Contributions to Semiparametric Inference to Biased-sampled and Financial Data

Tony Sit

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

ABSTRACT

Contributions to Semiparametric Inference to Biased-sampled and Financial Data

Tony Sit

This thesis develops statistical models and methods for the analysis of life-time and financial data under the umbrella of semiparametric framework. The first part studies the use of empirical likelihood on Lévy processes that are used to model the dynamics exhibited in the financial data. The second part is a study of inferential procedure for survival data collected under various biased sampling schemes in transformation and the accelerated failure time models.

During the last decade Lévy processes with jumps have received increasing popularity for modelling market behaviour for both derivative pricing and risk management purposes. Chan et al. (2009) introduced the use of empirical likelihood methods to estimate the parameters of various diffusion processes via their characteristic functions which are readily available in most cases. Return series from the market are used for estimation. In addition to the return series, there are many derivatives actively traded in the market whose prices also contain information about parameters of the underlying process. This observation motivates us to combine the return series and

the associated derivative prices observed at the market so as to provide a more reflective estimation with respect to the market movement and achieve a gain in efficiency. The usual asymptotic properties, including consistency and asymptotic normality, are established under suitable regularity conditions. We performed simulation and case studies to demonstrate the feasibility and effectiveness of the proposed method.

The second part of this thesis investigates a unified estimation method for semiparametric linear transformation models and accelerated failure time model under general biased sampling schemes. The methodology proposed is first investigated in Paik (2009) in which the length-biased case is considered for transformation models. The new estimator is obtained from a set of counting process-based unbiased estimating equations, developed through introducing a general weighting scheme that offsets the sampling bias. The usual asymptotic properties, including consistency and asymptotic normality, are established under suitable regularity conditions. A closed-form formula is derived for the limiting variance and the plug-in estimator is shown to be consistent. We demonstrate the unified approach through the special cases of left truncation, length-bias, the case-cohort design and variants thereof. Simulation studies and applications to real data sets are also presented.

Contents

List of Tables

Acknowledgments

This doctoral dissertation would not have been completed without Professor Zhiliang Ying's invaluable guidance and constant support. He has generously shared his insightful ideas and constructive suggestions, as well as his encouragement during my course of study. His scholarship and integrity have also set a role model to which I aspire. Feeling deeply honoured and fortunate, I would like to express my sincere gratitude towards him for being my advisor.

I am also deeply indebted to Professor Victor de la Peña, who has also mentored and advised me, Professor Steven Kou, Professor Mark Brown, and Professor Michael Shnaidman for serving on my dissertation committee. Not only did they invite some enlightening discussions that sharpened my understanding on this particular thesis, but they also opened the doors to a wide array of areas in statistics and probability which enriched my knowledge and research.

Studying at the Department of Statistics at Columbia University has been a rewarding experience. During my four years of study, I was exposed to a very stimulating learning environment where I acquired advanced knowledge at the frontier of statistical research. I would also like to extend my appreciation to Professor Ngai Hang Chan, Professor Minggao Gu, Professor Hoi Ying Wong, Professor Samuel Po Shing Wong and many other professors at The Chinese University of Hong Kong who were instrumental in helping me build a solid foundation for my future research. Financial support from Sir Edward Youde Memorial Fund is also gratefully acknowledged.

Coursework and research can sometimes be arduous. The heartwarming assistance, support and advice from Edward Cheng, Yongbum Cho, Vincent Dorie, Yang Feng, Chien-hsun Huang, Allen Tzu-Hsuan Hsu, Jane Paik Kim, Henry Kwai Hung Lam, Mengling Liu, Jingchen Liu, Heng Liu, Wenbin Lu, Radka Picková, Johannes Ruf, Ekaterina Vinkovskaya, Gongjun Xu, Chun Yip Yau, Yi Yu, Pengfei Zang, Junyi Zhang, Stephanie S Zhang and Haowen Zhong helped me overcome the challenges of the last four years.

Last, but not least, I would like to thank my parents, my brother and Yuet Ying for their unconditional support in encouraging me to achieve my goal in life, especially during the difficult and stressful times. It is their love and encouragement that has enabled me to complete this work.

To my parents and Ying

Chapter 1

Introduction

Semiparametric inference techniques have become increasingly important tools for solving statistical inference problems. These tools are particularly important when the statistical model for the data collected is semiparametric in the sense that it has one more unknown component that is of infinite dimension. Semiparametric models typically have one or more finite-dimensional Euclidean parameters of particular interest. This dissertation consists of semiparametric inference on data originated from finance and survival analysis. The first chapter analyses the use of empirical likelihood on financial market data. The second chapter focuses on the data arising from various types of biased sampling schemes in transformation models and accelerated failure time model.

It is desirable to lay down some necessary background before we start to study the details of the inferential procedure and the asymptotic properties of the estimators in various settings. In the following section, we will have a short introduction on Lévy processes and their applications in modelling financial data.

The two models studied in part two are transformation models and accelerated failure time model. They are two important models that are well suited for regression modelling of survival data and are relatively easy to fit. Since the counting process formulation given by Aalen (1975) in his Berkeley Ph.D. thesis, there has been a large number of literature including Andersen et al. (1993), Flemming and Harrington (1991) and Kalbfleisch and Prentice (2002) that place a strong emphasis on the counting process formulation. This formulation has become standard and it provides the framework under which we develop the second part of the thesis. A short review on those models and the counting process is included at the end of this chapter.

1.1 Lévy Processes and their Role in Financial Modelling

The Black-Scholes model has been very popular since its introduction in 1973. One should bear in mind, however, that this elegant theory is developed upon several crucial assumptions. They include the assumptions that there is no market friction, such as taxes and transaction costs, and that there are no constraints on the stock holding, etc. Moreover, empirical evidence suggests that the classical Black-Scholes model does not describe the statistical properties of financial time series very well. In Cont (2001), a more extended list of stylised features of financial data is given: (i) the log returns do not exhibit a normally-distributed pattern and (ii) the volatilities or the parameters of uncertainty estimated (or more generally the environment) change stochastically over time and are clustered.

Due to the fact that log returns of most financial assets do not follow a normal law, i.e. they are skewed and have an actual kurtosis higher than that of the normal distribution, other more flexible distributions are needed. We would like to have some models that are resemble Brownian motion in the sense that these processes still have the independent and stationary increment structure but they are based on a more general distribution than the normal. Such processes are called Lévy processes, in honour of Paul Lévy, the pioneer of the theory. Note that in order to define such a stochastic process with independent and stationary increments, the distribution has to be infinitely divisible; see Applebaum (2004) and Schoutens (2003).

Suppose $\phi(U)$ is the characteristic function of a distribution. If, for every positive integer, n, $\phi(u)$ is also the nth power of a characteristic function, we say that the distribution is infinitely divisible. One can define formally such a divisible distribution a stochastic process $\{X_t\}_{t\geq 0}$ as a Lévy process, which starts at zero and has independent and stationary increments such that the distribution of an increment over $[s, s+t], s, t \geq 0$. In other words, $X_{t+s} - X_s$ has $\{\phi(u)\}^t$ as its characteristic function.

Every Lévy process has a cádlág modification which is itself a Lévy process. We can hence assume that sample paths of a Lévy process, equipped with the natural filtration, are almost surely continuous from the right and have limits from the left.

The cumulant characteristic function $\psi(u) = \log \phi(u)$ is often called the characteristic exponent, which satisfies the following Lévy-Khintchine formula,

$$\psi(u) = i\gamma u - \frac{1}{2}\sigma^2 u^2 + \int \left[\exp\{iux\} - 1 - iuxI(|x| <)\right] \nu(dx),$$

where $\gamma \in \mathbb{R}$, $\sigma^2 \geq 0$ and ν is a measure on $\mathbb{R}$ $\{0\}$ with

$$\int \inf\{1, x^2\}\nu(dx) < \infty.$$

We call the Lévy process has a triplet of Lévy characteristics $(\gamma, \sigma^2, \nu(dx))$. The measure ν is called Lévy measure of $\{X_t\}_{t\geq 0}$. In many cases, Lévy processes are defined via the corresponding Lévy triplets. This makes the characteristic functions of these functions readily available. This is particularly important in our estimation procedure described in Chapter 2 of this dissertation.

Estimating the parameters of a Lévy process poses a nontrivial problem. Implementation of the most common and efficient estimation procedure, namely maximum likelihood estimation, may not be straightforward. Lévy-based models are usually described in terms of the Lévy density. As a consequence, the characteristic function of

X_t is known in closed form, but the corresponding density may not have a closed form expression. Even if, in some cases, the density enjoys a closed-form representation, it may still be intractable.

Although the evaluation of the density function of Lévy processes is cumbersome, their score function can be analytically tractable. This offers the background for quasi-likelihood estimator. One may consider the following approximation method of the score function: We adopt the orthogonal projection of the score function onto the real valued basis function $G = \{g(u_i, y), i = 1, 2, \ldots\}$ in Hilbert Space. One may include $G = \{\exp(u_i)y\}$ for example. The quasi-likelihood estimator can be obtained as the root of the quasi-likelihood equation:

$$0 = \frac{1}{n}\sum_{i=1}^{n} S_Q(Y_i;\boldsymbol{\theta}) = \left[\frac{\partial\gamma(\boldsymbol{\theta})}{\partial\boldsymbol{\theta})}\right]' \Sigma(\boldsymbol{\theta})^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} g(Y_i) - \gamma(\boldsymbol{\theta})\right],$$

where $S_Q(Y;\boldsymbol{\theta}) = E[S(Y;\boldsymbol{\theta})|1, g(X)]$, $\gamma(\boldsymbol{\theta}) = E[g(Y)]$ and $\Sigma(\boldsymbol{\theta}) = \mathrm{Var}(g(Y))$, $Y_i = X_i - X_{i-1}, i = 1, \ldots, n$. Readers are referred to Sueishi (2005) for more detail.

Another estimation procedure is (Continuum of) Genearlised Methods of Moments (GMM/CGMM). Feuerverger and McDunnough (1981a) proposed to choose a finite grid $\boldsymbol{u} = (u_1, \ldots, u_k)'$ and to use $2k$ moment conditions:

$$E[h_{\boldsymbol{\theta}}(Y)] = \mathbf{0},$$

where

$$\begin{aligned} h_{\boldsymbol{\theta}}(Y_i) \;\; = \;\; & \big[\cos(u_1Y_i) - \phi^R_{\boldsymbol{\theta}}(u_1), \ldots, \cos(u_kY_i) - \phi^R_{\boldsymbol{\theta}}(u_k), \\ & \sin(u_1Y_i) - \phi^I_{\boldsymbol{\theta}}(u_1), \ldots, \sin(u_kY_i) - \phi^I_{\boldsymbol{\theta}}(u_k)\big]', \end{aligned}$$

with $\phi^R(\cdot)$ and $\phi^I(\cdot)$ the real and imaginary parts of the characteristic function $\phi(\cdot)$. The GMM estimator is obtained by

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\beta}} \left[\widehat{h_n}(\boldsymbol{\theta})\right]' \boldsymbol{W}_n \left[\widehat{h_n}(\boldsymbol{\theta})\right],$$

where $\widehat{h_n}(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n} h_{\boldsymbol{\theta}}(Y_i)$ and $\boldsymbol{W}_n$ is an arbitrary weighting matrix. Essentially, the optimal solution to the above minimisation problem is given by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{K}_n^{-1/2}\widehat{h_n}(\boldsymbol{\theta})\|,$$

where $\boldsymbol{K}_n$ is the consistent estimator of the covariance matrix and $\|\cdot\|$ is the Euclidean norm.

Carrasco and Florens (2000) extended the GMM to the case of a continuum of moment conditions (CGMM). Let π be a probability density function. They introduced the new norm which takes into account the continuum set of moment conditions into the GMM framework. The norm is dened by

$$\|f\|^2 = \int f(u)\overline{f(u)}\pi du,$$

where $\bar{f}$ denotes the complex conjugate of f. Similar to the GMM case, the CGMM estimator can be obtained by minimising $\|K^{-1}h_n(u, \boldsymbol{\theta})\|$, where K is the covariance operator such that

$$Kf(u) = \int E[h(u, Y; \boldsymbol{\theta})\overline{h(u, Y; \boldsymbol{\theta})}]f(u)\pi(v)dv.$$

Since, however, the operator K^{-1} is not continuous, $K_n^1 f$ is not stable against small changes in f. Carrasco and Florens (2000) then replaced K_n^{-1} by the Tikhonov approximation and eventually define the CGMM estimator as follows:

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \|(K_n^{\alpha_n})^{-1/2}h_n(u, \boldsymbol{\theta})\|,$$

where $(K_n^{\alpha_n})^{-1} = (K_n^2 + \alpha_n I)^{-1}K_n$. However, the asymptotic result does not indicate how to choose α_n in practice. In Chapter 2, the empirical likelihood method is used for parameter estimation. We will present the details of the procedure as well as the reason why including observed derivative prices can provide a more responsive estimate of the model parameters.

1.2 Counting Processes and Semiparametric Models for Survival Data

The formal definition of a counting process is as follows: A counting process $\{N(t)\}$ is a stochastic process that is adapted to a filtration $\mathcal{F}_t$, cádlág with $N(0) = 0$ and $N(t) < \infty$ a.s.; and whose paths are piecewise constant with jumps of size 1.

In general setting, a counting process N is a local martingale. The corresponding compensator is denoted by Λ. Such a process is nondecreasing and predictable, zero at time zero and such that

$$M = N - \Lambda$$

is a local martingale with respect to $\mathcal{F}_t$. It holds that

$$E[N(t) - \Lambda(t)] = 0. \tag{1.1}$$

Under the so-called absolute continuous case, the compensator has the special form

$$\Lambda(t) = \int_0^t \lambda(s)ds,$$

where $\lambda(t)$ is regarded as the intensity process, which is a predictable process.

For right-censored survival data, where only the minimum $T \wedge C$ is observed with T and C denote respectively the event and the censoring times, the counting process $N(t)$ has a compensator $\Lambda(t) = \int_0^t Y(s)\lambda(s)ds$, where $Y(t) = I(\widetilde{T} \geq t)$ and $\lambda(t)$, which is also known as the hazard rate function, can be interpreted as the instantaneous failure rate and is given as follows

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \lim_{h \to 0} \frac{1}{h} \Pr\{t \leq T < t + h | T \geq t\}.$$

The mean zero property demonstrated in (1.1) provides the foundation upon which various estimating equations are developed.

Existing methods for analysis of survival data are largely based on the celebrated Coxs model (see Cox 1972) and pertained to a certain type of sampling design. The general linear transformation model, which includes Cox's model and proportional odds model as special cases makes a linear regression for the event time T on a scale given by the unknown strictly increasing function H given a p-dimensional covariate, say $\boldsymbol{Z} = (Z_1, \ldots, Z_p)$ such that

$$H(T) = -\boldsymbol{Z}'\boldsymbol{\beta} + \epsilon,$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is a p-dimensional regression parameter and the residual ϵ has a known distribution, say F.

