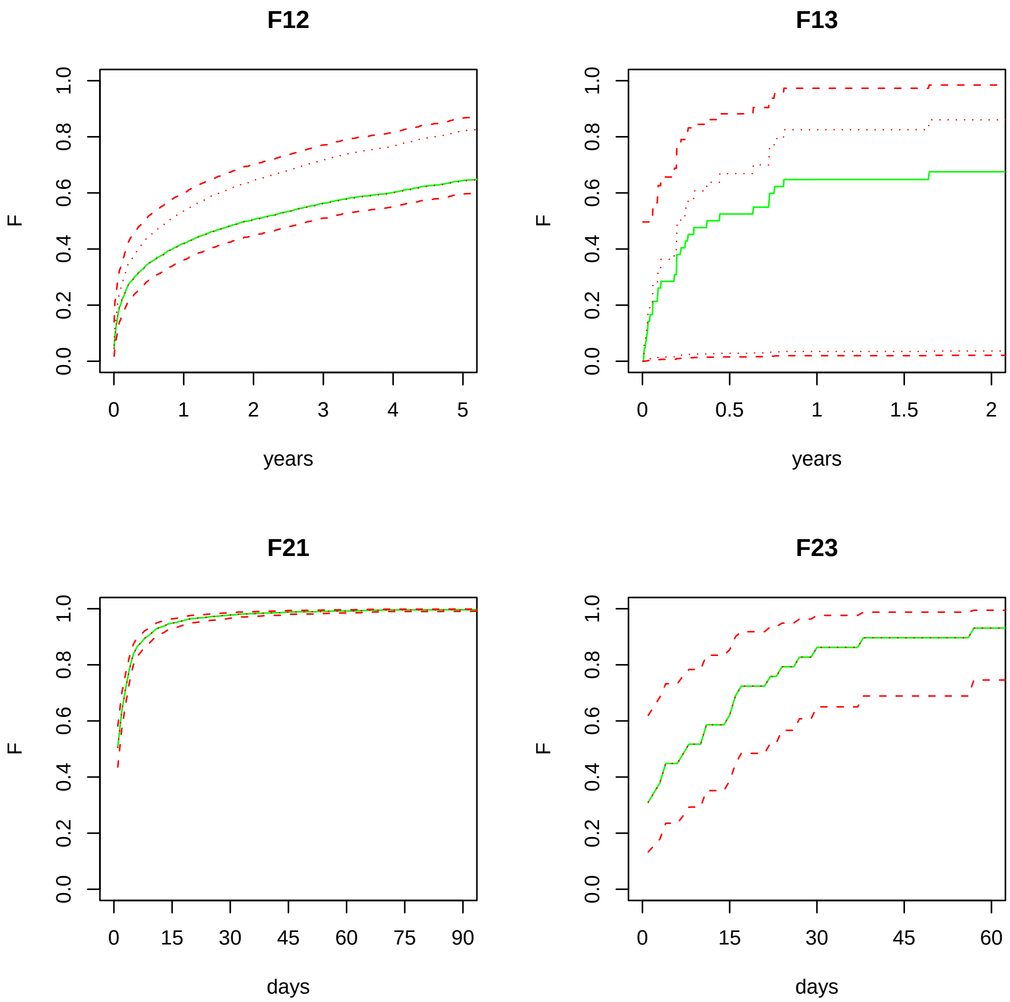

Figure 6.18: Estimated distribution of sojourn times for hospitalization data (Solid: normalized estimate; Dotted: estimated bound; Dashed: 95% transformed EP band)

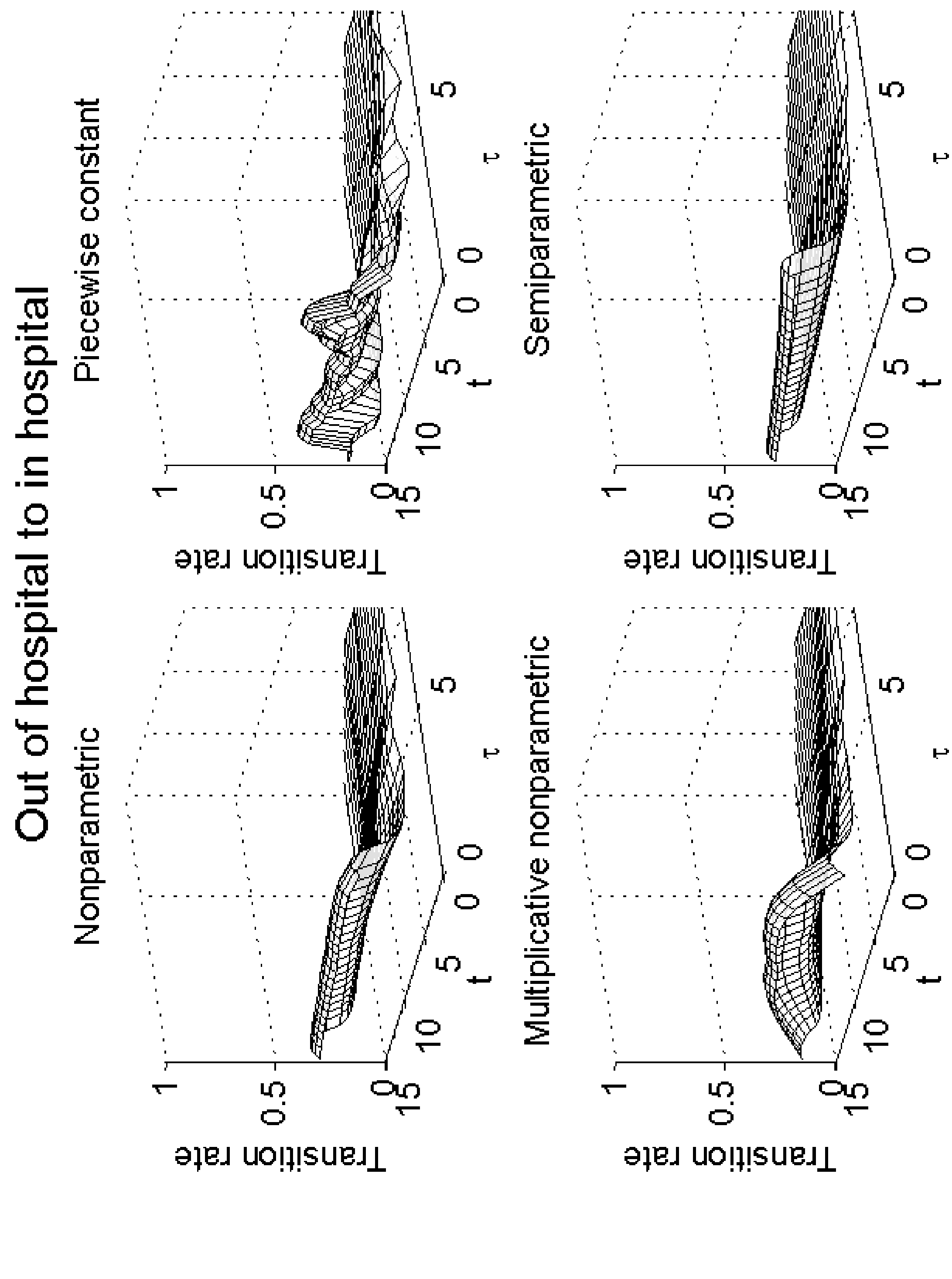

Figure 6.19: Estimated transition rate functions from "out of hospital" to "in hospital"

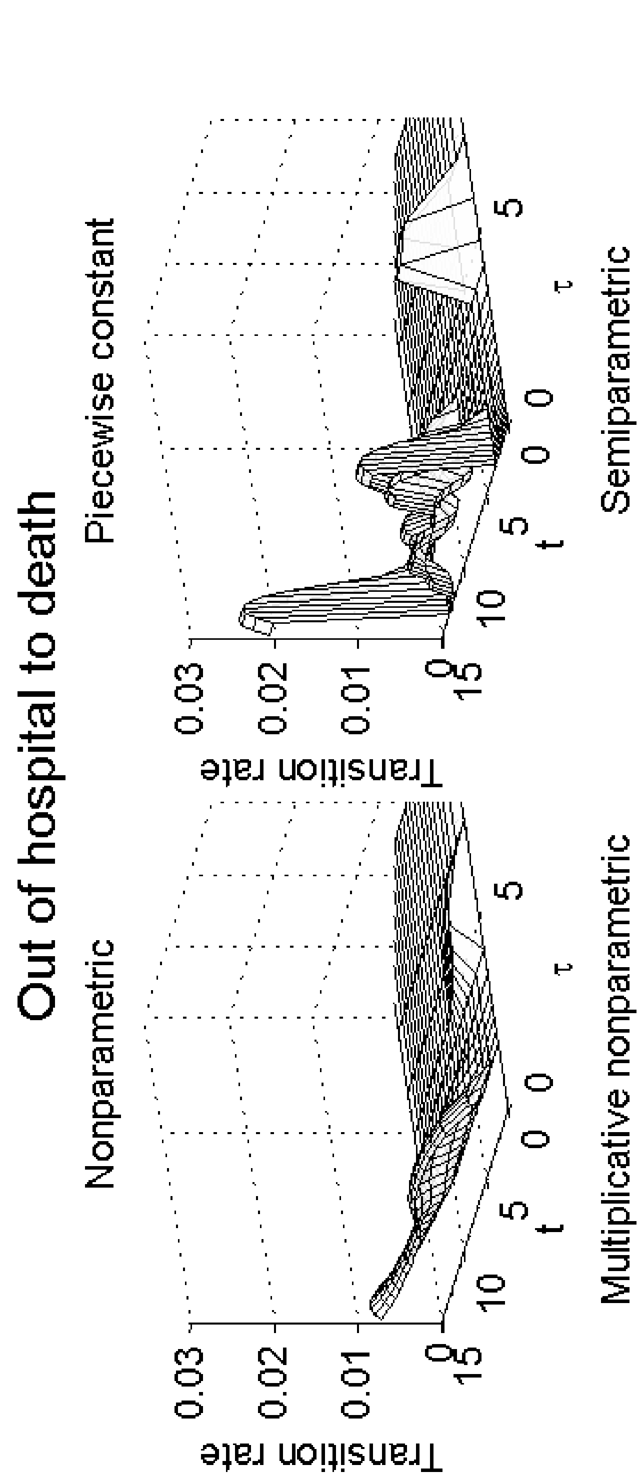

Figure 6.20: Estimated transition rate functions from “out of hospital” to “death”

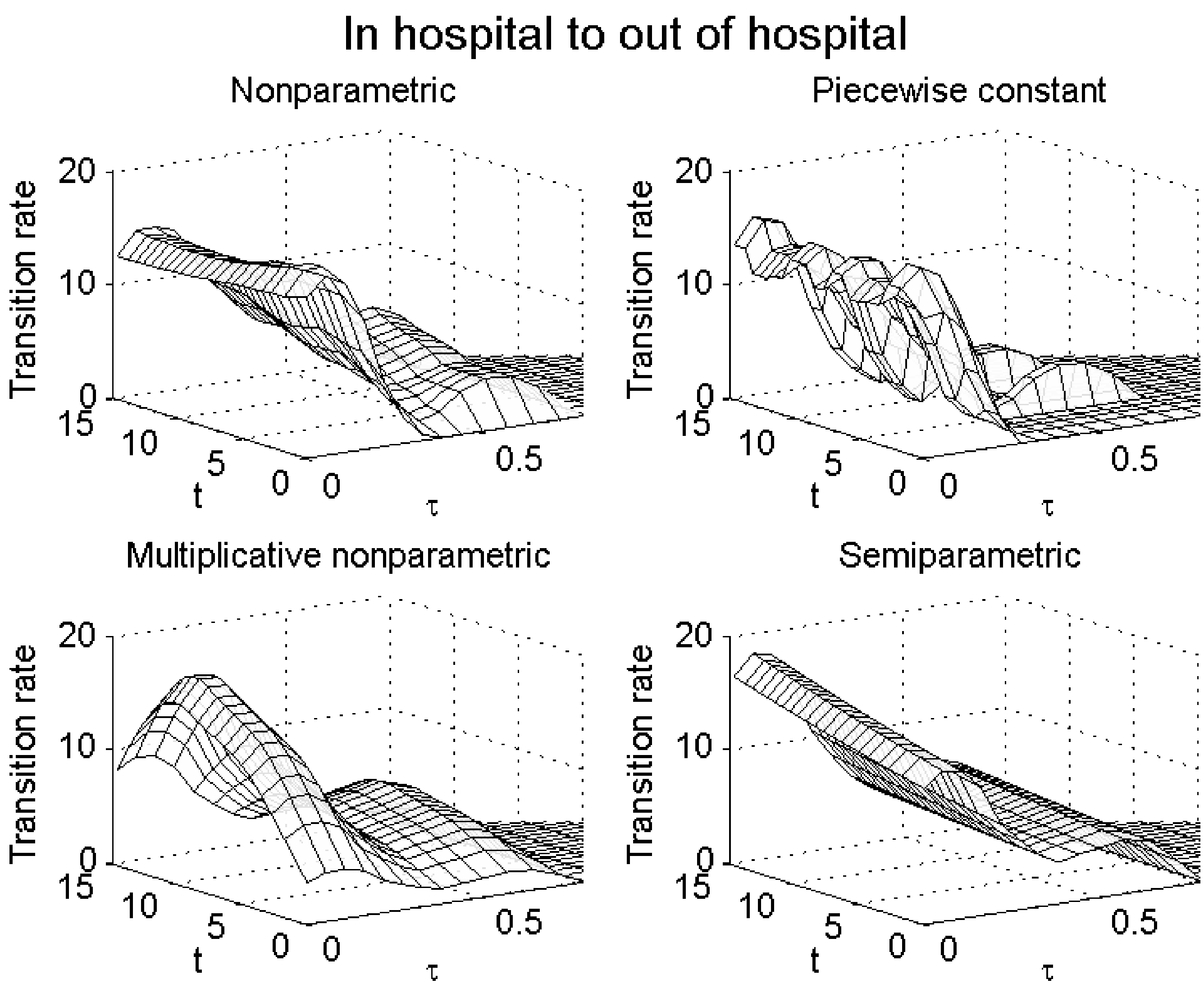

Figure 6.21: Estimated transition rate functions from “in hospital” to “out of hospital”