The above model can be reparametrised and be written as

$$h(T) = \exp\{-\boldsymbol{Z}'\boldsymbol{\beta}\} \exp\{\epsilon\},$$

where $h = \exp(H)$ is a strictly increasing positive function such that $h(0) = 0$ and $\lim_{n\to\infty} H(t) = \infty$. Denote $\dot{h} = dH(t)/dt$. The hazard of T given $\boldsymbol{Z}$ can then be written as

$$\lambda(t) = \dot{h}(t) \exp\{\boldsymbol{Z}'\boldsymbol{\beta}\}\lambda_0(\exp\{\boldsymbol{Z}'\boldsymbol{\beta}\}H(t)), \tag{1.2}$$

where $\lambda_0(t)$ is the hazard associated with $\exp\{\epsilon\}$. When ϵ has the extreme value distribution, then $\exp\{\epsilon\}$ is standard exponentially distributed, which gives $\lambda_0(t) = 1$. The hazard function (1.2) is then reduced to the Cox regression model. Furthermore, suppose we denote $S(t|\boldsymbol{Z})$ as the conditional survival function of T given $\boldsymbol{Z}$, then we can write the transformation model into

$$S_\epsilon^{-1}(S_Z(t)) = H(t) + \boldsymbol{Z}'\boldsymbol{\beta}.$$

When ϵ follows the standard logistic distribution, we can see that the model becomes

$$\text{logit}(1 - S_Z(t)) = \log(h(t)) + \boldsymbol{Z}'\boldsymbol{\beta},$$

since in which case $F_\epsilon = \exp\{x\}/(1+\exp\{x\})$. The above two examples that connects the two popular models under one framework reveals that transformation model has some attractive features. However, for other choices of ϵ the regression coefficients are more difficult to interpret because of the unknown monotone transformation H. There has been a substantial volume of literature studying inference procedure of the regression parameters (and the transformation function) under various settings. Readers may refer to Chapter 3 for the references therein.

An alternative class of semiparametric models for survival data is called the accelerated failure time model. This model assumes that

$$\log(T) = -\boldsymbol{Z}'\boldsymbol{\beta} + \epsilon,$$

where $\boldsymbol{\beta}$ is a set of regression parameters and ϵ is the residual term with an unspecified distribution. Such a formulation allows the model to provide users with an easy interpretation because it directly relates the covariates with the level of $\log(T)$. Sir D. Cox himself once remarked in Reid (1994) that

> *Of course, another issue is the physical or substantive basis for the proportional hazards model. I think that's one of its weakness, that accelerated life models are in many ways more appealing because of their quite direct physical interpretation, particularly in an engineering context.*

Despite its easy interpretation, the model is not as easy to fit as the regression models that are more commonly known. The asymptotics of the estimators is more difficult to obtain due to the fact that inference has to be conducted via the residual term $\widetilde{T} - \boldsymbol{Z}'\boldsymbol{\beta}_0$. The existence of non-differentiable terms with respect to $\boldsymbol{\beta}$ explains the difficulty. The lack of monotonicity in the terms that involve $\boldsymbol{\beta}$ gives rise to the possibility of multiple solutions to estimating equations. Recent advances in studying the properties of this model include Tsiatis (1990), Ying (1993) and Jin et al. (2003) for example.

There is a connection between the two models. Chen and Wang (2000) consider a hazard rate model

$$\lambda(t|\boldsymbol{Z}) = \lambda_0[t\exp\{\boldsymbol{Z}'\boldsymbol{\beta}\}],$$

which they refer to the accelerated hazards model. This model can be used to model a range of hazard ratio shapes despite the drawback that it lacks a clear regression parameter interpretation. Chen and Jewell (2001) then consider a class of hazard rate models which are defined as follows:

$$\lambda(t|\boldsymbol{Z}) = \lambda_0[t\exp\{\boldsymbol{Z}'\boldsymbol{\beta}_1\}]\exp\{\boldsymbol{Z}'\boldsymbol{\beta}_2\}.$$

This model generalises both the proportional hazards model ($\boldsymbol{\beta}_1 = \mathbf{0}$) and the accelerated failure time model ($\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$) as well as the accelerated hazards model ($\boldsymbol{\beta}_2 = \mathbf{0}$). This model may play a useful role in discriminating between a proportional hazards and the accelerated failure time models.

Chapter 3 is dedicated to the study of the semiparametric inference of the above-mentioned models in biased sampling settings. This work further extends the results developed in Paik (2009) to a more general framework and to the accelerated failure time model. More examples and explanations will be elaborated in that chapter so that readers can appreciate why the unified approach proposed is practically useful and versatile.

Chapter 2

Parameter Estimation using Empirical Likelihood combined with Market Information

2.1 Introduction

Brownian motion and normal distribution have been widely used in the BlackScholes option-pricing framework to model the return of assets. Stylised facts, however, contradict with the model assumptions specified in the Black-Scholes framework. This motivated studies to modify the Black-Scholes model to explain the empirical phenomena. One direction of extension is to model the asset return dynamics through Lévy processes which are able to capture jumps and the asymmetric leptokurtic features. Readers are referred to Schoutens (2003) for a full account of Lévy processes and their applications in finance.

Empirical likelihood, introduced by Owen (1988), provides an alternative, nonparametric approach to inference. By placing a probability p_j on the jth observation and computing a profile likelihood, the method can be used to construct nonparametric point estimation as well as confidence regions for the parameters of interest. Qin and Lawless (1994) first linked estimating equations with empirical likelihood. In the paper, they developed methods of combining information about parameters in over-constrained optimisation problems which are frequently discussed in econometrics or financial literature. When the number of estimating equations is larger than the number of parameters of interest, the empirical likelihood estimation procedure will automatically combine the constraints by assigning them appropriate weights and produce an efficient estimate. Owen (2001) gives a good review on the development and applications of empirical likelihood.

Statistical inference based on the characteristic functions was proposed by Feuerverger and Mureika (1977), Feuerverger and McDunnough (1981a) for independent observations. Chan et al. (2009) suggested using characteristic functions as constraints for the empirical likelihood estimation. Such an approach makes use of the advantage that the characteristic functions of many diffusion processes are readily available, but it does not incorporate information from the market including derivative prices, for instance, which provide informative and most-updated knowledge of the parameters of interest. The key goal of this paper is to discuss how one can make use of the market data in the empirical likelihood estimation procedure to obtain more accurate estimates.

Let $\{S_t\}_{t\geq 0}$ be a continuous-time Lévy process that records the evolution of a financial security over a period of time. Assuming that S's are observed over a collection of discrete time points: $0, \delta, 2\delta, \ldots, n\delta$, over a time span $[0, n\delta]$, we can treat the difference of any two consecutive observations, i.e. the increments, as a set of independent observations with the same distribution since increments of a Lévy process are independently and identically distributed. In other words

$$R_j := \log S_{j\delta} - \log S_{(j-1)\delta} \sim_{iid} F_{\boldsymbol{\theta}}, \text{ say}$$

whose characteristic function is given by

$$\phi(t;\boldsymbol{\theta}) = E^{\mathbb{P}}[\exp\{itR_j\};\boldsymbol{\theta}] = \int \exp\{itr\}F_{\boldsymbol{\theta}}(dr),$$

where $\mathbb{P}$ denotes the expectation taken under the physical measure and $\boldsymbol{\theta}$ denotes the parameters of interest that governs the process $\{\log S_t\}_{t\geq 0}$. Of course, using the maximum likelihood approach can produce the most efficient parameter estimates. This is, however, only possible when the density function is readily available, which is not the case for most of the Lévy processes. In this paper, we follow Chan et al. (2009) to formulate an estimation procedure using the empirical likelihood with characteristic functions as one of the constraints. Observe that a characteristic function contains the same amount of model information as what a probability density function can carry, it is sensible to incorporate them as one of the estimating equations. Instead of having the return sequence as the only source of data, we can, in fact, incorporate information from actively-traded derivatives in order to provide a more timely estimate of the model parameters. In this paper, prices of European call options on the same underlying asset are used as moment constraints for empirical likelihood estimation procedure. Due to put-call parity between European calls and their put counterparts, it suffices to include just call prices as the put counterparts should contain the same amount of information.

The remainder of this chapter is organized as follows: we first define the notation and describe the methodology needed in Section 2.2. Sections 2.3 provides readers with specific examples on how to apply the results in Section 2 to carry out the estimation procedure. Section 2.4 extends the model to multi-period case. A simulation study and a case study are given in Section 2.5, followed by a discussion in Section 2.6. Proofs are relegated to the Appendix.

2.2 Methodology

2.2.1 Single Period Model

Throughout this section, we assume that $R_1, \ldots, R_n$ are iid random variables with distribution F and the characteristic function $\phi(t; \boldsymbol{\theta})$, whose closed-form formulation can be readily obtained. To begin, we start with the simplest possible: In addition to the return series, we also observe risk-free rate r as well as a call option with maturity δ and strike K. Lévy processes have independent stationary increments and so the above set up fits Lévy process.

Following Qin and Lawless (1994) and Chan et al. (2009), we study the maximum empirical likelihood estimator (MELE) based on constraints due to both the characteristic function as well as option prices as follows. First, it is easy to see that the equation $\phi(t; \boldsymbol{\theta}) = E[e^{itR_j}]$ provides us with two constraints on $\boldsymbol{\theta}$:

$$\sum_{j=1}^{n} p_j \cos(tR_j) = Re[\phi(t; \boldsymbol{\theta})] \quad \text{and} \quad \sum_{j=1}^{n} p_j \sin(tR_j) = Im[\phi(t; \boldsymbol{\theta})],$$

where $Re(z)$ and $Im(z)$ denote respectively the real and the imaginary parts of z.

For the option constraint, denote $\widetilde{c}(S_{n\delta}, K, r, \delta, \boldsymbol{\theta})$ the call price observed at time $n\delta$, with the underlying asset price $S_{n\delta}$ and strike K that matures at $(n+1)\delta$. To simplify the notation, for the rest of the paper, we suppress the subscript δ and use S_n and R_n to denote the underlying asset price and the associated return at time $n\delta$ respectively.

Observe that

$$\begin{aligned} 0 &= E^{\mathbb{Q}}[e^{-r\delta} \max\{S_{n+1} - K, 0\} | S_n] - \widetilde{c}(S_n, K, r, \delta, \boldsymbol{\theta}) \\ &= E^{\mathbb{Q}}[e^{-r\delta} \max\{S_n e^{R_{n+1}} - K, 0\} | S_n] - \widetilde{c}(S_n, K, r, \delta, \boldsymbol{\theta}) \\ &= E^{\mathbb{P}}\left[e^{-r\delta} \max\{S_n e^{R_{n+1}} - K, 0\} \frac{d\mathbb{Q}}{d\mathbb{P}}(R_{n+1}; \boldsymbol{\theta}) - \widetilde{c}(S_n, K, r, \delta, \boldsymbol{\theta}) \middle| S_n \right], \quad (2.1) \end{aligned}$$

where $\frac{d\mathbb{Q}}{d\mathbb{P}}(R;\boldsymbol{\theta})$ represents the Radon-Nikodym derivative (or the density ratio) of R that adjusts the difference between the probabilities defined under the physical and a risk neutral measures.

Option prices are usually specified through moneyness which is denoted by $m = S_n/K$. (2.1) can be rewritten as

$$\begin{aligned} 0 &= E^{\mathbb{P}}\left[S_n\left(e^{-r\delta}\max\{e^{R_{n+1}}-m,0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R_{n+1};\boldsymbol{\theta})-\frac{\widetilde{c}(S_n,K,r,\delta,\boldsymbol{\theta})}{S_n}\right)\middle|S_n\right] \\ &= E^{\mathbb{P}}\left[e^{-r\delta}\max\{e^{R}-m,0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R;\boldsymbol{\theta})-c(m,r,\delta,\boldsymbol{\theta})\right], \end{aligned} \tag{2.2}$$

which gives an additional constraint. Note that $c(m,r,\delta,\boldsymbol{\theta}) = \widetilde{c}(S_n,K,r,\delta,\boldsymbol{\theta})/S_n$ is independent of S_n in most cases; see Section 3.

The above derivation differs from Stutzer (1996)'s canonical approach in which case historical returns are used to construct n possible values for the asset price one period from now, i.e.

$$S_{n+1} = S_n e^{R_j}, j = 1,\ldots,n.$$

That is, the previous realised returns are used to construct possible prices at $(n+1)\delta$. Walker and Haley (2010) used a similar approach to investigate alternative tilts for non-parametric pricing. They proposed the following estimating equation:

$$E^{\mathbb{Q}}\left[e^{-r\delta}\max\{S_n e^{R}-K,0\}\right] = c(S_n,K,r,\delta,\boldsymbol{\theta}).$$

which will, however, create bias in cases with small sample sizes because of the projected asset price. The difference between the magnitudes of S_n, thus the additional constraint, with the two constraints derived from considering the characteristic function of the return series may produce unstable numerical estimates.

2.2.2 Model Setup

Denote $p_1(t), \ldots, p_n(t)$ be probability weights allocated to the residuals $\{\boldsymbol{g}_j(t;\theta)\}_{j=1,\ldots,n}$, where

$$\boldsymbol{g}_j(t;\boldsymbol{\theta}) = \begin{pmatrix} \cos(tR_j) - \phi^R(t;\boldsymbol{\theta}) \\ \sin(tR_j) - \phi^I(t;\boldsymbol{\theta}) \\ e^{-r\delta}\max\{e^{R_j} - K, 0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R_j;\boldsymbol{\theta}) - c(m,r,\delta,\boldsymbol{\theta}) \end{pmatrix}. \tag{2.3}$$

An empirical likelihood for $\boldsymbol{\theta}$ at t is given by

$$L_n(\tau,\boldsymbol{\theta}) = \prod_{j=1}^{n} p_j(t), \tag{2.4}$$

subject to constraints $\sum_{j=1}^n p_j(t) = 1$ and $\sum_{j=1}^n p_j(t)\boldsymbol{g}_j(t;\boldsymbol{\theta}) = \mathbf{0}$. Applying Lagrange-multiplier approach as we usually see in maximum empirical likelihood derivation, we see that (2.4) is maximized when

$$p_j(t) = \frac{1}{n}\frac{1}{\boldsymbol{\lambda}(t;\boldsymbol{\theta})'\boldsymbol{g}_j(t;\boldsymbol{\theta})},$$

where $\boldsymbol{\lambda}(t;\boldsymbol{\theta})$ is a Lagrange multiplier in $\mathbb{R}^k$ satisfying

$$\boldsymbol{Q}_{1n} =: \frac{1}{n}\sum_{j=1}^{n}\frac{\boldsymbol{g}_j(t;\boldsymbol{\theta})}{1+\boldsymbol{\lambda}(t;\boldsymbol{\theta})'\boldsymbol{g}_j(t;\boldsymbol{\theta})} = \mathbf{0}.$$

Hence, the local log empirical likelihood ratio becomes

$$\ell_n(t;\boldsymbol{\theta}) = 2\sum_{j=1}^{n}\log\left\{1+\boldsymbol{\lambda}(t;\boldsymbol{\theta})'\boldsymbol{g}_j(t;\boldsymbol{\theta})\right\}.$$

Like Chan et al. (2009), we consider integrating $\ell_n(t;\boldsymbol{\theta})$ against a probability weight $\pi(t)$, an integrated empirical likelihood ratio for $\boldsymbol{\theta}$ is given by

$$\ell_n(\boldsymbol{\theta}) = \int_{t\in\mathbb{R}} \ell_n(t;\boldsymbol{\theta})\pi(t)dt.$$

The maximum empirical likelihood estimator (MELE) for θ is defined as

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell_n(\boldsymbol{\theta}).$$

Remark: arg min is considered because -2 has been multiplied to the EL ratio $\ell_n(\boldsymbol{\theta})$.