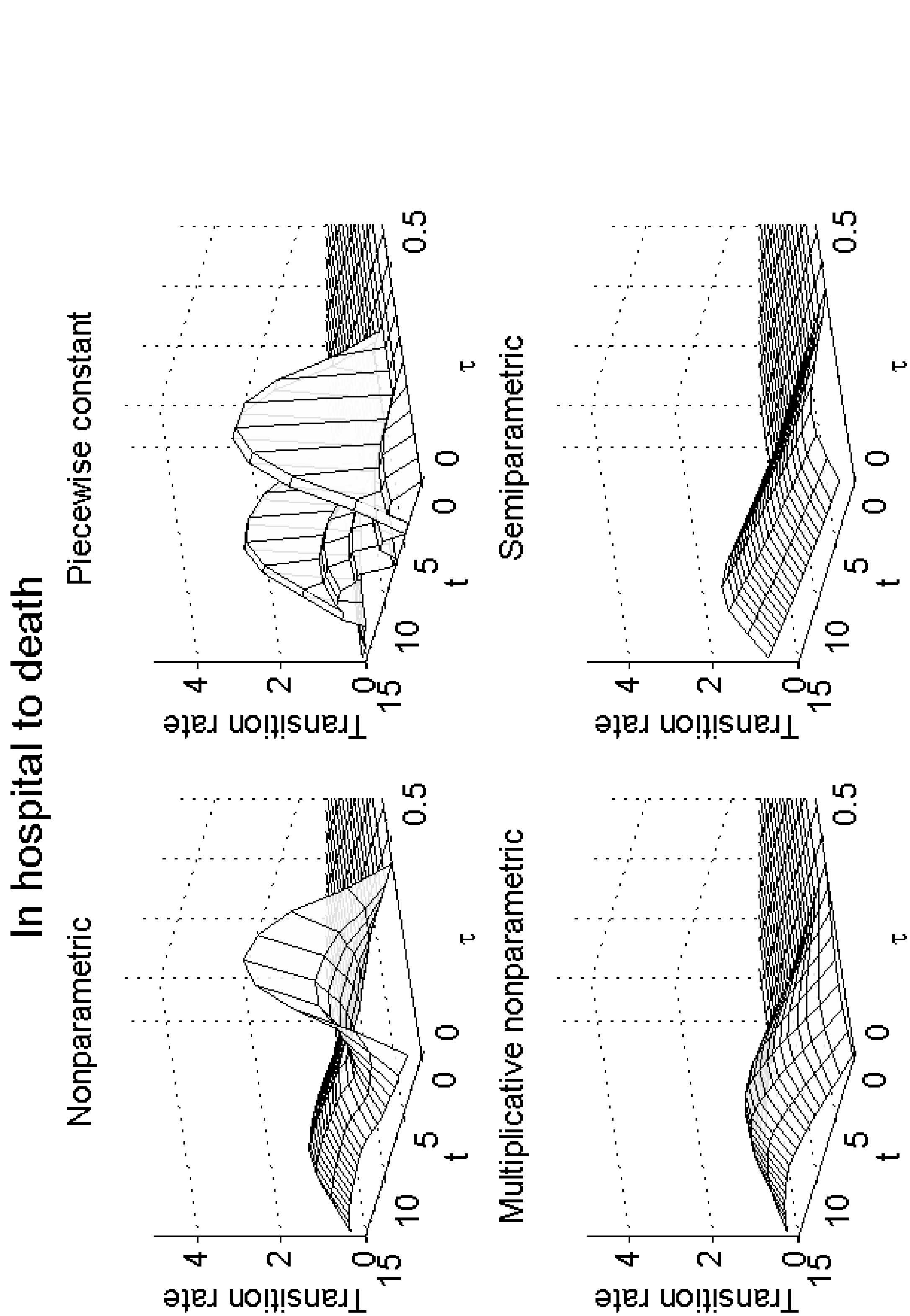

Figure 6.22: Estimated transition rate functions from “in hospital” to “death”

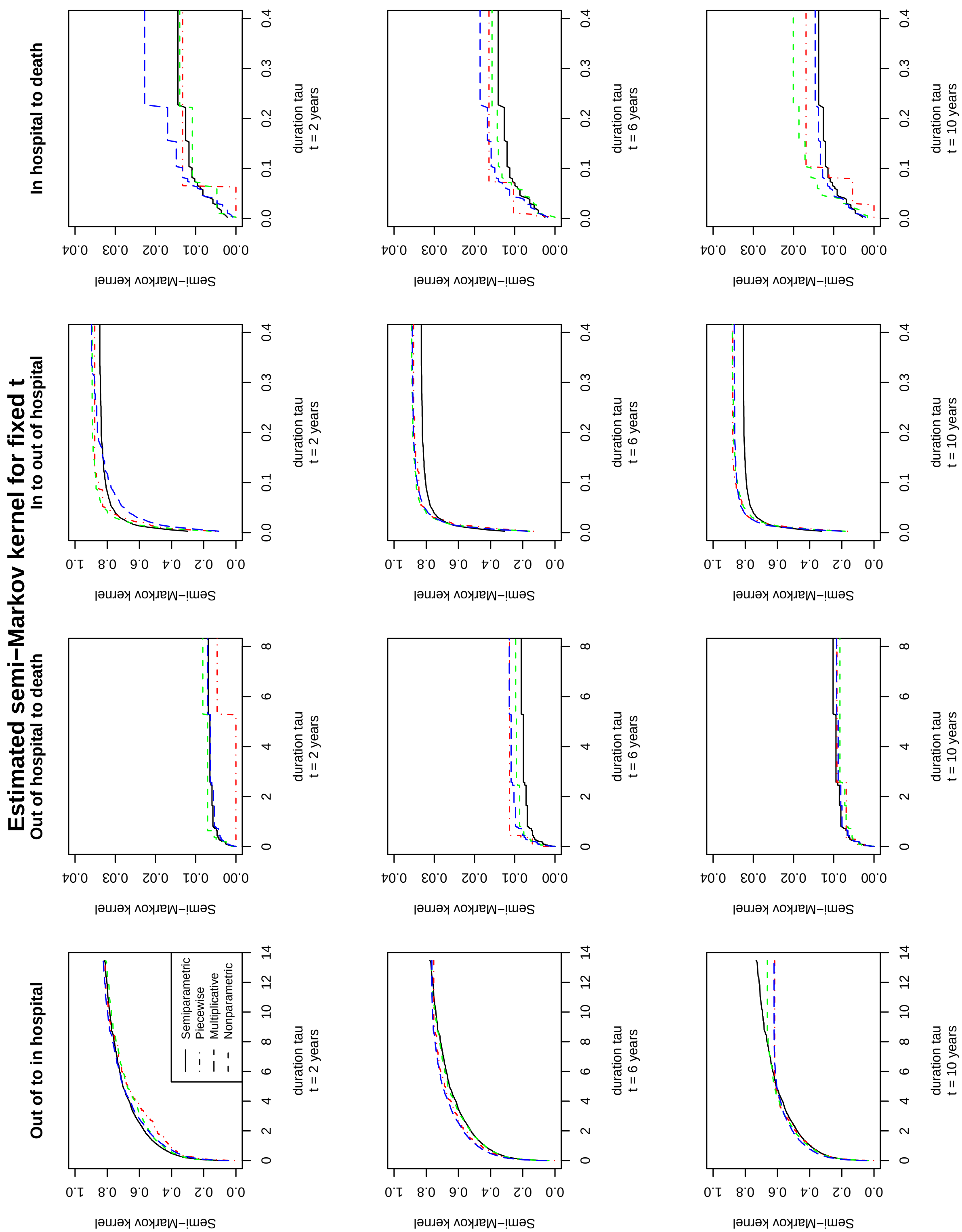

Figure 6.23: Estimated semi-Markov kernel for fixed t

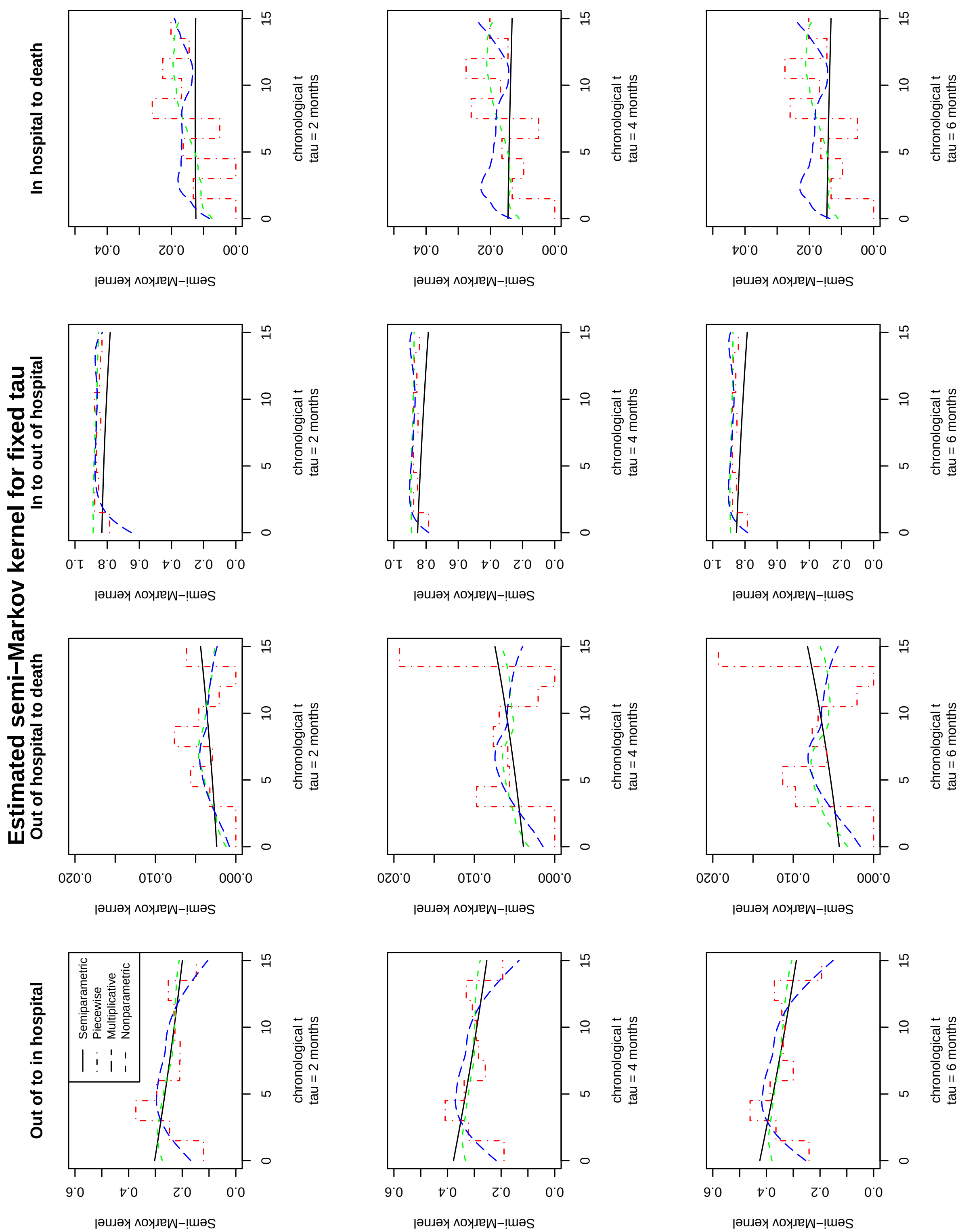

Figure 6.24: Estimated semi-Markov kernel for fixed τ

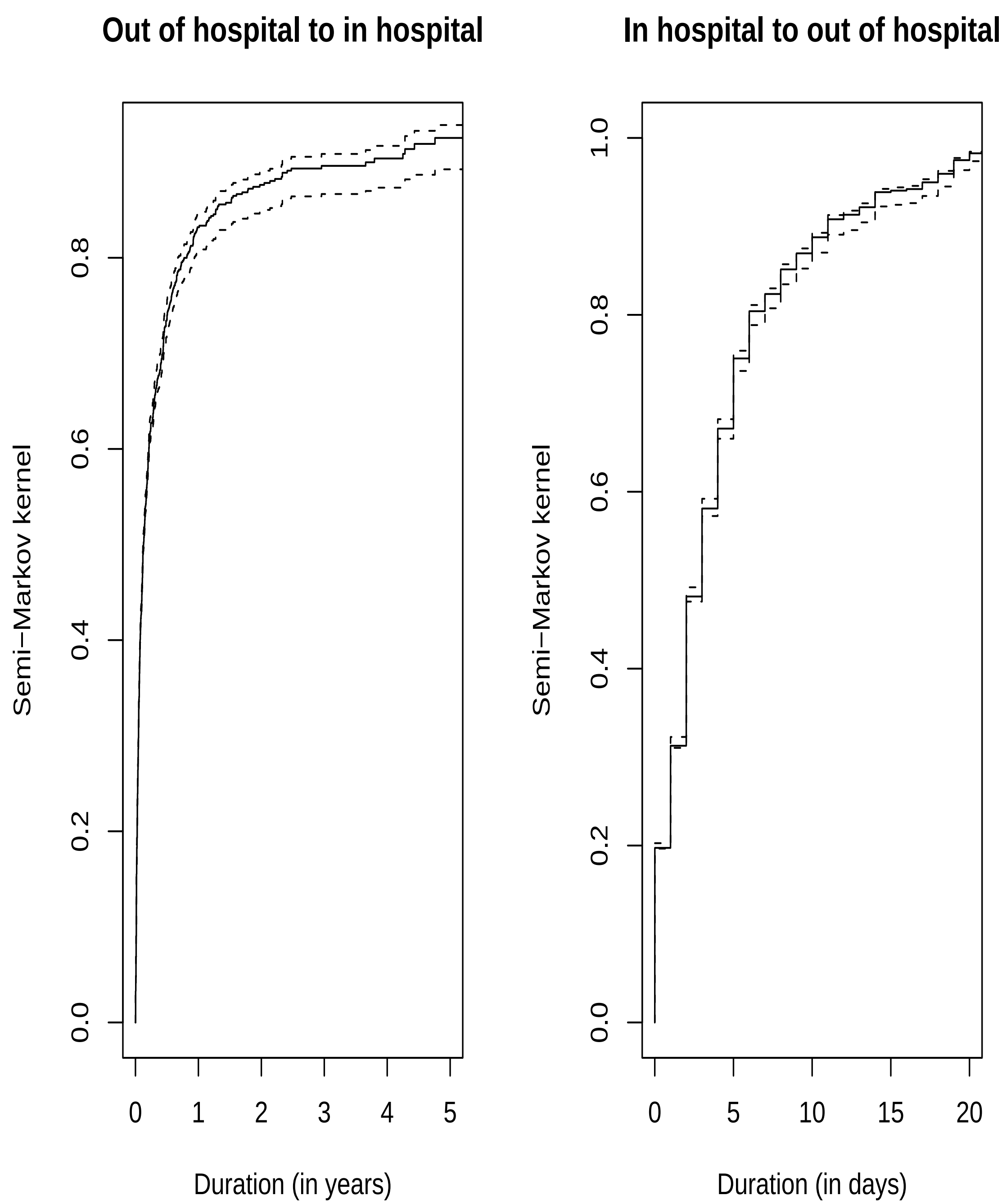

Figure 6.25: Estimated semi-Markov kernel of the two-state HSM process (Solid: estimates ignoring the possible dependent censoring; Dashed: plausible bounds based on the assumed copulas from Frank's family)