The maximum empirical likelihood estimator (MELE) for θ is defined as

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell_n(\boldsymbol{\theta}).$$

Remark: arg min is considered because -2 has been multiplied to the EL ratio $\ell_n(\boldsymbol{\theta})$.

The above estimation procedure can be easily extended to situation in which there is more than one option traded in the market. In other words, options with the same maturity but different strikes can be added as additional constraints. For multiple strike constraints, say there are k European calls with moneynesses m_i $(i = 1, \ldots, k)$ respectively, one can simply rewrite (2.3) as

$$\boldsymbol{g}_j(t;\boldsymbol{\theta}) = \begin{pmatrix} \cos(tR_j) - \phi^R(t;\boldsymbol{\theta}) \\ \sin(tR_j) - \phi^I(t;\boldsymbol{\theta}) \\ e^{-r\delta}\max\{e^{R_j} - m_1, 0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R_j;\boldsymbol{\theta}) - c(m_1, r, \delta, \boldsymbol{\theta}) \\ \vdots \\ e^{-r\delta}\max\{e^{R_j} - m_m, 0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R_j;\boldsymbol{\theta}) - c(m_m, r, \delta, \boldsymbol{\theta}) \end{pmatrix}$$

through which we can obtain $\widehat{\boldsymbol{\theta}}_{EL}$ using the same estimation procedure.

2.2.3 Multiple-period Model

The above framework can be further extended to incorporate options with different strikes as well as different maturities. Similar to the single-period case set-up, suppose we have observed a series of returns $\{R_j\}_{j=1,\ldots,n}$ with the current asset price S_n. In addition, we can also obtain prices for calls with different maturities M and moneynesses m.

The procedure will follow closely to the methodology proposed in Section 2.2. We start from the simplest case in which there are two groups of calls: one group contains

N_1 calls with different moneynesses but the same maturity δ while the other group containing N_2 calls with different moneynesses but the same maturity 2δ.

The single-period case can be dealt as what we have done in Section 2. For the double-period model, we can view each pair of consecutive (non-overlapping) returns as a single observation. In this case, the double-period model can be reduced to the single-period model with $n/2$ number of observations. Essentially, it means

$$R_j^{(2)} = R_{2j-1} + R_{2j}, \qquad j = 1, \ldots, n/2$$

The corresponding set of estimating equations for the calls that mature in 2δ can be written as

$$\boldsymbol{g}_j^{(2)}(t;\boldsymbol{\theta}) = \begin{pmatrix} \cos(tR_j^{(2)}) - \phi^R(t;\boldsymbol{\theta}) \\ \sin(tR_j^{(2)}) - \phi^I(t;\boldsymbol{\theta}) \\ e^{-r\delta}\max\{e^{R_j^{(2)}} - m_1^{(2)}, 0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R^{(2)};\boldsymbol{\theta}) - c^{(2)}(m_1^{(2)}, r, \delta, \boldsymbol{\theta}) \\ \vdots \\ e^{-r\delta}\max\{e^{R_j^{(2)}} - m_{N_2}^{(2)}, 0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R^{(2)};\boldsymbol{\theta}) - c^{(2)}(m_{N_2}^{(2)}, r, \delta, \boldsymbol{\theta}) \end{pmatrix},$$

where $c^{(2)}$, and $m^{(2)}$ denote respectively the call prices and their corresponding moneynesses.

We can define $p_j^{(2)}(t)$, $\boldsymbol{Q}_{1n}^{(2)}$ and $l_n^{(2)}(t;\boldsymbol{\theta})$ accordingly for double-period model. In general, we can also extend above extension to multiple-period case in which g becomes

$$\boldsymbol{g}_j^{(k)}(t;\boldsymbol{\theta}) = \begin{pmatrix} \cos(tR_j^{(k)}) - \phi^R(t;\boldsymbol{\theta}) \\ \sin(tR_j^{(k)}) - \phi^I(t;\boldsymbol{\theta}) \\ e^{-r\delta}\max\{e^{R_j^{(k)}} - K_1^{(k)}, 0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R_j^{(k)};\boldsymbol{\theta}) - c^{(k)}(m_1^{(k)}, r, \delta, \boldsymbol{\theta}) \\ \vdots \\ e^{-r\delta}\max\{e^{R_j^{(k)}} - K_{m^{(k)}}^{(k)}, 0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R_j^{(k)};\boldsymbol{\theta}) - c^{(k)}(m_{m^{(k)}}^{(k)}, r, \delta, \boldsymbol{\theta}) \end{pmatrix},$$

where $R_j^{(k)} = R_{(j-1)k+1} + \ldots + R_{kj}$.

Following the idea of Chan et al. (2009), we try to express our overall likelihood as a sum of all the sub-empirical likelihood. The maximum likelihood estimator for $\boldsymbol{\theta}$ can be defined similarly as

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{j=1}^{n_O} \ell_n^{(j)}(\boldsymbol{\theta}), \tag{2.5}$$

where n_O denotes the number of unique maturities of the options observed. Readers should be noted that, for simplicity, we just use the call prices as constraints. One can use any other option prices as long as they can write down the estimating equations. The inclusion of the puts may not help estimation due to the put-call parity. This methodology, of course, performs worse when the maximum maturity becomes long that leads to a huge reduction of the number of observations. One should note that, however, only options with short maturities are traded actively. These options, meanwhile, provide the most up-to-date, thus useful, information about the parameters.

2.3 Models and Examples

In this section, three commonly used models with known characteristic functions are considered. Discretely observed data are used to investigate the performance of the proposed empirical likelihood estimator to provide an accurate estimate of the unknown parameters of the continuous time models studied.

2.3.1 Black-Scholes Model

Suppose the stock price S_t follow the geometric Brownian motion

$$d\log S_t = (\mu - \frac{\sigma^2}{2})dt + \sigma dW_t, \tag{2.6}$$

where W_t is a $\mathbb{P}$-Brownian motion. Again, we denote the historical returns of the previous n trading period as $R_j = \log S_j - \log S_{(j-1)}, j = 1, \ldots, n$, we know that for each j,

$$R_j \overset{iid}{\sim} \mathcal{N}\left((\mu - \frac{\sigma^2}{2})\delta, \sigma^2\delta\right).$$

The characteristic function of S_j is given by

$$\phi(t; \boldsymbol{\theta}) = \exp\{\delta(it(\mu - \sigma^2/2) - \sigma^2 t^2/2)\},$$

where $\boldsymbol{\theta} = (\mu, \sigma)$. Hence, for any $j = 1, \ldots, n$,

$$E^{\mathbb{P}}[e^{itR} - \exp\{\delta(it(\mu - \sigma^2/2) - \sigma^2 t^2/2)\}] = 0 \tag{2.7}$$

is an estimating equation for $\boldsymbol{\theta}$.

In addition to the return series, we also observe option prices traded at time $n\delta$, each of them expires in the next period of length δ: $\{c(m_j, r, \boldsymbol{\theta})\}_{j=1,\ldots,k}$. From these k option prices, we can write down an estimating equation for the parameters $\boldsymbol{\theta} = (\mu, \sigma)$:

$$0 = E^{\mathbb{P}}\left[e^{-r\delta}\max\{e^R - m, 0\}\frac{d\mathbb{Q}}{d\mathbb{P}}(R; \boldsymbol{\theta}) - c(m, r, \delta, \boldsymbol{\theta})\right],$$

where, if (2.6) holds,

$$\begin{aligned}
\frac{d\mathbb{Q}}{d\mathbb{P}}(R; \boldsymbol{\theta}) &= \left[\frac{1}{\sqrt{2\pi\sigma^2\delta}}\exp\left\{-\frac{(R - (r - \sigma^2/2)\delta)^2}{2\sigma^2\delta}\right\}\right] \\
&\quad \times \left[\frac{1}{\sqrt{2\pi\sigma^2\delta}}\exp\left\{-\frac{(R - (\mu - \sigma^2/2)\delta)^2}{2\sigma^2\delta}\right\}\right]^{-1} \\
&= \exp\left\{-\frac{1}{2\sigma^2\delta}\left[(R - (r - \sigma^2/2)\delta)^2 - (R_j - (\mu - \sigma^2/2)\delta)^2\right]\right\} \\
&= \exp\left\{\frac{r - \mu}{\sigma^2}R - \frac{r^2 - \mu^2}{2\sigma^2}\delta + \frac{(r - \mu)\delta}{2}\right\}.
\end{aligned}$$

This leads to the following estimating equation

$$E\left[\left\{e^{-r\delta}\max\{e^R - m, 0\} - c(m, r, \delta, \boldsymbol{\theta})\right\}\exp\left\{\frac{r - \mu}{\sigma^2}R - \frac{r^2 - \mu^2}{2\sigma^2}\delta + \frac{(r - \mu)\delta}{2}\right\}\right] = 0. \tag{2.8}$$

2.3.2 Black-Scholes Model with Merton Jumps (BS-MJ)

Empirical studies suggest that log return sequences usually exhibit skewness and an excess kurtosis (compared with a normal distribution). In order to devise a model that can provide a better fit to the financial market data, Merton (1976), believing that the Black-Scholes solution is not valid as the stock prices dynamics should not be presented by a stochastic process with a continuous path, proposed Black-Scholes Model with jumps (BS-MJ), which is specified as follows:

$$dS_t = (\mu - \lambda\kappa)S_t dt + \sigma S_t dW_t + (J_t - 1)S_t dN_t, \tag{2.9}$$

where N_t is a Poisson process with intensity parameter $\lambda > 0$ and J_t is the jump size following a lognormal distribution $\log - \mathcal{N}(\mu_J, \sigma_J^2)$ and is independent of W_t. $\lambda\kappa := \lambda E[J_t - 1] = \lambda(\exp\{\mu_J + \sigma_J^2/2\} - 1)$ is the compensator of the compound Poisson process $(J_t - 1)S_t dN_t$.

By Ito's lemma for jump diffusion processes (see Shreve 2004), (2.9) can be rewritten as

$$d\log S_t = (\mu - \lambda\kappa - \sigma^2/2)dt + \sigma dW_t + \log J_t dN_t, \tag{2.10}$$

under the physical measure $\mathbb{P}$. Despite the fact that there is no closed form density for $\log S_t$, its characteristic function is given as follows:

$$\phi(t;\boldsymbol{\theta}) = \exp\left\{\delta\left[it(\mu - \lambda\kappa - \sigma^2/2) - \sigma^2 t^2/2 + \lambda\left(e^{i\mu_J t - \sigma_J^2 t^2/2} - 1\right)\right]\right\}. \tag{2.11}$$

By constructing a hedging portfolio, Merton (1976) proposed that the European call option price on an equity that follows the dynamics given by (2.10) $V(S_t, t)$ should be the solution of

$$\frac{\partial V}{\partial t} + \frac{\sigma^2 S_t^2}{2}\frac{\partial^2 V}{\partial S_t^2} + rS_t\frac{\partial V}{\partial S_t} - rV + \lambda E[V(J_t S_t, t) - V(S_t, t)] - \lambda S_t \frac{\partial V}{\partial S_t} E[J_t - 1] = 0,$$

which is equal to

$$
\begin{aligned}
V(t, S_t) &= e^{-r\tau} E^{\mathbb{Q}_M}[\max\{S_(t+\delta) - K, 0\} | S_t] \\
&= \sum_{n \geq 0} \frac{e^{-\bar{\lambda}\delta}(\bar{\lambda}\delta)^n}{n!} V^{BS}\left(\delta, S_t; \sigma_n, r_n\right),
\end{aligned}
$$

with

$$
\begin{aligned}
\bar{\lambda} &= \lambda(1+\kappa) = \lambda \exp\{\mu_J + \sigma_J^2/2\} \\
\sigma_n &= \sqrt{\sigma^2 + n\sigma_J^2/\delta} \\
r_n &= r - \lambda\kappa + \frac{n\mu_J + n\sigma_J^2/2}{\delta} \\
V^{BS}(\delta, S, \sigma, r) &= S\mathcal{N}\left(\frac{\log(S/K) + (r+\sigma^2/2)\delta}{\sigma\sqrt{\delta}}\right) \\
&\quad -Ke^{-r\delta}\mathcal{N}\left(\frac{\log(S/K) + (r-\sigma^2/2)\delta}{\sigma\sqrt{\delta}}\right).
\end{aligned}
$$

Again, we need to compute the Radon-Nikodym derivative between the two measures $\mathbb{P}$ and $\mathbb{Q}_M$. Using the inverse Fourier transform formula, we can express the density of the Merton's jump diffusion model under physical measure $\mathbb{P}$ as follows:

$$
\begin{aligned}
f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t; \boldsymbol{\theta}) e^{-itx} dt \\
&= \frac{e^{\lambda\delta}}{2\pi} \int_{-\infty}^{\infty} e^{i\delta t(\mu - \lambda\kappa - \sigma^2/2) - \sigma^2 t^2/2} \exp\{e^{it\mu_J - t^2\sigma_J^2/2}\lambda\delta\} dt \\
&= \frac{e^{\lambda\delta}}{2\pi} \int_{-\infty}^{\infty} e^{i\delta t(\mu - \lambda\kappa - \sigma^2/2) - \sigma^2 t^2/2} \sum_{n \geq 0} \frac{(\lambda\delta)^n}{n!} (e^{it\mu_J - t^2\sigma_J^2/2})^n dt \\
&= \frac{e^{\lambda\delta}}{2\pi} \sum_{n \geq 0} \frac{(\lambda\delta)^n}{n!} \int_{-\infty}^{\infty} \exp\{i\delta t(\mu - \lambda\kappa - \sigma^2/2) \\
&\quad -\sigma^2 t^2/2 - itx + int\mu_J - nt^2\sigma_J^2/2\} dt.
\end{aligned}
$$

Using the identity that $\int_{-\infty}^{\infty} e^{-az^2 + ibz} dz = \sqrt{\frac{\pi}{a}} e^{-b^2/4a}$, putting $a = \delta\sigma^2/2 + n\sigma_J^2/2$, $b = (\mu - \sigma^2/2 - \lambda\kappa)\delta - x + n\mu_J$, we can write

$$
f(x) = \frac{e^{-\lambda\delta}}{\sqrt{2\pi}} \sum_{n \geq 0} \frac{(\lambda\delta)^n}{n!} \frac{e^{-\frac{[(\mu - \sigma^2/2 - \lambda\kappa)\delta + n\mu_J - x]^2}{2(\delta\sigma^2 + n\sigma_J^2)}}}{\sqrt{\delta\sigma^2 + n\sigma_J^2}},
$$

which is a fast converging sequence. So, the Radon-Nikodym derivative required is

$$\begin{aligned}\frac{d\mathbb{Q}_M}{d\mathbb{P}}(R_j;\boldsymbol{\theta}) &= \left[\sum_{n\geq 0}\frac{(\lambda\delta)^n}{n!}\frac{e^{-\frac{[(\mu-\sigma^2/2-\lambda\kappa)\delta+n\mu_J-R_j]^2}{2(\delta\sigma^2+n\sigma_J^2)}}}{\sqrt{\delta\sigma^2+n\sigma_J^2}}\right] \\ &\quad\times\left[\sum_{n\geq 0}\frac{(\lambda\delta)^n}{n!}\frac{e^{-\frac{[(r-\sigma^2/2-\lambda\kappa)\delta+n\mu_J-R_j]^2}{2(\delta\sigma^2+n\sigma_J^2)}}}{\sqrt{\delta\sigma^2+n\sigma_J^2}}\right]^{-1}.\end{aligned}$$