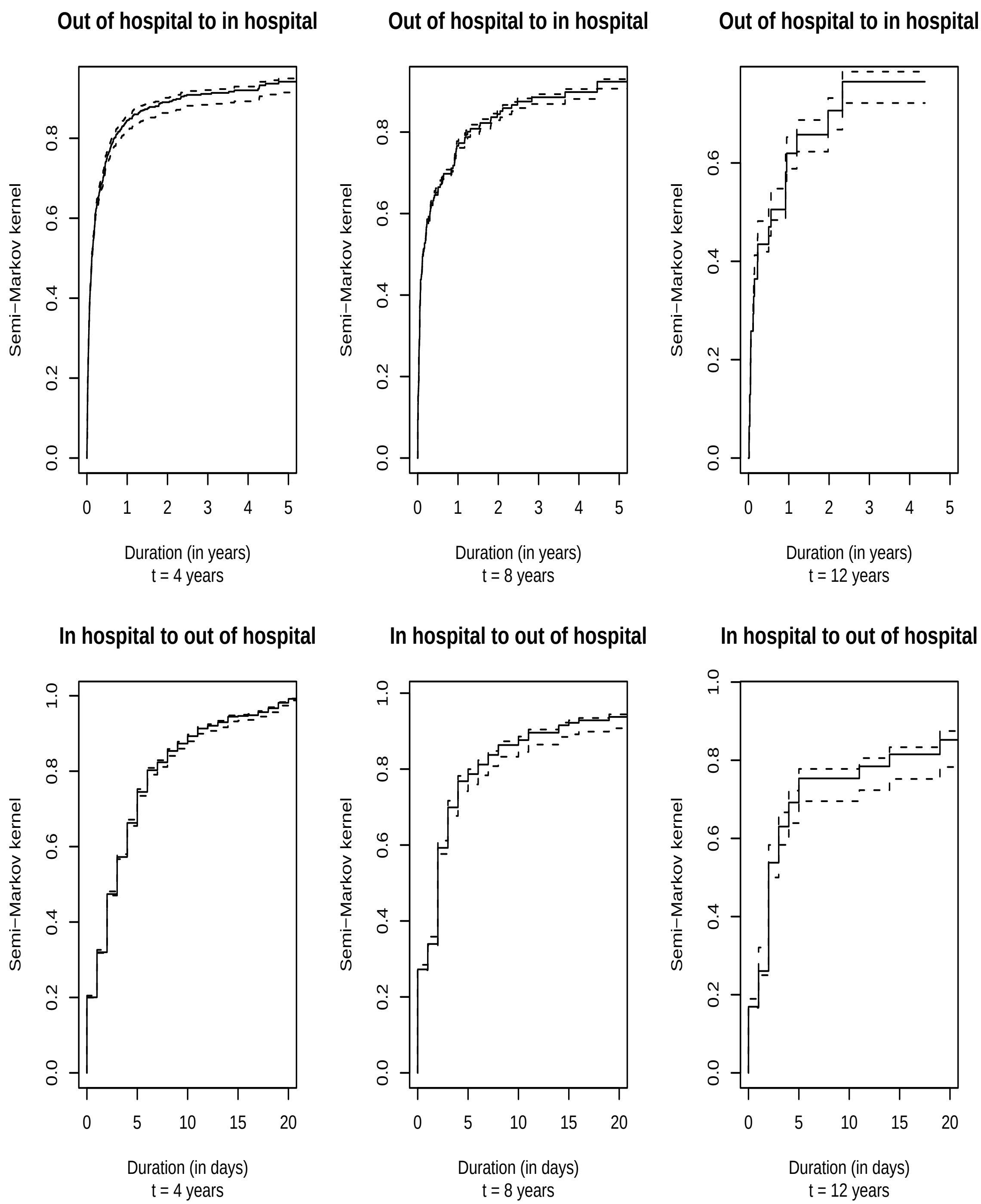

Figure 6.26: Estimated semi-Markov kernel of the two-state NHSM process (Solid: estimates ignoring the possible dependent censoring; Dashed: plausible bounds based on the assumed copulas from Frank's family)

Chapter 7

Discussion

7.1 Summary

Markov models have been widely used for multi-state processes because of (i) the simplicity of model specification and interpretation, (ii) the availability of counting process martingale theory to derive asymptotics, and (iii) the product integration to link the transition intensities with transition probabilities. However, due to their memoryless property, strict Markov models can not deal with duration dependence, and have been found inadequate in many practical applications. In this thesis, we have developed statistical methods for multi-state processes with duration-dependent transition intensities.

We start from the homogeneous semi-Markov (HSM) process, as a generalization of the classical homogeneous Markov processes, which assumes that the transition intensities depend on the history only through the current state and the duration time. Gill (1980) derives the consistency and weak convergence of the estimator of the semi-Markov kernel proposed by Lagakos et al. (1978). However, the asymptotic Gaussian process does not have an independent increment structure, thus it can not be transformed into the standard Brownian bridge or Brownian motion to construct confidence bands for the semi-Markov kernel. We propose two simulation based algorithms for this purpose. In addition, we show that the existing estimators for the transition probabilities of the embedded Markov chain and the sojourn time distributions can be biased when right censoring is involved. A robust estimation procedure is proposed to address the concern. Simulation studies show that the proposed methods perform well with finite sample size. The efficiency of the robust approach can be comparable with the existing approaches.

We then study the modulated semi-Markov (MSM) process, which extends the HSM process to a Cox regression setting. The dependence of the baseline transition intensities on the duration time scale makes the model fall outside the framework of Aalen's multiplicative intensity models and invalidates the usual martingale methods. Dabrowska *et al.* (1994) consider MSM processes with covariates depend on the duration time in the present state only, which precludes the study time variable. As a generalization, Chapter 3 allows general time-dependent covariates and proposes estimating equations for the regression parameters. Using empirical process theory, we establish the consistency, asymptotic normality and efficiency of the estimators for the regression parameter. The large sample approximation of the limiting distribution is adequate with sample size as small as 50, as shown in the simulation.

As a further generalization, the nonhomogeneous semi-Markov (NHSM) process assumes its transition intensity can involve both the study and the duration time scales. We consider statistical inferences for the NHSM processes with four different model specifications. The first one is a MSM model, in which the the study time is included as a time-dependent covariate. The second model assumes that the transition intensities are piecewise constant in the study time scale, and can vary arbitrarily in the duration time scale. The third model is structured nonparametric, where the transition intensities depend on the two time scales in a multiplicative form. The last model is fully nonparametric in that the transition intensities can depend freely on the two time scales. Simulation studies show that the more structured models are more efficient if they are correctly specified. However, they can lead to biased inferences when the model assumptions are violated. The four models are nested, which can be utilized to select the most parsimonious model.

Dependent censoring problem is challenging in event history data analysis. We consider a particular type of informative right censoring scheme with the observation of a NHSM process. Motivated by the competing risks formulation of the HSM processes, we model the informative censoring mechanism as another competing risk. Under this model assumption, the censored process becomes a new NHSM process with the censoring included as a new absorbing state of the original process. We then adapt a copula based approach for dependent competing risks to the setting. An advantage of the copula approach is that the marginal distributions need not to be specified, and can be estimated nonparametrically.

We applied the proposed methods to the two real data sets described in Chapter 1. The analysis outcomes suggest that the transition intensities of both the human sleep and the hospitalization processes vary in both the study and the duration time scales. Thus the NHSM model is a plausible model for the two data sets.

7.2 Further Investigations

In what follows, we outline some directions for future research.

7.2.1 Interval Transition Probabilities

An easily interpretable quantity of a multi-state process, which is often of interest in practice, is the interval transition probability

$$P_{hj}(s,t;\mathcal{F}_s) = P\{\mathcal{S}(t) = j|\mathcal{S}(s) = h, \mathcal{F}_{s-}\}. \quad (7.2.1)$$

With nonhomogeneous Markov processes, (7.2.1) reduces to

$$P_{hj}(s,t) = P\{\mathcal{S}(t) = j|\mathcal{S}(s) = h\},$$

which is linked with the transition intensities through product integration. The inference procedures have been well studied (Aalen and Johansen, 1978; Andersen et al., 1993). Dabrowska *et al.* (1994) and Dabrowska (1995) consider estimation procedures for the interval transition probabilities with HSM processes and modulated semi-Markov processes with time-independent covariates.

It can be of practical interest to estimate the interval transition probabilities with NHSM processes. Based on the stochastic feature of NHSM processes, it is equivalent to estimate the following interval transition probabilities:

$$P_{hj}(t;\tau_0,\tau) = P\left[\mathcal{S}(t+\tau) = j|\mathcal{S}(t+\tau_0) = h, B(t+\tau_0) = \tau_0\right], \quad (7.2.2)$$

which is the probability of the process being in state j at time $t+\tau$, given the process enters state h at time t and remains in state h by time $t+\tau_0$. Let $P_{hj}(t;\tau) = P_{hj}(t;0+,\tau)$ be the probability that the process in state j at time $t+\tau$ given it enters state h at time t.

Applying a conditional argument, we can show that $P_{hj}(t;\tau)$'s satisfy the following system of Volterra integral equations:

$$P_{hj}(t;\tau) = \delta_{hj}S_h(\tau;t) + \sum_{k=1}^{r}\int_0^{\tau} q_{hk}(u;t)P_{kj}(t+u;\tau-u)du \quad (7.2.3)$$

for all h and j, where $\delta_{hj} = 1$ if $h = j$ and 0 otherwise, and $S_h(\tau; t) = 1 - H_h(\tau; t)$. These equations are the direct counterpart of Kolmogorov equations for the Markov processes. Let

$$Q_{hj}(\tau_0, \tau; t) = \frac{Q_{hj}(\tau; t) - Q_{hj}(\tau_0; t)}{1 - Q_{hj}(\tau_0; t)},$$

and $q_{hj}(\tau_0, \tau; t)$ be the partial derivative of $Q_{hj}(\tau_0, \tau; t)$ with respect to τ. Then

$$P_{hj}(t; \tau_0, \tau) = \delta_{hj} S_h(\tau_0, \tau; t) + \sum_{k=1}^{r} \int_{\tau_0}^{\tau} q_{hk}(\tau_0, u; t) P_{kj}(t + u; \tau - u) du \qquad (7.2.4)$$

where $S_h(\tau_0, \tau; t) = 1 - \sum_k Q_{hk}(\tau_0, \tau; t)$.

Lucas et al. (2006) approximate numerically the solution $\hat{P}_{hj}(t; \tau)$ of (7.2.3) by means of a finite system of algebra equations. Here we consider an alternative method based on Monte Carlo simulation to estimate $P_{hj}(t; \tau_0, \tau)$. We first apply the methods in Chapter 4 to obtain estimates $\hat{\alpha}_{hj}(\tau; t)$ and $\hat{Q}_{hj}(\tau; t)$ of $\alpha_{hj}(\tau; t)$ and $Q_{hj}(\tau; t)$, respectively, for $h, j \in \mathcal{E}$ and $t, \tau \geq 0$. Denote $\sum_{j \neq h} \hat{Q}_{hj}(\tau; t)$ by $\hat{H}_h(\tau; t)$. Then an algorithm to estimate $P_{hj}(t; \tau_0, \tau)$ is as follows.

Step 1. Let $(J_0, T_0) = (h, t)$ and $m = 0$.

Step 2. Generate X_{m+1} from $\hat{H}_{J_m}(\cdot; T_m)$, and then generate J_{m+1} from state j with probability $\hat{\alpha}_{J_m j}(X_{m+1}; T_m) / \sum_j \hat{\alpha}_{J_m j}(X_{m+1}; T_m)$. Record (J_{m+1}, T_{m+1}), where $T_{m+1} = T_m + X_{m+1}$.

Step 3. Repeat Step 2 for each $m = 1, 2, \ldots$, until either J_{m+1} is an absorbing state or $T_{m+1} > t + \tau$.

Step 4. Repeat Step 1 to Step 3 to obtain M sample paths.

We can then estimate $P_{hj}(t; \tau_0, \tau)$ by the proportion of sample paths with $X_1 > \tau_0$ and being in state j at time $t + \tau$, denoted by $\hat{P}_{hj}(t; \tau_0, \tau)$. The standard deviation of $\hat{P}_{hj}(t; \tau_0, \tau)$ can be obtained by the bootstrap approach.

7.2.2 Goodness-of-fit Tests

We have considered several nested models for NHSM processes: semiparametric, structured nonparametric, and fully nonparametric. They can be applied to conduct goodness-of-fit tests to select the most parsimonious model. In the context of a nonparametric hazard model for survival data, McKeague and Utikal (1991) consider some goodness-of-fit tests based on the differences between estimates of the doubly cumulative hazard function. The idea can be used here by comparing the cumulative transition rate function

$$A_{hj}(\tau;t) = \int_0^\tau \alpha_{hj}(u;t)du,$$

or the doubly cumulative transition rate function

$$\mathcal{A}_{hj}(\tau;t) = \int_0^t \int_0^\tau \alpha_{hj}(u;s)dsdu.$$

For instance, to test whether the multiplicative nonparametric model fit the data adequately, we may plot the differences of the estimated cumulative transition rate function under the multiplicative nonparametric model and the fully nonparametric model. If the multiplicative nonparametric model is appropriate, the difference should be fluctuate around zero without any pattern.

We can also conduct formal hypothesis tests based on the standardized difference of the estimated cumulative transition rate functions under the nested models, which can be shown to converge to a Gaussian random field. The limiting process with possibly complex covariance structure may be approximated via bootstrap. The hypothesis tests can be based on Kolmogorov-Smirnov type or Cramér-von Mises type statistics.

7.2.3 General Modulated Semi-Markov Models

In Chapter 3, we considered estimation procedures based on the modulated semi-Markov models where the transition intensities are assumed to have the form

$$\alpha_{hj}(t|\mathcal{F}_t, Z_{hj}(t)) = \alpha_{0hj}\left(B(t)\right) e^{\theta' Z_{hj}(t)},$$

or

$$\alpha_{hj}(t|\mathcal{F}_t, Z_{hj}(t)) = \alpha_{0hj}\left(B(t), \tilde{N}(t-)\right) e^{\theta' Z_{hj}(t)}.$$

The possible dependence of the transition intensities on the study time is modeled parametrically.

More flexible models, where the baseline transition rate function can depend on both the study and duration time scales freely, are

$$\alpha_{hj}(t|\mathcal{F}_t, Z_{hj}(t)) = \alpha_{0hj}\left(B(t);t\right) e^{\theta' Z_{hj}(t)},$$

and

$$\alpha_{hj}(t|\mathcal{F}_t, Z_{hj}(t)) = \alpha_{0hj}\left(B(t); t, \tilde{N}(t-)\right) e^{\theta' Z_{hj}(t)}.$$

With a fixed θ, we can apply the methodology in Chapter 4 to obtain an estimate of the baseline transition rate function. We can then consider some profile estimator for the regression parameter θ.

7.2.4 Alternative Observation Schemes

In this thesis, we have focused on the right censoring observation scheme. In practice, the data may be collected periodically, leading to panel data. Kalbfleisch and Lawless (1985) fit Markov models with panel data. The difficulty to fit semi-Markov models with panel data, as shown in Kang and Lagakos (2007), is that the likelihood function is very complicated to work with. Kang and Lagakos (2007) consider a simplified situation where the transition intensity from at least one of the states of the underlying process is time homogeneous, in which case they show that the likelihood function is tractable.

Another problem is the possible unknown duration time in the initial state, when the subjects are already in certain state before they enter the study. In this thesis, we have assumed that the subjects start a new state when enter the study, or the duration time in the current state prior to the entrance into the study is known. Otherwise, the methods need to be modified to account for the duration of time in the initial state prior to the entrance of the study. Methods similar to those in the work by Satten and Sternberg (1999) and Cai et al. (2008) might be adopted to address the issue.

7.2.5 Other Further Investigations

Listed below are other research topics closely related to the thesis project.

7.2.5.1 Robustness of the Semiparametric Approach

As shown in the literature, approaches associated with Cox regression may have certain robustness properties against model misspecification in Markov models. We will examine the performances of the approaches in Chapter 3 when the model is misspecified with the semi-Markov model.

7.2.5.2 Nonparametric Additive Model

In Chapter 4, we consider a particular nonparametric structured model, the nonparametric multiplicative model, with NHSM processes. Another nonparametric structured model for future research is the nonparametric additive model, in which the transition intensities depend on the study time and the duration time additively, and the functional forms are left unspecified. The iterative algorithm developed in Section 4.4 can be adapted. Theoretical justification for convergence of the algorithms warrants further investigation.

7.2.5.3 Bandwidth Selection

Kernel smoothing methods are used in the estimation procedures with NHSM processes. It is not clear, however, how to choose the optimal bandwidth. The asymptotic distribution of the estimators may be too complicated to be used in selecting the bandwidth by the plug-in method. Bandwidth selection has been well studied in the context of hazard rate estimation with survival data. Patil (1993) proposes the least squares cross-validation method. González-Manteiga *et al.* (1996) introduce a bootstrap approach. These methods can be potentially adapted to the setting.