In other words, the corresponding estimating equation that is derived from an option is given by

$$\begin{aligned}0 &= E\Bigg[\left(e^{-r\delta}\max\{e^R-m,0\}-\sum_{n\geq 0}\frac{e^{-\bar{\lambda}\delta}(\bar{\lambda}\delta)^n}{n!}V^{BS}(\delta,S_t;\sigma_n,r_n)\right) \\ &\quad\left(\sum_{n\geq 0}\frac{(\lambda\delta)^n}{n!}\frac{e^{-\frac{[(\mu-\sigma^2/2-\lambda\kappa)\delta+n\mu_J-R_j]^2}{2(\delta\sigma^2+n\sigma_J^2)}}}{\sqrt{\delta\sigma^2+n\sigma_J^2}}\right)\left(\sum_{n\geq 0}\frac{(\lambda\delta)^n}{n!}\frac{e^{-\frac{[(r-\sigma^2/2-\lambda\kappa)\delta+n\mu_J-R_j]^2}{2(\delta\sigma^2+n\sigma_J^2)}}}{\sqrt{\delta\sigma^2+n\sigma_J^2}}\right)^{-1}\Bigg].\end{aligned}$$

To generate $\log S_t$ from (2.10), we use a sequence of Bernoulli processes to approximate the Poisson jump process. Discretised sample paths can be generated through

$$\log S_{n+1} = \log S_n + (\mu-\lambda\kappa-\sigma^2/2)\delta+\sigma\sqrt{\delta}Z+\sum_{l=1}^{200}N_lJ_l,$$

where Z denotes a standard normal random variable, $J_l \sim \mathcal{N}(\mu_J,\sigma_J^2)$ and $N_l \sim \text{Bernoulli}\left((\lambda\delta/200)\exp\{-\lambda\delta/200\}\right)$.

2.3.3 Double-Exponential Jump Model

Kou (2002) proposed a jump-diffusion similar Merton's, where the jump size is double-exponentially distributed. The double-exponential jump diffusion (DEJD) model is designed to capture the leptokurtic feature of the empirical return distributions as well as the volatility smile in option markets which cannot be successfully modeled by BS-MJ model. The canonical decomposition of the driving process of Kou's model is

$$dS_t = \mu S_t dt + \sigma S_t dW_t + S_t d\left(\sum_{i=1}^{N(t)}(V_i - 1)\right), \tag{2.12}$$

where W_t is a standard Brownian motion, $N(t)$ is a Poisson process with rate λ and $\{V_i\}$ is a sequence of independent identically distributed non-negative random variables such that $Y \triangleq \log(V)$ has an asymmetric double exponential distribution with the density

$$f_Y(y) = p\eta_1 e^{-\eta_1 y}\mathbf{1}_{\{y\geq 0\}} + (1-p)\eta_2 e^{\eta_2 y}\mathbf{1}_{\{y<0\}},$$

with $\eta_1, \eta_2 > 0$, where $p \geq 0$ represent the probabilities of upward and downward jumps, i.e.

$$\log(V) = Y =_d \begin{cases} \xi^+ & \text{, with probability p} \\ -\xi^- & \text{, with probability 1-p,} \end{cases}$$

where ξ^+ and ξ^- are exponential random variables with means η_1^{-1} and η_2^{-1} respectively. In model (2.12), all sources of randomness, $N(t)$, $W(t)$ and Y's are assumed to be independent.

The analytical solution of a call option whose price is determined by an underlying asset that is driven by DEJD model also incorporates a psychological interpretation of investors. As we can see in (2.12), this model has six parameters, namely μ, the drift parameter, σ, the diffusion volatility, λ, the Poisson rate, p, the probability of having an upward jump, η_1, the rate of an upward exponential jump and η_2, the rate of a downward exponential jump. By incorporating option prices observed with different strikes and maturities, we can improve the estimation, compared with incorporating merely the characteristic function of the model. In addition, the option prices used can also enable the estimation of the parameters involved in the utility function.

Given (2.12), one can write down the dynamic of $d\log S_t$ by using Ito's Lemma:

$$d\log S_t = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dW_t + d\left(\sum_{i=1}^{N(t)} Y_i\right),$$

from which we can derive the characteristic function of $\log S_t$ (see Cont and Tankov 2004) under the risk-neutral probability measure without taking the jump risk into account:

$$\phi_T(u) = E[e^{iu\log S_t}] = \exp\left\{t\left(\left[\log S_0 + r - \frac{\sigma^2}{2}\right]u + iu\lambda\left[\frac{p}{\lambda_+ - iu} - \frac{1-p}{\lambda_- + iu}\right]\right)\right\},$$

since the Lévy density of the jump is

$$\nu(x) = p\lambda\eta_1 e^{-\eta_1 x}\mathbf{1}_{\{x>0\}} + (1-p)\lambda\eta_2 e^{\eta_2 x}\mathbf{1}_{\{x\leq 0\}}.$$

The corresponding European call price can be obtained via Carr and Madan (1999) method, which is specified as follows:

$$C(S_0, K, T, r) = \frac{e^{-\alpha\log K}}{2\pi}\int_{-\infty}^{\infty} e^{-iu\log K}\frac{e^{-rT}\phi_T(u-(\alpha+1)i)}{\alpha^2+\alpha-u^2+i(2\alpha+1)u}du.$$

Using the independence between the exponential and normal distributions used in the model and formulae for the sum of double exponential random variables, Kou (2002) obtains the probability density function of the return, which can be approximated by the following density function:

$$\begin{aligned} f_R(x;\mu) &:= \frac{1-\lambda\delta}{\sigma\sqrt{\delta}}\phi\left(\frac{x-\mu\delta}{\sigma\sqrt{\delta}}\right) \\ &= \lambda\delta\left\{p\eta_1 e^{(\sigma^2\eta_1^2\delta)/2}e^{-(x-\mu\delta)\eta_1}\Phi\left(\frac{x-\mu\delta-\sigma^2\eta_1\delta}{\sigma\sqrt{\delta}}\right)\right. \\ &\quad \left. + q\eta_2 e^{(\sigma^2\eta_2^2\delta)/2}e^{(x-\mu\delta)\eta_2}\ \Phi\left(-\frac{x-\mu\delta+\sigma^2\eta_2\delta}{\sigma\sqrt{\delta}}\right)\right\}, \end{aligned}$$

which can be used to define the Radon-Nikodym derivative to adjust for the difference between a risk-free probability measure (in Merton's sense) and the physical measure since

$$\frac{d\mathbb{Q}^M}{d\mathbb{P}} = \frac{f_R(x;r)}{f_R(x;\mu)}.$$

The estimating equation derived from the option price is

$$0 \;=\; E^{\mathbb{P}}\left[\left(e^{-r\delta}\max\{e^{R}-m,0\}-C(S_n,K,T,r)/S_n\right)\frac{d\mathbb{Q}^M}{d\mathbb{P}}(R;\boldsymbol{\theta})\right].$$

2.3.4 With Jump Risk Premium

Kou (2002) considered a typical rational expectations economy (Lucas (1978)) in which a representative investor has the utility function of the special form, as in Naik and Lee (1990):

$$U(c,t)=\begin{cases} e^{-\kappa t}c^{\alpha}/\alpha & \text{, if } 0<\alpha<1 \\ e^{-\kappa t}\log(c) & \text{, if } \alpha=0, \end{cases} \tag{2.13}$$

with $U_c(c,t) \triangleq \frac{\partial U(c,t)}{\partial c}$. The goal of the representative investor is to obtain $\max_c E[\int_0^\infty U(c(t),t)dt]$. In his model, Kou also assumed E_t, an endowment process, which is, under the physical measure $\mathbb{P}$, specified as follows:

$$\frac{dE_t}{E_t}=\mu_1 dt+\sigma_1 dW_t^{(1)}+d\left[\sum_{l=1}^{N(t)}(\tilde{V}_l-1)\right]; \tag{2.14}$$

given the endowment process (2.14), the asset price will have the dynamic of the form

$$\frac{dS_t}{S_t}=\mu dt+\sigma\{\rho dW_t^{(1)}+\sqrt{1-\rho^2}dW_t^{(2)}\}+d\left[\sum_{l=1}^{N(t)}(V_l-1)\right], \tag{2.15}$$

where $dW_t^{(2)}$ is a Brownian motion independent of $dW_t^{(1)}$ and $V_l=\tilde{V}_l^{\beta}$. Furthermore, α and κ in (2.13) are related as follows:

$$\mu=\kappa+(1-\alpha)\left\{\mu_1-\frac{1}{2}\sigma_1^2(2-\alpha)+\sigma_1\sigma\rho\right\}-\lambda\zeta_1^{(\alpha+\beta-1)},$$

where $\zeta_1^{(a)} \triangleq E[\tilde{V}^a-1]$.

It can been shown (see, for example, Stokey and Lucas 1989) that, under mild conditions, the rational expectations equilibrium price, or the “shadow” price, of the security $p(t)$, must satisfy the Euler equation

$$p(t)=\frac{E[e^{-\theta T(\delta(T))^{\alpha-1}}p(T)|\mathcal{F}_t]}{e^{-\theta t}(\delta(t))^{\alpha-1}},\quad \forall T\in[t,T_0], \tag{2.16}$$

where U_c is the partial derivative of U with respect to c. To simplify the model, we assume $E_t = S_t$, i.e. $\mu_1 = \mu$, $\sigma_1 = \sigma$ and $\rho = \beta = 1$. It follows that, as shown in (10) of Kou (2002), the Radon-Nikodym derivative between the risk-free measure $\mathbb{Q}$ and the physical measure $\mathbb{P}$ is given by

$$\begin{aligned}\frac{d\mathbb{Q}}{d\mathbb{P}}(R_{(j+1)\delta};\theta) &= \frac{e^{r(j+1)\delta}U_c(S_{(j+1)\delta},(j+1)\delta))}{e^{rj\delta}U_c(S_{j\delta},j\delta)} = e^{(r-\kappa)\delta}\left[\frac{S_{(j+1)\delta}}{S_{j\delta}}\right]^{\alpha-1} \\ &= e^{\delta\left[(1-\alpha)\mu-\frac{1}{2}\sigma^2(1-\alpha)(2-\alpha))\right]}e^{R_{(j+1)\delta}(\alpha-1)}.\end{aligned}$$

Here λ is a Poisson process with rate λ. The jump sizes $\{Y_1, Y_2, \ldots\}$ are independent identically distributed random variables such that $Y_i = \log(V_i)$. The moment generating function of $X(t) := \log(S_t/S_0)$ can be obtained as

$$E\left[e^{\theta X(t)}\right] = \exp\left\{G(\theta)t\right\},$$

where $G(x) = \widetilde{\mu}x + \frac{1}{2}x^2\sigma^2 + \lambda\left(E\left[e^{xY}\right] - 1\right)$. In the case of Merton's normal jump-diffusion model,

$$G(x) = \widetilde{\mu}x + \frac{1}{2}x^2\sigma^2 + \lambda\left\{\mu_J x + \frac{x^2\sigma_J^2}{2} - 1\right\};$$

and in the case of double exponential jump-diffusion model

$$G(x) = \widetilde{\mu}x + \frac{1}{2}x^2\sigma^2 + \lambda\left(\frac{p\eta_1}{\eta_1 - x} + \frac{(1-p)\eta_2}{\eta_2 + x} - 1\right).$$

Under the risk-neutral probability $\mathbb{Q}$, we have

$$\widetilde{\mu} = r - \frac{1}{2}\sigma^2 - \lambda\zeta,$$

where $\zeta := E\left[e^Y\right] - 1$. In the Merton's model

$$\zeta = E^{\mathbb{Q}}\left[e^Y\right] - 1 = \mu_J + \frac{\sigma_J^2}{2} - 1,$$

while in the double exponential jump-diffusion model

$$\zeta = \frac{p\eta_1}{\eta_1 - 1} + \frac{(1-p)\eta_2}{\eta_2 + 1} - 1.$$

Kou, Petrella and Wang (2005) adapted the method in Carr and Madan (1999), which is based on a change of the order of integration, to price European call and put options via Laplace transforms. The Laplace transform with respect to k of $C(S, e^k, r, T)$ is given by

$$\begin{aligned}\widehat{f}_C(\xi) &:= \int_{-\infty}^{\infty} e^{-\xi k} C(S, e^k, r, T) dk \\ &= e^{-rT} \frac{S^{\xi+1}}{\xi(\xi+1)} \exp\left\{G(\xi+1)T\right\}, \quad \xi > 0.\end{aligned}$$

This leads to the following estimating equation:

$$E\left[\left(e^{-r\delta} \max\left\{e^R - m, 0\right\} - c(m, r, \delta, \boldsymbol{\theta})\right) e^{\delta\left[(1-\alpha)\mu - \frac{1}{2}\sigma^2(1-\alpha)(2-\alpha)\right)\right]} e^{R(\alpha-1)}\right] = 0.$$

Note that the option price c under the double exponential jump diffusion dynamics can also be obtained directly using the method proposed by Kou (2002),

$$\begin{aligned}c(m; r, \delta, \boldsymbol{\theta}) &= \Upsilon\left(r + \frac{1}{2}\sigma^2 - \lambda\zeta, \sigma, \tilde{\lambda}, \tilde{p}, \tilde{\eta}_1, \tilde{\eta}_2; \log(1/m), \delta\right) \\ &+ me^{-r\delta}\Upsilon\left(r - \frac{1}{2}\sigma^2 - \lambda\zeta, \sigma, \tilde{\lambda}, \tilde{p}, \tilde{\eta}_1, \tilde{\eta}_2; \log(1/m), \delta\right),\end{aligned}$$

where the definitions of $\tilde{\lambda}, \tilde{p}, \tilde{\eta}_1, \tilde{\eta}_2$ and $\Upsilon(\cdot)$ can be found in Kou (2002).

The characteristic function of a return derived from the price driven by the process (2.15) can be obtained similarly as in Sections 3.2 and 3.3. The corresponding estimating equations are thus

$$E[\cos(tR)] - \mathrm{Re}(\phi(t; \boldsymbol{\theta})) = E[\sin(tR)] - \mathrm{Im}(\phi(t; \boldsymbol{\theta})) = 0.$$

2.4 Asymptotic Results

Regularity conditions:

1. $E[\boldsymbol{g}(t, X_1; \boldsymbol{\theta}_0)\boldsymbol{g}(t, X_1; \boldsymbol{\theta}_0)']$ is positive definite for $t \in [-a, a], a > 0$;

2. $\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{g}(t, x; \boldsymbol{\theta})$ is continuous in a neighbourhood of $\boldsymbol{\theta}_0$, for $t \in [-a, a], x \in \mathbb{R}$;

3. $\sup_{\boldsymbol{\theta} \in \Theta} \|\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{g}(t, x; \boldsymbol{\theta})\| \leq H(t, x)$, where

$$\int_{t=-a}^{a} \int_{x=-\infty}^{\infty} H(t, x) dF(x) dG_1(t) < \infty;$$

4. The rank of $E[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{g}(t, X_1; \boldsymbol{\theta}_0)]$ is $\min\{2, d\}$ for all $t \in [-a, a]$, where d is the dimension of $\boldsymbol{\theta}$;

5. $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \boldsymbol{g}(t, x; \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ for $\boldsymbol{\theta} \in \Theta$, $t \in [-a, a]$ and $x \in \mathbb{R}$;

6. $\sup_{\boldsymbol{\theta} \in \Theta} \|\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \boldsymbol{g}(t, x; \boldsymbol{\theta})\| \leq H(t, x)$, where H is given in 3.

Proposition 2.4.1. *Under conditions 1-4, with probability one, denote* $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} T_1(\boldsymbol{\theta})$ *which satisfies* $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \leq n^{-1/3}$*,*

$$\begin{aligned} \boldsymbol{Q}_{1n}(t, \widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}_1(t; \widehat{\boldsymbol{\theta}})) &= 0 \\ \int_{-a}^{a} \boldsymbol{Q}_{2n}(t, \widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}_1(t; \widehat{\boldsymbol{\theta}})) dG_1(t) &= 0, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{Q}_{1n}(t; \boldsymbol{\theta}, \boldsymbol{\lambda}) &= \frac{1}{n} \sum_{j=1}^{n} \frac{\boldsymbol{g}(t; x_j, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}' \boldsymbol{g}(t; x_j, \boldsymbol{\theta})}, \\ \boldsymbol{Q}_{2n}(t; \boldsymbol{\theta}, \boldsymbol{\lambda}) &= \frac{1}{n} \sum_{j=1}^{n} \frac{1}{1 + \boldsymbol{\lambda}' \boldsymbol{g}(t; x_j; \boldsymbol{\theta})} \frac{\partial \boldsymbol{g}'(t, x_j; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\lambda}. \end{aligned}$$

Proposition 2.4.2. *Under conditions 1-6, for the estimator* $\widehat{\boldsymbol{\theta}}$ *given in Proposition 2.1, we have as* $n \to \infty$*,*

$$\begin{aligned} \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= -\left\{ \int_{-a}^{a} \boldsymbol{s}_{21}(t) \boldsymbol{s}_{11}^{-1}(t) \boldsymbol{s}_{12}(t) dG_1(t) \right\}^{-1} \\ &\quad \times \left\{ \int_{-a}^{a} \boldsymbol{s}_{21}(t) \boldsymbol{s}_{11}^{-1}(t) \sqrt{n} \boldsymbol{Q}_{1n}(t; \boldsymbol{\theta}_0, 0) dG_1(t) \right\} + o_p(1) \\ &\to_d N(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{s}_{11}(t) &= -E[\boldsymbol{g}(t,R_1;\boldsymbol{\theta}_0)\boldsymbol{g}(t,R_1;\boldsymbol{\theta}_0)'],\\
\boldsymbol{s}_{12}(t) &= \boldsymbol{s}_{21}'(t) = E[\frac{\partial}{\partial\boldsymbol{\theta}}g(t,R_1;\boldsymbol{\theta}_0)],\\
\boldsymbol{\Sigma} &= \left\{\int_{-a}^{a}\boldsymbol{s}_{21}(t)\boldsymbol{s}_{11}^{-1}(t)\boldsymbol{s}_{12}(t)dG_1(t)\right\}^{-1}\\
&\quad\times\left\{\int_{-a}^{a}\int_{-a}^{a}\boldsymbol{s}_{21}(t_1)\boldsymbol{s}_{11}^{-1}(t_1)\boldsymbol{\Gamma}(t_1,t_2)\boldsymbol{s}_{11}^{-1}(t_2)\boldsymbol{s}_{12}(t_2)dG_1(t_1)dG_2(t_2)\right\}\\
&\quad\times\left\{\int_{-a}^{a}\boldsymbol{s}_{21}(t)\boldsymbol{s}_{11}^{-1}(t)\boldsymbol{s}_{12}(t)dG_1(t)\right\}^{-1},
\end{aligned}$$

where

$$\boldsymbol{\Gamma}(t_1,t_2) = E[\boldsymbol{g}(t_1,R_t;\boldsymbol{\theta}_0)\boldsymbol{g}(t_1,R_t;\boldsymbol{\theta}_0)'].$$

Corollary 2.4.1. *When $k_1 > k_2$, the asymptotic variance $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_k$ of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ cannot decrease if an estimating equation is dropped.*

2.5 Numerical Results

2.5.1 Simulations

For each model, 500 sample paths with size $n = 125, 250, 500$ and 1000 starting at initial value $\log S_0 = 100$ with frequency $\delta = 1/52$ were simulated. Similar to Chan et al. (2009) approach, we also choose the uniform weight function $G(t)$ and l_n can be approximated by the Riemann sum of $l_n(t)$ evaluated at $t \in [-5.0, 5.0]$ with the number of grids set to be 100. [1] Simulation results for BS, BS-MJ and DEJD are tabulated in Tables 2.1,2.2 and 2.4 respectively. As we can see from the simulation results, by incorporating more option prices, the estimated standard deviation of the estimates are reduced, which is due to the result of Corollary 2.4.1.

[1] In Chan et al. (2009), they chose the interval to be $[-0.5, 0.5]$. In the simulation studies, we found that using $[-0.5, 0.5]$ produced poor estimations. A wider interval chosen allows the data to provide more information about the parameter values.

2.5.2 Case Study

We examine empirically whether the proposed methodology can be applied to the real data set and what insights call prices can reveal when we incorporate them into the model. Historical S&P 500 index values and corresponding call option prices between 2 January 1987 and 31 December 2008 were downloaded from Wharton Research Data Services (WRDS). The index is sampled at 4-day frequency and in total we have 1,260 data points. The duration of four days is chosen so as to match the time to maturity of the call prices. The calls were traded on Chicago Board Options Exchange (CBOE).

We included, in our simulations, from one to four call prices that were most frequently traded on the last day of our analyses so as to reflect the market information on that particular trading day. The mean annual rate of return is 0.0531 with the associated volatility equals 0.1328. In addition to the market crash of 1987, the tech- and credit-bubble between the late 90's and mid 2000's as well as September 11 attack in 2001, the sample period also covered the recent Lehmann Brother's collapse as a result of credit crunch in 2008. In particular, our data analysis was done with the last day selected as September 29, 2008 - the day on which the largest single day plunge was recorded shortly after Lehman brothers' and Washington Mutual's bankruptcy. Furthermore, on that day, the Volatility S&P (VIX), a measure of market volatility, has the record highest jump in history. Estimated values of the parameters and the associated estimated asymptotic variances are tabulated in Tables 6 - 8.

It can be seen from the tables that by incorporating constraints due to observed option prices, one can lower the variance of the estimates. It should be also noted that in order for Chan et al. (2009)'s approach to achieve the same magnitude of variance as what we can see by including additionally one call price, the sample size should have to be roughly doubled. In other words, using call prices as constraints reduces the required sample size at the expense that the equity price dynamics are specified by a particular model. Since the option prices are considered as a summary of the current market view on the underlying equity price dynamics, our methodology can successfully capture more updated estimate of the current market condition. This can be seen in the data analysis results in which the volatility and/or jump size estimates are both larger than the estimates that Chan et al. (2009) provided, which can be interpreted as the consequence because of the late-2000's financial crisis. Finally, we comment that, due to the small number of option prices included, it is challenging to produce an accurate estimate α, the risk-preference parameter of investors of which information can be only derived from the option prices.

2.6 Conclusion

Lévy processes are an excellent tool for modeling price processes in mathematical finance. Its popularity arises from its flexibility and simple structure in comparison with general semimartingales. Estimation for Lévy processes are challenging statistical inference problems because of the lack of analytical expression for the transitional density function. Inspired by Chan et al. (2009) that uses integrated empirical likelihood approach for parameter estimation, we propose in this chapter incorporating call prices as constraints in addition to using the characteristic function associated with the process. This method provides a more efficient estimate that can reflect the recent market condition more accurately which is demonstrated via simulations and real data analyses. The idea of using derivative prices as one of the estimating equations is not restricted to call prices only; in fact, any price that can be expressed in terms of expectation of an independent random variable that follows the same distribution as specified by the underlying process are eligible for being included as one of constraints. The approach, therefore, has robust theoretical and versatility for a wide range of processes including processes with jump components.

2.7 Appendix

Proof of Proposition 4.1. It follows closely the proof of lemma 1 of Qin and Lawless (1994). □

Proof of Proposition 4.2. Similar to Chan et al. (2009), we can show that

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{Q}_{1n}(t;\boldsymbol{\theta};0) &\stackrel{P}{\to} s_{12}(t), \\
\frac{\partial}{\partial \boldsymbol{\lambda}'} \boldsymbol{Q}_{1n}(t;\boldsymbol{\theta};0) &\stackrel{P}{\to} s_{11}(t), \\
\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{Q}_{2n}(t;\boldsymbol{\theta};0) &= \mathbf{0}, \\
\frac{\partial}{\partial \boldsymbol{\lambda}'} \boldsymbol{Q}_{2n}(t;\boldsymbol{\theta};0) &\stackrel{P}{\to} s_{21}(t)
\end{aligned}$$

uniformly in $t \in [-a, a]$. Denote $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| + \sup_{t\in[-a,a]} \|\boldsymbol{\lambda}_1(t;\widehat{\boldsymbol{\theta}})\|$. Then, we can expand $\boldsymbol{Q}_{1n}$ and $\boldsymbol{Q}_{2n}$ using Taylor series expansions and yield

$$\begin{aligned}
\boldsymbol{\lambda}_1(t;\widehat{\boldsymbol{\theta}}) &= -s_{11}^{-1}(t)\boldsymbol{Q}_{1n}(t;\boldsymbol{\theta}_0,0) - s_{11}^{-1}(t)s_{12}(t)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(\delta_n), \quad \text{and} \\
0 &= \int_{-a}^{a} \boldsymbol{Q}_{2n}(t;\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}_1(t;\widehat{\boldsymbol{\theta}}))dG_1(t) \\
&= \int_{-a}^{a} \Big\{ \boldsymbol{Q}_{2n}(t;\boldsymbol{\theta}_0,0) + \frac{\partial \boldsymbol{Q}_{2n}(t;\boldsymbol{\theta}_0,0)}{\partial \theta}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\
&\quad + \frac{\partial \boldsymbol{Q}_{2n}(t;\boldsymbol{\theta}_0,0)}{\partial \boldsymbol{\lambda}'}\boldsymbol{\lambda}_1(t;\widehat{\boldsymbol{\theta}}) \Big\} dG_1(t) + o_p(\delta_n) \\
\Rightarrow \quad \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 &= -\left\{ \int_{-a}^{a} s_{21}(t)s_{11}^{-1}(t)s_{12}(t)dG_1(t) \right\}^{-1} \\
&\quad \times \left\{ \int_{-a}^{a} s_{21}(t)s_{11}^{-1}(t)\boldsymbol{Q}_{1n}(t;\boldsymbol{Q}_{1n}(t;\boldsymbol{\theta}_0,0))dG_1(t) \right\} \\
&\quad + o_p(\delta_n),
\end{aligned}$$

which completes the proof. □

Proof of Corollary 5.1. To prove the Corollary, it suffices to show

$$\begin{aligned}
&\int_{-a}^{a}\int_{-a}^{a} s_{21}(t_1)s_{11}^{-1}(t_1)\boldsymbol{\Gamma}(t_1,t_2)s_{11}^{-1}(t_2)s_{12}(t_2)dG_1(t_1)dG_1(t_2) \\
&- \int_{-a}^{a} s_{21}(t_1)s_{11}^{-1}(t_1)s_{12}(t_1)dG_1(t_1) \le 0,
\end{aligned}$$

where $A \leq B$ denotes $A - B$ is a negative-semidefinite. Observe that, by Cauchy-Schwarz inequality,

$$\begin{aligned}
& \int_{-a}^{a}\int_{-a}^{a} s_{21}(t_1)s_{11}^{-1}(t_1)\mathbf{\Gamma}(t_1,t_2)s_{11}^{-1}(t_2)s_{12}(t_2)dG_1(t_1)dG_1(t_2) \\
\leq & \int_{-a}^{a}\int_{-a}^{a} \left[s_{21}(t_1)s_{11}^{-1}(t_1)s_{11}(t_1)s_{11}(t_1)s_{21}(t_1)\right]^{1/2} \\
& \left[s_{11}(t_2)s_{11}^{-1}(t_2)s_{12}(t_2)s_{11}(t_2)s_{21}(t_2)\right]^{1/2} dG_1(t_1)dG_1(t_2) \\
= & \left[\int_{-a}^{a} \left(s_{21}(t_1)s_{11}^{-1}(t_1)s_{21}(t_1)\right)^{1/2} dG_1(t_1)\right]^{\otimes 2} \\
\leq & \int_{-a}^{a} s_{21}(t_1)s_{11}^{-1}(t_1)s_{21}(t_1)dG_1(t_1).
\end{aligned}$$