7.2.5.4 Estimation of Marginal Quantities

This thesis focuses on intensity-based models with the counting process formulation of multi-state processes. For the special case of recurrent events, robust inference procedures based on marginal rate functions have been proposed (Lin *et al.*, 2000; Cook and Lawless, 2007). Datta and Satten (2001, 2002) consider estimation of the state occupation probabilities with multi-state processes based on marginal transition rate functions. More generally, we will study estimation of marginal transition probabilities. Meira-Machado *et al.* (2006) propose estimators with a special illness-death process without recovery.

Bibliography

Akritas, M. G. (**1986**), "Bootstrapping the Kaplan-Meier Estimator," Journal of the American Statistical Association **81**, 1032–1038.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (**1993**), *Statistical models based on counting processes* (New York: Springer-Verlag).

Andersen, P. K., Esbjerg, S., and Sorensen, T. A. (**2000**), "Multi-state models for bleeding episodes and mortality in liver cirrhosis," Statistics in Medicine **19**, 587–599.

Andersen, P. K. and Gill, R. D. (**1982**), "Cox's Regression Model for Counting Processes: A Large Sample Study," The Annals of Statistics **10**, 1100–1120.

Andersen, P. K. and Keiding, N. (**2002**), "Multi-state models for event history analysis," Statistical Methods in Medical Research **11**, 91–115.

Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A. (**1983**), "Information and asymptotic efficiency in parametric-nonparametric models," The Annals of Statistics **11**, 432–452.

Beran, R. (**1981**), "Nonparametric regression with randomly censored survival data," Tech. rep., University of California, Berkeley.

Bilias, Y., Gu, M., and Ying, Z. (**1997**), "Towards a general asymptotic theory for Cox model with staggered entry," The Annals of Statistics **25**, 662–682.

Borgan, Ø. and Liestøl, K. (**1990**), "A Note on Confidence Intervals and Bands for the Survival Function Based on Transformations," Scandinavian Journal of Statistics **17**, 35–41.

Bowe, T. R. and Anders, T. F. (**1979**), "The Use of the Semi-Markov Model in the Study of the Development of Sleep-Wake States in Infants," Psychophysiology **16**, 41–48.

Braekers, R. and Veraverbeke, N. (**2005**), "A copula-graphic estimator for the conditional survival function under dependent censoring," The Canadian Journal of Statistics **33**, 429–447.

Brémaud, P. (**1981**), *Point processes and queues, martingale dynamics* (New York: Springer-Verlag).

Cai, L., Schenker, N., Lubitz, J., Diehr, P., Arnold, A., and Fried, L. P. (**2008**), "Evaluation of a method for fitting a semi-Markov process model in the presence of left-censored spells using the Cardiovascular Health Study," Statistics in Medicine **27**, 5509–5524.

Carrière, J. F. (**1995**), "Removing cancer when it is correlated with other causes of death," Biometrical Journal **37**, 339–350.

Chang, I. S. and Hsiung, C. A. (**1994**), "Information and asymptotic efficiency in some generalized proportional hazards models for counting processes," The Annals of Statistics **22**, 1275–1298.

Chen, P. and Tien, H. (**2004**), "Semi-Markov Models for Multistate Data Analysis with Periodic Observations," Communications in Statistics - Theory and Methods **33**, 475–486.

Clayton, D. G. (**1978**), "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence," Biometrika **65**, 141–151.

Clayton, D. G. (**1988**), "The analysis of event history data: a review of progress and outstanding problems," Statistics in Medicine **7**, 819–841.

Commenges, D. (**1999**), "Multi-state Models in Epidemiology," Lifetime Data Analysis **5**, 315–327.

Cook, R. J. and Lawless, J. F. (**2007**), *The statistical analysis of recurrent events* (New York: Springer).

Cox, D. R. (**1973**), "The statistical analysis of dependencies in point processes," in *Symposium on Point Processes*, edited by P. A. W. Lewis (New York: Wiley), pp. 55–66.

Dabrowska, D. M. (**1995**), "Estimation of transition probabilities and bootstrap in a semiparametric markov renewal model," Journal of Nonparametric Statistics **5**, 237–259.

Dabrowska, D. M. and Ho, W. T. (**2006**), "Estimation in a semiparametric modulated renewal process," Statistica Sinica **16**, 93–119.

Dabrowska, D. M. and Lee, W. (**1996**), "Nonparametric estimation of transition probabilities in a two-stage duration model," Journal of Nonparametric Statistics **7**, 75–103.

Dabrowska, D. M. and Lee, W. (**1997**), "Smoothed Cox regression," The Annals of Statistics **25**, 1510–1540.

Dabrowska, D. M., Sun, G., and Horowitz, M. M. (**1994**), "Cox Regression in a Markov Renewal Model: An Application to the Analysis of Bone Marrow Transplant Data," Journal of the American Statistical Association **89**, 867–877.

Datta, S. and Satten, G. A. (**2001**), "Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models," Statistics & probability letters **55**, 403–411.

Datta, S. and Satten, G. A. (**2002**), "Estimation of Integrated Transition Hazards and Stage Occupation Probabilities for Non-Markov Systems Under Dependent Censoring," Biometrics **58**, 792–802.

Dinse, G. E. and Larson, M. G. (**1986**), "A note on semi-Markov models for partially censored data," Biometrika **73**, 379–386.

Fermanian, J. D. (**2003**), "Nonparametric estimation of competing risks models with covariates," Journal of Multivariate Analysis **85**, 156–191.

Fiocco, M., Putter, H., and Houwelingen, H. C. V. (**2008**), "Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models," Statistics in Medicine **27**, 4340–4358.

Fisher, L. and Kanareck, P. (**1974**), "Presenting censored survival data when censoring and survival times may not be independent," in *Reliability and Biometry*, edited by F. Proschan and R. J. Serfling (Philadelphia, PA: SIAM), pp. 303–326.

Frank, M. J. (**1979**), "On the simultaneous association of $F(x, y)$ and $x+y-F(x, y)$," Aequationes Mathematics **21**, 37–38.

Gill, R. D. (**1980**), "Nonparametric Estimation Based on Censored Observations of a Markov Renewal Process," Z. Wahrscheinlichkeitstheorie verw. Gebiete **53**, 97–116.

Glidden, D. V. (**2002**), "Robust Inference for Event Probabilities with Non-Markov Event Data," Biometrics **58**, 361–368.

González-Manteiga, W., Cao, R., and Marron, J. S. (**1996**), "Bootstrap Selection of the Smoothing Parameter in Nonparametric Hazard Rate Estimation," Journal of the American Statistical Association **91**, 1130–1140.

Gumbel, E. J. (**1961**), "Distributions des valeurs extrêmes en plusiers dimensions," Publications Institute Statistical University Paris **9**, 171–173.

Hall, W. J. and Wellner, J. A. (**1980**), "Confidence bands for a survival curve from censored data," Biometrika **67**, 133–143.

Härdle, W., Müler, M., Sperlich, S., and Werwatz, A. (**2004**), *Nonparametric and Semiparametric Models* (Berlin: Springer-Verlag).

Heckman, J. J. and Honore, B. E. (**1989**), "The identifiability of the competing risks model," Biometrika **76**, 325–330.

Hougaard, P. (**1986**), "A class of multivariate failure time distributions," Biometrika **73**, 671–678.

Hougaard, P. (**1999**), "Multi-state Models: A Review," Lifetime Data Analysis **5**, 239–264.

Hu, X. J., Lorenzi, M., Spinelli, J., Ying, S. C., and McBride, M. L. (**2008**), "Analysis of Recurrent Events with Non-negligible Event Duration, with Application to Evaluating Hospital Utilization," Submitted for publication.

Huang, X. and Zhang, N. (**2008**), "Regression Survival Analysis with an Assumed Copula for Dependent Censoring: A Sensitivity Analysis Approach," Biometrics **64**, 1090–1099.

Iosifescu Manu, A. (**1972**), "Non-homogeneous Semi-Markov Processes," Studii si Cercetuari Matematice **24**, 529–533.

Jacod, J. (**1975**), "Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales," Z. Wahrscheinlichkeitstheorie verw. Gebiete **31**, 235–253.