It follows that $\mathbf{\Sigma} \leq \left[\int_{-a}^{a} s_{21}(t)s_{11}^{-1}(t)s_{21}(t)dG_1(t)\right]^{-1}$. The proof can be completed following Qin and Lawless (1994). Write

$$\begin{aligned}
s_{12}(t;\boldsymbol{\theta}) &= \left[\left(\frac{\partial \boldsymbol{g}_1}{\partial \boldsymbol{\theta}}\right)', \ldots, \left(\frac{\partial \boldsymbol{g}_{k-1}}{\partial \boldsymbol{\theta}}\right)', \left(\frac{\partial \boldsymbol{g}_k}{\partial \boldsymbol{\theta}}\right)'\right] \triangleq \left[s_{12}^{-}(t;\boldsymbol{\theta}), \left(\frac{\partial \boldsymbol{g}_k}{\partial \boldsymbol{\theta}}\right)'\right] \\
s_{11}(t;\boldsymbol{\theta}) &= \begin{pmatrix} s_{11,a}(t,\boldsymbol{\theta}) & s_{11,b}(t,\boldsymbol{\theta}) \\ s_{11,c}(t,\boldsymbol{\theta}) & s_{11,d}(t,\boldsymbol{\theta}) \end{pmatrix},
\end{aligned}$$

where $s_{11,a}(t,\boldsymbol{\theta})$ is a $(k-1)\times(k-1)$ matrix. Then, for all $t \in [-a,a]$,

$$\begin{aligned}
s_{21}(t)s_{11}^{-1}(t)s_{21}(t) &= \left[s_{12}^{-}(t;\boldsymbol{\theta}), \left(\frac{\partial \boldsymbol{g}_k}{\partial \boldsymbol{\theta}}\right)'\right] \begin{pmatrix} s_{11,a}(t,\boldsymbol{\theta}) & s_{11,b}(t,\boldsymbol{\theta}) \\ s_{11,c}(t,\boldsymbol{\theta}) & s_{11,d}(t,\boldsymbol{\theta}) \end{pmatrix} \\
&\quad \times \left[s_{12}^{-}(t;\boldsymbol{\theta}), \left(\frac{\partial \boldsymbol{g}_k}{\partial \boldsymbol{\theta}}\right)'\right]' \\
&\geq \left[s_{12}^{-}(t;\boldsymbol{\theta}), \left(\frac{\partial \boldsymbol{g}_k}{\partial \boldsymbol{\theta}}\right)'\right] \begin{pmatrix} s_{11,a}(t,\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \left[s_{12}^{-}(t;\boldsymbol{\theta}), \left(\frac{\partial \boldsymbol{g}_k}{\partial \boldsymbol{\theta}}\right)'\right]' \\
&= s_{12}^{-}(t;\boldsymbol{\theta})s_{11,a}(t,\boldsymbol{\theta})s_{12}^{-}(t;\boldsymbol{\theta})'
\end{aligned}$$

which completes the proof. $\square$

	0 strike	1 strike	2 strikes	4 strikes
K	NA (Chan et al. 2009)	$0.99S$	$0.99S$, $1.01S$,	0.98, $0.99S$ $1.01S$, $1.02S$
$n = 125$	$\widehat{\mu} = 0.044(0.192)$ $\widehat{\sigma} = 0.298(0.019)$	$\widehat{\mu} = 0.056(0.190)$ $\widehat{\sigma} = 0.2992(0.012)$	$\widehat{\mu} = 0.045(0.189)$ $\widehat{\sigma} = 0.2998(0.009)$	$\widehat{\mu} = 0.041(0.117)$ $\widehat{\sigma} = 0.299(0.008)$
$n = 250$	$\widehat{\mu} = 0.047(0.133)$ $\widehat{\sigma} = 0.299(0.014)$	$\widehat{\mu} = 0.054(0.132)$ $\widehat{\sigma} = 0.2996(0.009)$	$\widehat{\mu} = 0.050(0.130)$ $\widehat{\sigma} = 0.2999(0.0088)$	$\widehat{\mu} = 0.042(0.130)$ $\widehat{\sigma} = 0.2869(0.073)$
$n = 500$	$\widehat{\mu} = 0.051(0.097)$ $\widehat{\sigma} = 0.300(0.010)$	$\widehat{\mu} = 0.052(0.091)$ $\widehat{\sigma} = 0.2998(0.0070)$	$\widehat{\mu} = 0.051(0.0984)$ $\widehat{\sigma} = 0.2999(0.0071)$	$\widehat{\mu} = 0.053(0.0873)$ $\widehat{\sigma} = 0.2996(0.0064)$
$n = 1000$	$\widehat{\mu} = 0.047(0.068)$ $\widehat{\sigma} = 0.3000(0.007)$	$\widehat{\mu} = 0.054(0.069)$ $\widehat{\sigma} = 0.2949(0.0069)$	$\widehat{\mu} = 0.054(0.0069)$ $\widehat{\sigma} = 0.295(0.0066)$	$\widehat{\mu} = 0.052(0.0071)$ $\widehat{\sigma} = 0.295(0.0062)$

Table 2.1: Black-Scholes Model (BS)

	0 strike	1 strike	2 strikes	4 strikes
K	NA (Chan et al. 2009)	$0.99S$	$0.99S$, $1.01S$	0.98, $0.99S$ $1.01S$, $1.02S$
$n = 125$	$\widehat{\mu} = 0.068(0.5829)$ $\widehat{\sigma} = 0.1597(0.3086)$ $\widehat{\lambda} = 2.7485(10.0591)$ $\widehat{\mu_J}$ =-0.3121(0.6929) $\widehat{\sigma_J} = 0.4440(0.4946)$	$\widehat{\mu} = 0.0329(0.1903)$ $\widehat{\sigma} = 0.2305(0.1608)$ $\widehat{\lambda} = 2.8674(6.0220)$ $\widehat{\mu_J}$ =-0.3445(0.4306) $\widehat{\sigma_J} = 0.4665(0.2547)$	$\widehat{\mu} = 0.0103(0.1281)$ $\widehat{\sigma} = 0.2392(0.1602)$ $\widehat{\lambda} = 3.1887(5.4130)$ $\widehat{\mu_J}$ =-0.2749(0.2912) $\widehat{\sigma_J} = 0.4474(0.2468)$	$\widehat{\mu} = 0.0244(0.1644)$ $\widehat{\sigma} = 0.2709(0.1378)$ $\widehat{\lambda} = 3.0066(2.8460)$ $\widehat{\mu_J}$ =-0.2670(0.2848) $\widehat{\sigma_J} = 0.5009(0.2273)$
$n = 250$	$\widehat{\mu} = 0.0639(0.3790)$ $\widehat{\sigma} = 0.1613(0.2518)$ $\widehat{\lambda} = 1.8401(3.5736)$ $\widehat{\mu_J}$ =-0.3172(0.4775) $\widehat{\sigma_J} = 0.4229(0.4559)$	$\widehat{\mu} = 0.0443(0.1435)$ $\widehat{\sigma} = 0.2304(0.1424)$ $\widehat{\lambda} = 2.3803(2.9758)$ $\widehat{\mu_J}$ =-0.2993(0.3115) $\widehat{\sigma_J}$ =0.5126(0.2216)	$\widehat{\mu} = 0.0258(0.085)$ $\widehat{\sigma} = 0.2561(0.1303)$ $\widehat{\lambda} = 2.5194(2.0394)$ $\widehat{\mu_J}$ =-0.2811(0.2445) $\widehat{\sigma_J} = 0.5047(0.2254)$	$\widehat{\mu} = 0.0235(0.0824)$ $\widehat{\sigma} = 0.2806(0.1250)$ $\widehat{\lambda} = 2.5063(2.0116)$ $\widehat{\mu_J}$ =-0.2723(0.2397) $\widehat{\sigma_J} = 0.5269(0.2057)$
$n = 500$	$\widehat{\mu} = 0.0740(0.2405)$ $\widehat{\sigma} = 0.2033(0.2094)$ $\widehat{\lambda} = 1.7045(0.9860)$ $\widehat{\mu_J}$ =-0.2411(0.3034) $\widehat{\sigma_J} = 0.5153(0.2159)$	$\widehat{\mu} = 0.0650(0.1228)$ $\widehat{\sigma} = 0.2259(0.1379)$ $\widehat{\lambda} = 2.0594(1.4227)$ $\widehat{\mu_J}$ =-0.2864(0.2636) $\widehat{\sigma_J} = 0.5203(0.2143)$	$\widehat{\mu} = 0.0524(0.1057)$ $\widehat{\sigma} = 0.3290(0.1097)$ $\widehat{\lambda} = 1.7884(1.2292)$ $\widehat{\mu_J}$ =-0.3097(0.2377) $\widehat{\sigma_J} = 0.5344(0.2119)$	$\widehat{\mu} = 0.0597(0.1076)$ $\widehat{\sigma} = 0.3498(0.1226)$ $\widehat{\lambda} = 1.9321(1.0289)$ $\widehat{\mu_J}$ =-0.3049(0.2280) $\widehat{\sigma_J} = 0.5748(0.1987)$
$n = 1000$	$\widehat{\mu} = 0.0315(0.1900)$ $\widehat{\sigma} = 0.2349(0.1658)$ $\widehat{\lambda} = 1.8302(0.9038)$ $\widehat{\mu_J}$ =-0.2807(0.2423) $\widehat{\sigma_J} = 0.5800(0.2155)$	$\widehat{\mu} = 0.0317(0.0884)$ $\widehat{\sigma} = 0.2553(0.1282)$ $\widehat{\lambda} = 1.9637(0.8853)$ $\widehat{\mu_J}$ =-0.3068(0.2161) $\widehat{\sigma_J} = 0.6030(0.1720)$	$\widehat{\mu} = 0.0326(0.0825)$ $\widehat{\sigma} = 0.3432(0.1017)$ $\widehat{\lambda} = 1.6987(0.7044)$ $\widehat{\mu_J}$ =-0.2999(0.1857) $\widehat{\sigma_J} = 0.6095(0.1618)$	$\widehat{\mu} = 0.0396(0.0813)$ $\widehat{\sigma} = 0.3407(0.0975)$ $\widehat{\lambda} = 1.9238(0.7021)$ $\widehat{\mu_J} = -0.2776(0.1627)$ $\widehat{\sigma_J} = 0.6051(0.1470)$

Table 2.2: Black-Scholes model with Merton Jumps (BSMJ)

	1 strike	2 strikes	4 strikes
K	$0.99S$	$0.99S$ $1.01S$	$0.98S, 0.99S$ $1.01S, 1.02S$
$n = 125$	$\widehat{\mu} = 0.0974(0.1840)$ $\widehat{\sigma} = 0.2607(0.1305)$ $\widehat{\lambda} = 1.9939(0.0349)$ $\widehat{\mu_J} = -0.0874(0.1070)$ $\widehat{\sigma_J} = 0.0370(0.1530)$ $\widehat{\alpha} = 0.6956(1.4105)$	$\widehat{\mu} = 0.0794(0.1682)$ $\widehat{\sigma} = 0.2741(0.1193)$ $\widehat{\lambda} = 1.9956(0.0306)$ $\widehat{\mu_J} = -0.1004(0.0994)$ $\widehat{\sigma_J} = 0.0564(0.1612)$ $\widehat{\alpha} = 0.6460(1.2068)$	$\widehat{\mu} = 0.0586(0.1603)$ $\widehat{\sigma} = 0.2826(0.1069)$ $\widehat{\lambda} = 1.9980(0.0311)$ $\widehat{\mu_J} = -0.1088(0.1051)$ $\widehat{\sigma_J} = 0.0592(0.1642)$ $\widehat{\alpha} = 0.6429(1.1079)$
$n = 250$	$\widehat{\mu} = 0.0966(0.1422)$ $\widehat{\sigma} = 0.2792(0.1140)$ $\widehat{\lambda} = 1.9934(0.0250)$ $\widehat{\mu_J} = -0.0811(0.0847)$ $\widehat{\sigma_J} = 0.0461(0.1668)$ $\widehat{\alpha} = 0.6744(1.1766)$	$\widehat{\mu} = 0.0866(0.1382)$ $\widehat{\sigma} = 0.2789(0.1053)$ $\widehat{\lambda} = 1.9920(0.0511)$ $\widehat{\mu_J} = -0.0927(0.0818)$ $\widehat{\sigma_J} = 0.0496(0.1659)$ $\widehat{\alpha} = 0.6382(1.0552)$	$\widehat{\mu} = 0.0732(0.1315)$ $\widehat{\sigma} = 0.2921(0.0756)$ $\widehat{\lambda} = 1.9955(0.0247)$ $\widehat{\mu_J} = -0.1009(0.0853)$ $\widehat{\sigma_J} = 0.0523(0.1641)$ $\widehat{\alpha} = 0.6449(0.9589)$
$n = 500$	$\widehat{\mu} = 0.1027(0.1121)$ $\widehat{\sigma} = 0.2949(0.0699)$ $\widehat{\lambda} = 1.9938(0.0203)$ $\widehat{\mu_J} = -0.0841(0.0693)$ $\widehat{\sigma_J} = 0.0607(0.1655)$ $\widehat{\alpha} = 0.5703(0.7906)$	$\widehat{\mu} = 0.0995(0.1087)$ $\widehat{\sigma} = 0.2982(0.0552)$ $\widehat{\lambda} = 1.9932(0.0154)$ $\widehat{\mu_J} = -0.0897(0.0644)$ $\widehat{\sigma_J} = 0.0605(0.1635)$ $\widehat{\alpha} = 0.5499(0.7051)$	$\widehat{\mu} = 0.0876(0.1047)$ $\widehat{\sigma} = 0.2982(0.0391)$ $\widehat{\lambda} = 1.9939(0.0136)$ $\widehat{\mu_J} = -0.0990(0.0623)$ $\widehat{\sigma_J} = 0.0819(0.1608)$ $\widehat{\alpha} = 0.5846(0.6543)$
$n = 1000$	$\widehat{\mu} = 0.1044(0.0860)$ $\widehat{\sigma} = 0.3032(0.0461)$ $\widehat{\lambda} = 1.9946(0.0096)$ $\widehat{\mu_J} = -0.0822(0.0594)$ $\widehat{\sigma_J} = 0.1007(0.1488)$ $\widehat{\alpha} = 0.5846(0.6543)$	$\widehat{\mu} = 0.1040(0.0869)$ $\widehat{\sigma} = 0.3026(0.0369)$ $\widehat{\lambda} = 1.9947(0.0080)$ $\widehat{\mu_J} = -0.0832(0.0551)$ $\widehat{\sigma_J} = 0.1042(0.1493)$ $\widehat{\alpha} = 0.5678(0.6310)$	$\widehat{\mu} = 0.0918(0.0855)$ $\widehat{\sigma} = 0.3040(0.0.0279)$ $\widehat{\lambda} = 1.9960(0.0068)$ $\widehat{\mu_J} = -0.0946(0.0494)$ $\widehat{\sigma_J} = 0.1212(0.1386)$ $\widehat{\alpha} = 0.5722(0.5827)$

Table 2.3: Merton Jump-diffusion model

	1 strike	2 strikes	4 strikes
K	$0.99S$	$0.99S$ $1.01S$	$0.98S, 0.99S$ $1.01S, 1.02S$
$n = 125$	$\widehat{\mu} = 0.1592(0.1997)$ $\widehat{\sigma} = 0.2274(0.1999)$ $\widehat{\lambda} = 1.7825(0.9630)$ $\widehat{p} = 0.1107(0.7369)$ $\widehat{\eta_1} = 7.6935(0.3580)$ $\widehat{\eta_2} = 8.8874(0.7513)$ $\widehat{\alpha} = 0.3415(1.7920)$	$\widehat{\mu} = 0.1473(0.1713)$ $\widehat{\sigma} = 0.3109(0.0658)$ $\widehat{\lambda} = 1.8095(0.6815)$ $\widehat{p} = 0.2571(0.9921)$ $\widehat{\eta_1} = 7.6935(0.3580)$ $\widehat{\eta_2} = 8.9638(0.5798)$ $\widehat{\alpha} = 0.1705(1.3924)$	$\widehat{\mu} = 0.1129(0.1777)$ $\widehat{\sigma} = 0.3141(0.0528)$ $\widehat{\lambda} = 1.9343(0.6037)$ $\widehat{p} = 0.3422(0.5358)$ $\widehat{\eta_1} = 7.5872(0.4652)$ $\widehat{\eta_2} = 8.9658(0.4665)$ $\widehat{\alpha} = 0.2717(1.2334)$
$n = 250$	$\widehat{\mu} = 0.1152(0.1876)$ $\widehat{\sigma} = 0.2969(0.0641)$ $\widehat{\lambda} = 1.8130(0.6496)$ $\widehat{p} = 0.3424(0.4645)$ $\widehat{\eta_1} = 7.5572(0.6307)$ $\widehat{\eta_2} = 8.9710(0.7906)$ $\widehat{\alpha} = 0.1835(1.2546)$	$\widehat{\mu} = 0.1115(0.1527)$ $\widehat{\sigma} = 0.3018(0.0443)$ $\widehat{\lambda} = 1.8397(0.6693)$ $\widehat{p} = 0.3934(0.4164)$ $\widehat{\eta_1} = 7.5784(0.3063)$ $\widehat{\eta_2} = 9.0307(0.4292)$ $\widehat{\alpha} = 0.0800(1.1524)$	$\widehat{\mu} = 0.1007(0.1469)$ $\widehat{\sigma} = 0.3023(0.0606)$ $\widehat{\lambda} = 1.9636(0.5531)$ $\widehat{p} = 0.3897(0.3653)$ $\widehat{\eta_1} = 7.5292(0.4211)$ $\widehat{\eta_2} = 8.9892(0.4004)$ $\widehat{\alpha} = 0.2256(1.0019)$
$n = 500$	$\widehat{\mu} = 0.1044(0.1478)$ $\widehat{\sigma} = 0.3065(0.0355)$ $\widehat{\lambda} = 1.8205(0.6064)$ $\widehat{p} = 0.4018(0.3516)$ $\widehat{\eta_1} = 7.5446(0.3409)$ $\widehat{\eta_2} = 9.0713(0.3104)$ $\widehat{\alpha} = 0.0764(0.8658)$	$\widehat{\mu} = 0.1206(0.1239)$ $\widehat{\sigma} = 0.3036(0.0329)$ $\widehat{\lambda} = 1.8907(0.7034)$ $\widehat{p} = 0.4041(0.3138)$ $\widehat{\eta_1} = 7.4794(0.4696)$ $\widehat{\eta_2} = 9.0671(0.3318)$ $\widehat{\alpha} = 0.0220(0.8857)$	$\widehat{\mu} = 0.0929(0.1245)$ $\widehat{\sigma} = 0.2976(0.0487)$ $\widehat{\lambda} = 2.0139(0.5391)$ $\widehat{p} = 0.4030(0.3050)$ $\widehat{\eta_1} = 7.4912(0.3298)$ $\widehat{\eta_2} = 8.9822(0.2867)$ $\widehat{\alpha} = 0.1735(0.7628)$
$n = 1000$	$\widehat{\mu} = 0.1109(0.1201)$ $\widehat{\sigma} = 0.2889(0.0366)$ $\widehat{\lambda} = 2.0648(0.6382)$ $\widehat{p} = 0.4004(0.2498)$ $\widehat{\eta_1} = 7.4740(0.2350)$ $\widehat{\eta_2} = 8.9613(0.3358)$ $\widehat{\alpha} = 0.1242(0.6115)$	$\widehat{\mu} = 0.1169(0.1031)$ $\widehat{\sigma} = 0.2891(0.0400)$ $\widehat{\lambda} = 2.1036(0.7695)$ $\widehat{p} = 0.4026(0.2132)$ $\widehat{\eta_1} = 7.4330(0.3336)$ $\widehat{\eta_2} = 8.9589(0.3884)$ $\widehat{\alpha} = 0.0254(0.5691)$	$\widehat{\mu} = 0.0725(0.0955)$ $\widehat{\sigma} = 0.2982(0.0343)$ $\widehat{\lambda} = 2.0227(0.4550)$ $\widehat{p} = 0.4398(0.2069)$ $\widehat{\eta_1} = 7.4613(0.3112)$ $\widehat{\eta_2} = 9.0012(0.2128)$ $\widehat{\alpha} = 0.1828(0.4250)$

Table 2.4: Kou Double-exponential Jump-diffusion model

0 strike Chan et al. (2009)	1 strike	2 strikes	4 strikes
$\widehat{\mu} = 0.0620(0.0296)$	$\widehat{\mu} = -3.5483 \times 10^{-10}(0.0179)$	$\widehat{\mu} = 0.0133(0.0030)$	$\widehat{\mu} = 0.0133(7.7640 \times 10^{-6})$
$\widehat{\sigma} = 0.1327(0.0036)$	$\widehat{\sigma} = 0.2058(0.0025)$	$\widehat{\sigma} = 0.2257(0.0019)$	$\widehat{\sigma} = 0.2340(6.2043 \times 10^{-4})$

Table 2.5: Empirical estimation for the S&P500 index between January 2, 1987 and September 29, 2008: Black-Scholes Model.

0 strike Chan et al. (2009)	1 strike	2 strikes	4 strikes
$\widehat{\mu} = 0.0800(0.0296)$	$\widehat{\mu} = -0.0763(0.0031)$	$\widehat{\mu} = -0.0731(0.0009)$	$\widehat{\mu} = -0.0755(0.0005)$
$\widehat{\sigma} = 0.12715(0.0137)$	$\widehat{\sigma} = 0.2101(0.0177)$	$\widehat{\sigma} = 0.2374(0.0080)$	$\widehat{\sigma} = 0.2379(0.0048)$
$\widehat{\lambda} = 1.24319(0.0049)$	$\widehat{\lambda} = 1.8688(0.0002)$	$\widehat{\lambda} = 1.8760(0.0011)$	$\widehat{\lambda} = 1.876(0.0001)$
$\widehat{\mu_J} = 0.0213(2.3519)$	$\widehat{\mu_J} = 0.0177(0.9440)$	$\widehat{\mu_J} = -0.0050(0.0622)$	$\widehat{\mu_J} = -0.0061(0.0283)$
$\widehat{\sigma_J} = 0.0267(2.5951)$	$\widehat{\sigma_J} = 0.0821(0.943)$	$\widehat{\sigma_J} = 0.0223(0.2606)$	$\widehat{\sigma_J} = 0.0226(0.1270)$
NA	$\widehat{\alpha} = 0.2860(2.2553)$	$\widehat{\alpha} = 0.2889(1.5988)$	$\widehat{\alpha} = 0.2934(1.3815)$

Table 2.6: Empirical estimation for the S&P500 index between January 2, 1987 and September 29, 2008: Black-Scholes with Merton Jumps Model.

0 strike Chan et al. (2009)	1 strike	2 strikes	4 strikes
$\widehat{\mu} = 0.0606(0.0279)$	$\widehat{\mu} = -0.0438(0.0173)$	$\widehat{\mu} = 0.0035(0.0034)$	$\widehat{\mu} = -0.0775(0.0014)$
$\widehat{\sigma} = 0.0924(0.0048)$	$\widehat{\sigma} = 0.09447(0.0043)$	$\widehat{\sigma} = 0.1956(0.0014)$	$\widehat{\sigma} = 0.2246(0.0014)$
$\widehat{\lambda} = 2.160(0.1067)$	$\widehat{\lambda} = 2.1493(0.0828)$	$\widehat{\lambda} = 2.1684(0.0371)$	$\widehat{\lambda} = 1.6002(0.0323)$
$\widehat{p} = 0.2715(0.0001)$	$\widehat{p} = 0.2408(0.0001)$	$\widehat{p} = 0.4576(0.0001)$	$\widehat{p} = 0.4633(0.0001)$
$\widehat{\eta_1} = 16.1389(0.2696)$	$\widehat{\eta_1} = 16.1391(0.2761)$	$\widehat{\eta_1} = 16.1434(0.0775)$	$\widehat{\eta_1} = 21.5180(0.0589)$
$\widehat{\eta_2} = 25.2204(1.6940)$	$\widehat{\eta_2} = 25.2201(1.5885)$	$\widehat{\eta_2} = 20.1736(1.5765)$	$\widehat{\eta_2} = 30.2644(1.3122)$
NA	$\widehat{\alpha} = 0.09093(7.9389)$	$\widehat{\alpha} = 0.2084(0.4855)$	$\widehat{\alpha} = 0.2329(0.1077)$

Table 2.7: Empirical estimation for the S&P500 index between January 2, 1987 and September 29, 2008: Kou Double Exponential Jump-Diffusion Model.

Chapter 3

A Unified Approach to Semiparametric Transformation Models and Accelerated Failure Time Model under General Biased Sampling Schemes

3.1 Introduction

Linear transformation models are a rich class of semiparametric regression models that are especially useful for the analysis of failure time data. They include the well-known proportional hazards model and proportional odds model as special cases (Clayton and Cuzick, 1985; Cuzick, 1988; Bickel, Klaassen, Ritov and Wellner, 1993; Cheng, Wei and Ying, 1995). Various inferential procedures have been proposed for the estimation of the regression parameters and the transformation function, including rank-based estimating equations, martingale estimating equations, and nonparametric maximum likelihood (Cheng et al., 1995; Chen, Jin and Ying, 2002; Zeng and Lin, 2007). These methods deal with data that are obtained via simple random sampling, in which case the sampling probability does not depend on the data. In many cases, either naturally or by design, data are not randomly sampled from the target population.

The accelerated failure time (AFT) model or accelerated life model relates the logarithm of the failure time linearly to the covariates; see Wei, Ying and Lin (1990), Kalbfleisch and Prentice (2002) and Cox and Oakes (1984). Under the AFT model, the effect of the covariates on the failure time is directly related to the acceleration or deceleration of time to failure. This feature facilitates an easy interpretation for clinicians and hence the model provides an alternative to the celebrated Cox (1972) proportional hazards model for the regression analysis of censored failure time data.

Several estimation and inferential procedures have been proposed for the estimation of the regression parameters including rank-based estimating equations and martingale estimating equations. Prentice (1978) proposed a score test for the marginal likelihood of generalised ranks, followed by Buckley and James (1979) who provided a modification of the least-squares estimator to incorporate censoring. The large-sample properties of the Buckley-James-type and rank estimators were then studied by Ritov (1990), Tsiatis (1990), Lai and Ying (1991a;b), Ying (1993) and Jin et al. (2003). Despite the theoretical advances, semiparametric methods for the accelerated failure time model have rarely been developed to deal with data that are not randomly sampled from the target population, either due to natural setting or trial design.

The purpose of this article is to propose a unified approach for dealing with many commonly encountered biased sampling schemes where the sampling probabilities are data dependent. This work extends Paik (2009) in which length-biased case is considered. The usefulness and generality of the proposed approach are seen from the fact that it not only handles all semiparametric linear transformation models and the accelerated failure time model but also covers such commonly encountered biased sampling schemes as length-biased sampling, left-truncation, the case-cohort design, as well as variants of the case-cohort design.

There is an extensive literature addressing various biased sampling schemes. Left truncation occurs naturally in astronomy on red shift (Segal, 1976) and in studies of HIV infection (Lagakos et al., 1988). It pertains to the existence of a second random variable, in addition the variable of interest, such that the observation is truncated if the latter falls below the former. In other words, left truncation arises when individuals come under observation only some known time after the time origin of the phenomenon under study. These data arise naturally from large-scale panel studies, when entry into the study depends on some event occurring before the event of interest. For left-truncated data, nonparametric estimators of the survivor function in the one-sample problem can be found in Turnbull (1976), Vardi (1982), Woodroofe (1985), Wang (1987), Tsai, Jewell and Wang (1987). Furthermore, Wang, Jewell and Tsai (1986), Keiding and Gill (1990) and Lai and Ying (1991a) derived large sample properties. For semiparametric regression models, see Bhattacharrya, Chernoff and Yang (1983), Tsui, Jewell and Wu (1988), Lai and Ying (1991b), Wang, Brookmeyer and Jewell (1993) and Gross (1996).

Inference on length-biased data has been discussed in studies of ecology (McFadden, 1962), electron tube life (Blumenthal, 1967), fiber length (Cox, 1969), as well as in shrub data (Muttlak and MacDonald, 1990) and economic duration data (Kiefer, 1988; Helsen and Schmittlein, 1993; de Uña Álvarez, 2004). Under the length biased sample, the density of the observed sample is the original density multiplied by the length. The problem that Cox (1969) studied is a situation in which assemblies of fibres is gripped at a sampling point with those fibres which are not gripped (not crossing the sampling point) are combed out. The fibres remaining are the ones in the sample. Each fibre, therefore, has a probability of selection proportional to its length. Data can arise in a sample of patients in hospitals - experimental units with longer stays have greater likelihoods of being sampled. The one-sample problem of estimating the survivor function has been explored in Vardi (1982, 1985), Bhattacharyya, Franklin and Richardson (1988), Jones (1991), Asgharian (2004), Assgharian, M'Lan and Wolfson (2002) and Asgharian and Wolfson (2005). In the context of regression analysis, Wang (1996) proposed inference for length-biased data using the Cox model with time-varying covariates but without censoring. More recently, Luo and Tsai (2009) proposed a pseudo-partial likelihood estimator for the Cox model and derived two nonparametric estimators; see also Huang and Qin (2011). Qin and Shen (2010) proposed estimating equations for the Cox model and Chen (2010) proposed inference for size-biased data using an accelerated failure time model. Shen, Ning and Qin (2009) extended a rank-based approach used by Cheng et al. (1995) to construct an unbiased estimating function for the parameters in an accelerated failure time model and linear transformation model.

The case cohort design was proposed by Prentice (1986) to save time and cost for large scale epidemiological studies. It is equivalent to collecting complete information for all failures and a simple random sample of the non-failures. These designs are useful in designing large cohort studies for rare diseases when certain covariate information is difficult or costly to obtain. Its basic large sample properties established in Self and Prentice (1988). Further developments can be found in Lin and Ying (1993), Chen and Lo (1999) and Chen (2001) among others. For the semiparametric linear transformation models, Kong, Cai and Sen (2004) extended the rank-based estimator of Cheng et al. (1995) to the case-cohort design, while Lu and Tsiatis (2006) extended the martingale estimating equations of Chen et al. (2002). Extensions of the classical case-cohort design to more complex sampling schemes can be found in Borgan et al. (2000), Kulich and Lin (2004), Breslow and Wellner (2007) and Samuelsen, Ånestad and Skrondal (2007).

In this chapter, we develop a unified approach to linear transformation models and the AFT model under a general formulation of biased sampling schemes. We show that our approach leads to estimators that are consistent and asymptotically normal and we provide simple consistent variance estimators. The generality and usefulness of our approach are demonstrated through four special cases of biased sampling schemes, namely left truncation, length-biased sampling, case-cohort design and generalized case cohort designs.

The rest of this chapter is structured as follows. Section 2 introduces notation and specifies the models. Sections 3 presents details on the weight function for each specific biased sampling scheme, the estimation procedure as well as the large sample properties of the estimators. The algorithm and implementation are also explained. Simulation results together with applications on shrub data and nickel refinery data are included in Sections 4 and 5, respectively, followed by a discussion in Section 6. All the technical proofs are presented in the Appendix.

3.2 Notation and Methodology

Throughout this paper, we use T to denote the failure time of interest, C the censoring time and $\boldsymbol{Z}$ the p-vector of covariates. Let $\widetilde{T} = \min\{T, C\}$ and $\Delta = I(T \leq C)$.