Janssen, J. and De Dominicis, R. (**1984**), "Finite non-homogeneous semi-Markov processes: theoretical and computational aspects," Insurance: Mathematics and Economics **3**, 157–165.

Kalbfleisch, J. D. and Lawless, J. F. (**1985**), "The Analysis of Panel Data Under a Markov Assumption," Journal of the American Statistical Association **80**, 863–871.

Kang, M. and Lagakos, S. W. (**2007**), "Statistical methods for panel data from a semi-Markov process, with application to HPV," Biostatistics **8**, 252–264.

Kaplan, E. L. and Meier, P. (**1958**), "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association **53**, 457–481.

Keiding, N. and Andersen, P. K. (**1989**), "Nonparametric Estimation of Transition Intensities and Transition Probabilities: A Case Study of a Two-State Markov Process," Applied Statistics **38**, 319–329.

Klein, J. P. and Moeschberger, M. L. (**1988**), "Bounds on Net Survival Probabilities for Dependent Competing Risks," Biometrics **44**, 529–538.

Kneib, T. and Hennerfeind, A. (**2008**), "Bayesian semiparametric multi-state models," Statistical Modelling **8**, 169–198.

Lagakos, S. W., Sommer, C., and Zelen, M. (**1978**), "Semi-Markov models for partially censored data," Biometrika **65**, 311–317.

Lin, D. Y. (**1997**), "Non-parametric inference for cumulative incidence functions in competing risks studies," Statistics in Medicine **16**, 901–910.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (**2000**), "Semiparametric regression for the mean and rate functions of recurrent events," Journal of the Royal Statistical Society Series B **62**, 711–730.

Lin, D. Y., Wei, L. J., and Ying, Z. (**1993**), "Checking the Cox model with cumulative sums of martingale-based residuals," Biometrika **80**, 557–572.

Linton, O. B., Nielsen, J. P., and Van De Geer, S. (**2003**), "Estimating Multiplicative and Additive Hazard Functions by Kernel Methods," The Annals of Statistics **31**, 464–492.

Lucas, A., Monteiro, A., and Smirnov, G. V. (**2006**), "Nonparametric Estimation for Non-Homogeneous Semi-Markov Processes: An Application to Credit Risk," Tinbergen Institute Discussion Paper No. 06-024/2. Available at SSRN: http://ssrn.com/abstract=892782 .

Mathieu, E., Foucher, Y., Dellamonica, P., and Daures, J. P. (**2007**), "Parametric and Non Homogeneous Semi-Markov Process for HIV Control," Methodology and Computing in Applied Probability **9**, 389–397.

Matthews, D. E. (**1984**), "Some Observations on Semi-Markov Models for Partially Censored Data," The Canadian Journal of Statistics **12**, 201–205.

McKeague, I. W. and Utikal, K. J. (**1990**a), "Identifying nonlinear covariate effects in semimartingale regression models," Probability Theory and Related Fields **87**, 1–25.

McKeague, I. W. and Utikal, K. J. (**1990**b), "Inference for a nonlinear counting process regression model," The Annals of Statistics **18**, 1172–1187.

McKeague, I. W. and Utikal, K. J. (**1991**), "Goodness-of-Fit Tests for Additive Hazards and Proportional Hazards Models," Scandinavian Journal of Statistics **18**, 177–195.

Meira-Machado, L., Uña-Álvarez, J. D., and Cadarso-Suárez, C. (**2006**), "Nonparametric estimation of transition probabilities in a non-Markov illnessCdeath model," Lifetime Data Analysis **12**, 325–344.

Meira-Machado, L., Uña-Álvarez, J. D., Cadarso-Suárez, C., and Andersen, P. K. (**2009**), "Multi-state models for the analysis of time-to-event data," Statistical Methods in Medical Research **18**, 195–222.

Nair, V. N. (**1984**), "Confidence Bands for Survival Functions with Censored Data: A Comparative Study," Technometrics **26**, 265–275.

Nelsen, R. B. (**1999**), *An introduction to copulas* (New York: Springer-Verlag).

Nielsen, J. P. and Linton, O. B. (**1995**), "Kernel Estimation in a Nonparametric Marker Dependent Hazard Model," The Annals of Statistics **23**, 1735–1748.

Oakes, D. and Cui, L. (**1993**), "On semiparametric inference for modulated renewal processes," Biometrika **81**, 83–90.

Ouhbi, B. and Limnios, N. (**1999**), "Nonparametric Estimation for Semi-Markov Processes Based on its Hazard Rate Functions," Statistical Inference for Stochastic Processes **2**, 151–173.

Park, Y., Tian, L., and Wei, L. J. (**2006**), "One- and two-sample nonparametric inference procedures in the presence of a mixture of independent and dependent censoring," Biostatistics **7**, 252–267.

Patil, P. N. (**1993**), "On the Least Squares Cross-Validation Bandwidth in Hazard Rate Estimation," The Annals of Statistics **21**, 1792–1810.

Peterson, A. V. (**1976**), "Bounds for a Joint Distribution Function with Fixed Sub-Distribution Functions: Application to Competing Risks," Proceedings of the National Academy of Sciences of USA **73**, 11–13.

Phelan, M. J. (**1990**a), "Bayes Estimation from a Markov Renewal Process," The Annals of Statistics, **18**, 603–616.

Phelan, M. J. (**1990**b), "Estimating the transition probabilities from censored Markov renewal processes," Statistics & probability letters **10**, 43–47.

Pollard, D. (**1990**), *Empirical processes: theory and applications* (Hayward: Institute of Mathematical Statistics).

Pons, O. (**2004**), "Estimation in a model for a semi-Markov process with covariates under right-censoring," Statistics **38**, 273–293.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (**1981**), "On the regression analysis of multivariate failure time data," Biometrika **68**, 373–379.

Rivest, L. P. and Wells, M. T. (**2001**), "A Martingale Approach to the Copula-Graphic Estimator for the Survival Function under Dependent Censoring," Journal of Multivariate Analysis **79**, 138–155.

Satten, G. and Sternberg, M. (**1999**), "Fitting Semi-Markov Models to Interval-Censored Data with Unknown Initiation Times," Biometrics **55**, 507–513.

Shu, Y., Klein, J. P., and Zhang, M. (**2007**), "Asymptotic theory for the Cox semi-Markov illness-death model," Lifetime Data Analysis **13**, 91–117.

Slud, E. V. and Rubinstein, L. V. (**1983**), "Dependent competing risks and summary survival curves," Biometrika **70**, 643–649.

Sternberg, M. and Satten, G. (**1999**), "Discrete-time nonparametric estimation for semi-Markov models of chain-of-events data subject to interval censoring and truncation," Biometrics **55**, 514–522.

Tsiatis, A. (**1975**), "A Nonidentifiability Aspect of the Problem of Competing Risks," Proceedings of the National Academy of Sciences of USA **72**, 20–22.

Voelkel, J. G. and Crowley, J. (**1984**), "Nonparametric Inference for a Class of Semi-Markov Processes with Censored Observations," The Annals of Statistics, **12**, 142–160.

Williams, J. S. and Lagakos, S. W. (**1977**), "Models for censored survival analysis: Constant-sum and variable-sum models," Biometrika **64**, 215–224.

Yang, M. C. K. and Hursch, C. J. (**1973**), "The Use of a Semi-Markov Model for Describing Sleep Patterns," Biometrics **29**, 667–676.

Ying, S. C. (**2006**), "Generalized longitudinal data analysis, with application to evaluating hospital utilization based on administrative database," Master's thesis, Simon Fraser University, Burnaby, BC, Canada.

Zheng, M. and Klein, J. P. (**1995**), "Estimates of marginal survival for dependent competing risks based on an assumed copula," Biometrika **82**, 127–138.

Zung, W. W., Naylor, T. H., Gianturco, D. T., and Wilson, W. P. (**1965**), "Computer simulation of sleep EEG patterns with a Markov chain model," Recent Advances in Biological Psychiatry **8**, 335–355.