To introduce our biased sampling scheme, we first consider the situation of the usual random sampling from a population. Let $q_Z(t, \delta)$ $(t \geq 0, \delta \in \{0, 1\})$ denote the joint conditional density of $(\widetilde{T}, \Delta)$ given covariates $\boldsymbol{Z}$. Furthermore, let f_Z $(\bar{F}_Z)$ and g_Z $(\bar{G}_Z)$ denote the conditional density (survival) functions of T and C, respectively. Since T and C are assumed to be conditionally independent given Z, it follows that

$$q_Z(t, \delta) = \{f_Z(t)\bar{G}_Z(t)\}^{\delta}\{g_Z(t)\bar{F}_Z(t)\}^{1-\delta}, \qquad t \geq 0, \quad \delta \in \{0, 1\}.$$

Now suppose we have a biased sample from the population with biasing function

$w(t, \delta), t \geq 0, \delta \in \{0, 1\}$. Following Bickel et al. (1993, p. 86), the conditional joint density of $(\widetilde{T}, \Delta)$ given $\boldsymbol{Z}$ then becomes

$$\widetilde{q}_Z(t, \delta) = \frac{w(t, \delta) q_Z(t, \delta)}{\int w(s, 0) q_Z(s, 0) ds + \int w(s, 1) q_Z(s, 1) ds}. \tag{3.1}$$

Note that such a sampling scheme depends on the outcome variables $(\widetilde{T}, \Delta)$. Common examples include length-biased sampling with $w(t, \delta) = t$ (Vardi, 1982; Gill, Vardi and Wellner, 1988) and case-cohort sampling with $w(t, \delta) = \delta + (1 - \delta)p$ (Prentice, 1986), where $p \in (0, 1)$ is a constant. In addition, we would like to point out that our general approach also handles the situation in which the biasing function is allowed to depend on $\boldsymbol{Z}$ and other observed covariates.

Throughout the rest of the paper, we will suppress the subscript $\boldsymbol{Z}$ in q and $\widetilde{q}$ when no ambiguity arises. In the absence of sampling bias, according to the counting process theory

$$N(t) - \int_0^t I(\widetilde{T} \geq t) d\Lambda(t),$$

where $\Lambda(\cdot)$ is the corresponding cumulative hazard function for the counting process $N(t)$, is a martingale process. In particular, this process has zero mean, a key property that gives unbiased estimating equations. Under the biased-sampling scheme, however, it is no longer a zero mean process and proper adjustment needs to be made.

As we will see in Lemma 3.2.1, one such adjustment is to insert into the integrand of the compensator the following weight function

$$\omega(t, \widetilde{T}, \Delta) = \frac{q(T, \Delta)}{\widetilde{q}(T, \Delta)} \times \frac{\widetilde{q}(t, 1)}{q(t, 1)}, \tag{3.2}$$

which is a product of two terms,$q(T, \Delta)/\widetilde{q}(T, \Delta)$ and $\widetilde{q}(t, 1)/q(t, 1)$. These two terms can be viewed as the Radon-Nikodym derivatives between the true and the biased densities for the risk set and the counting process respectively. Since both the counting process and the risk set are observed under biased sampling scheme, whereas the hazard function corresponds to the true density, we have to convert both $dN(t)$ and the risk set $I(\widetilde{T} \geq t)$ by the corresponding Radon-Nikodym derivatives so that all the components in estimating equation (3.3) are evaluated under the same measure. Our method resembles the idea of risk-set re-sampling first investigated in Wang (1996) that corrects the bias resulting from biased sampling. Note that the weight function ω may depend on $\boldsymbol{Z}$.

With the argument discussed above, we arrive at the following lemma which helps us obtain unbiased estimating equations.

Lemma 3.2.1. *Under the biased sampling scheme, i.e.* $(\widetilde{T}, \Delta)$ *follows* $\widetilde{q}_Z$ *given by* (3.1) *and* $\omega(t, \widetilde{T}, \Delta)$ *is defined by* (3.2)*, we have*

$$E_Z[dN(t)] = E_Z[\omega(t, \widetilde{T}, \Delta)Y(t)\lambda(t)dt], \tag{3.3}$$

where E_Z *denotes the conditional expectation given* $\boldsymbol{Z}$*.*

A formal proof of (3.3) is given in the Appendix. In the following text, special examples that can aid understanding the generality and the scope of applicability of the new approach are provided.

3.3 Examples and Special Cases

Biased sampling appears in many applications, either naturally or by design. Here we present six special cases involving biased-sampling that can be dealt with by our proposed method to obtain explicit expressions for the weight functions.

3.3.1 Length-biased Sampling

Under the length-biased sampling, the density of $(\widetilde{T}, \Delta)$ can be expressed as

$$\widetilde{q}(t, \delta) \propto tq(t, \delta).$$

In this case, $\omega(t, \widetilde{T}, \Delta)$, the bias-adjustment weight function is, therefore, given by

$$\begin{aligned}\omega(t, \widetilde{T}, \Delta) &= \frac{q(\widetilde{T}, \Delta)\widetilde{q}(t,1)}{\widetilde{q}(\widetilde{T}, \Delta)q(t,1)} \\ &= \frac{t}{\widetilde{T}}.\end{aligned}$$

Note that $\widetilde{T}$ is the length of follow-up time. It is noteworthy to mention that the current set up is designed for handling censoring first and followed by length bias sampling. The reason is twofold: first, this setting occurs naturally when cross-sectional sampling (censoring) is done in which the probability for a sample to be selected is proportional to the follow-up period $\widetilde{T}$ instead of the event time T; second, there may exist a non-identifiablity issue if the biasing function depends on the unobserved T.

3.3.2 Left Truncation

Left truncation arises in situations in which individuals come under observation only when their survival times are beyond some pre-specified time points; see, for example, Kalbfleisch and Prentice (2002, p. 14). The risk set just prior to an event time does not include individuals whose left truncation times exceed the given event time. In this case, denoting by U the truncation variable, the biased joint conditional density of $(\widetilde{T}, \Delta)$ given U can be obtained by

$$\widetilde{q}(t, \delta) \propto I(U < t)q(t, \delta).$$

Writing $\widetilde{q}(t, \delta) = KI(U < t)q(t, \delta)$, where K is the normalization constant, it follows that

$$\frac{q(\widetilde{T}, \Delta)}{\widetilde{q}(\widetilde{T}, \Delta)} = \frac{1}{\kappa} I(U < \widetilde{T}) = \frac{1}{\kappa}$$

Note that $I(U < \widetilde{T}) = 1$ since, for every observation, $\widetilde{T} > U$ always holds. Furthermore,

$$\frac{q(t, \delta)}{\widetilde{q}(t, \delta)} = \frac{1}{\kappa} I(U < t).$$

Combining the two ratios yields

$$\omega(t, \widetilde{T}, \Delta) = I(U < t). \tag{3.4}$$

3.3.3 Case-cohort Design

Under the case-cohort design, complete covariate information is collected only on all cases ($\Delta = 1$) and a random subset of censored subjects ($\Delta = 0$). Suppose that the probability of selecting a censored individual into the sub-cohort is p, the weight function can be obtained, again, via considering the ratio between the biased and the unbiased conditional joint densities.

Since

$$\widetilde{q}(t, \delta) \propto q(t, \delta)\delta + pq(t, \delta)(1 - \delta) = q(t, \delta)\left[\delta + p(1 - \delta)\right],$$

we have $q(t, 1)/\widetilde{q}(t, 1) = K$ and

$$\begin{aligned} \frac{q(\widetilde{T}, \Delta)}{\widetilde{q}(\widetilde{T}, \Delta)} &= \frac{\kappa q(\widetilde{T}, \Delta)}{q(\widetilde{T}, \Delta)[\Delta + p(1 - \Delta)]} \\ &= \frac{\kappa}{\Delta + p(1 - \Delta)}, \end{aligned}$$

where κ is a normalization constant. This leads to the following weight function

$$\omega(t, \widetilde{T}, \Delta) = \frac{1}{\Delta + p(1 - \Delta)}.$$

Note that in our model, $(\tilde{T}_i, \Delta_i)$ refer to the samples selected in the subcohort, which is slightly different from the set up specified in Lu and Tsiatis (2006).

3.3.4 Case-cohort Sampling on a Length-biased Sample

Suppose that a case-cohort design is applied to length-biased data arising from a cross sectional study. As a result, the biasing function is proportion to $t[\delta + p(1 - \delta)]$ and

$$\widetilde{q}(t, \delta) \propto t[\delta + p(1 - \delta)]q(t, \delta).$$

The corresponding weight function is, therefore, given by

$$\omega(t, \widetilde{T}, \Delta) = \frac{t}{\widetilde{T}[\Delta + p(1-\Delta)]}.$$

3.3.5 Stratified Case-cohort Design

Borgan et al. (2000) and Kulich and Lin (2004) proposed a stratified case-cohort design, in which the probability of selecting a censored observation into the subcohort is dependent on $\boldsymbol{X}$, a vector of covariates that may or may not overlap with $\boldsymbol{Z}$. Let $p(\boldsymbol{X})$ denote this selection probability. Then, proceeding as in previous examples, we get

$$\widetilde{q}(t, \delta) \quad \propto \quad q(t, \delta)\delta + p(\boldsymbol{X})q(t, \delta)(1-\delta).$$

Hence $\omega(t, \widetilde{T}, \Delta) = [\Delta + p(\boldsymbol{X})(1-\Delta)]^{-1}$.

3.3.6 Generalized Case-cohort Design

We now propose a generalized case-cohort design that covers the sampling schemes discussed in Subsections 3.3.3 and 3.3.5 as special cases. Under this design, cases are sampled with the sampling probability $p_1(\widetilde{T}, \boldsymbol{X})$ whereas controls are sampled into the subcohort with the selection probability $p_2(\widetilde{T}, \boldsymbol{X})$. It should be noted that the sampling probabilities now depend on Δ , $\widetilde{T}$ and $\boldsymbol{X}$.

The joint density of $(\widetilde{T}, \Delta)$ can be shown to be

$$\widetilde{q}(t, \delta) \propto q(t, \delta)[p_1(t, \boldsymbol{X})\delta + (1-\delta)p_2(t, \boldsymbol{X})]$$

and the weight function thus becomes

$$\omega(t, \widetilde{T}, \Delta) = \Big[p_1(\widetilde{T}, \boldsymbol{X})\Delta + (1-\Delta)p_2(\widetilde{T}, \boldsymbol{X})\Big]^{-1}.$$

3.4 Semiparametric Transformation Models

3.4.1 Model Specifications

We assume that T satisfies the transformation model which is specified through

$$H(T) = -\boldsymbol{Z}'\boldsymbol{\beta} + \epsilon, \tag{3.5}$$

where $H(\cdot)$ is an unknown monotone increasing function, $\boldsymbol{\beta}$ a p-vector of regression coefficients and ϵ an error term with a known distribution. In particular, when ϵ is specified to follow the extreme value distribution, (3.5) becomes the Cox (1972) proportional hazards regression model; when ϵ follows the logistic distribution, it becomes the proportional odds model (Bennett, 1983). When the error distribution is also not specified, only the direction of $\boldsymbol{\beta}$ is identifiable and we refer to Han (1987), Sherman (1993) and Chen (2002) for details about parameter estimation.

3.4.2 Estimating Equations and Asymptotic Results

In this section, we first derive the estimating equations for $\boldsymbol{\beta}$ and $H(\cdot)$ and establish the usual asymptotic properties for the resulting estimators. We then present an algorithm and discuss the implementation of the estimation procedure.

Following the counting process notation commonly used in survival analysis, we let $Y(t) = I(\widetilde{T} \geq t)$ be the at-risk indicator and $N(t) = I(\widetilde{T} \leq t, \Delta = 1)$ be the counting process that jumps to 1 when a failure occurs. Hazard and cumulative hazard functions of ϵ, which are completely specified under model (3.5), are denoted by $\lambda(\cdot)$ and $\Lambda(\cdot)$, respectively.

Based on the argument presented earlier, under model (3.5), in the absence of sampling bias,

$$N(t) - \int_0^t Y(s) d\Lambda \left\{ \boldsymbol{Z}'\boldsymbol{\beta}_0 + H_0(s) \right\}$$

is a martingale process, where $\boldsymbol{\beta}_0$ and H_0 denote the true values of $\boldsymbol{\beta}$ and H. Under the biased-sampling scheme, however, it is no longer a zero mean process and the adjustment that involves the inclusion of the weight function needs to be made.

For truncation and case-cohort sampling, one may view this problem from missing data perspective in the following sense: Let $D = 1$ or 0 be the indicator of observing an individual or not. Then, with a slight abuse of notation for $\widetilde{T} = t$, we can write

$$P(\widetilde{T} = t, \delta | D = 1, \boldsymbol{Z}) \;=\; \frac{\{\pi(t,1) f_z(t) \bar{G}_z(t)\}^{\delta} \{\pi(t,0) \bar{F}_z(t) g_z(t)\}^{1-\delta}}{P(D = 1|\boldsymbol{Z})},$$

where $\pi(\widetilde{T}, \Delta) = P(D = 1|\widetilde{T}, \Delta)$ and
$P(D = 1|\boldsymbol{Z}) = \int_0^{\tau} \left[\pi(t,1) f_z(t) \bar{G}_z(t) + \pi(t,0) \bar{F}_z(t) g_z(t)\right] dt$. It follows that

$$E[\Delta I(\widetilde{T} = t)|\boldsymbol{Z}] \;=\; \lambda(\beta' \boldsymbol{Z} + H_0(t)) E\left[\frac{I(\widetilde{T} \geq t)\pi(t,1)}{\pi(\widetilde{T}, \Delta)}\middle| \boldsymbol{Z}\right]$$

and hence

$$E\left[\Delta dN(t) - \frac{\pi(t,1)}{\pi(\widetilde{T}, \Delta)} Y(t) d\Lambda(\beta' \boldsymbol{Z} + H_0(t)) \middle| \boldsymbol{Z}\right] = 0.$$

Equation (3.3) leads to the following:

$$\sum_{i=1}^{n} \left[dN_i(t) - \omega(t, \widetilde{T}_i, \Delta_i) Y_i(t) d\Lambda\left\{\boldsymbol{Z}_i' \boldsymbol{\beta} + H(t)\right\}\right] \;=\; 0 \quad 0 \leq t \leq \tau, \tag{3.6}$$

$$\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i \left[dN_i(t) - \omega(t, \widetilde{T}_i, \Delta_i) Y_i(t) d\Lambda\left\{\boldsymbol{Z}_i' \boldsymbol{\beta} + H(t)\right\}\right] \;=\; 0, \tag{3.7}$$

where H is a nondecreasing function satisfying $H(0) = -\infty$ and τ is a pre-specified constant such that $\Pr\{\widetilde{T} \geq \tau\} > 0$. They are analogous to the martingale estimating equations derived in Chen et al. (2002). Note that the condition on τ is common and is imposed to avoid possible tail instability with censored data.

For a fixed $\boldsymbol{\beta}$, equation (3.6) entails that H be uniquely defined and be a monotone increasing step function with jumps only at observed failure times $t_1, \ldots, t_K$ and $H(t) = -\infty$ for all $t < t_1$. Let $\widehat{H}(\boldsymbol{\beta}; \cdot)$ be the unique solution to (3.6). Thus, the resulting estimator of $\boldsymbol{\beta}_0$ satisfies $U(\boldsymbol{\beta}) = 0$, where

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i \left[dN_i(t) - \omega(t, \widetilde{T}_i, \delta_i) Y_i(t) d\Lambda \left\{ \boldsymbol{Z}_i' \boldsymbol{\beta} + \widehat{H}(t, \boldsymbol{\beta}) \right\} \right]. \tag{3.8}$$

We let $\widehat{\boldsymbol{\beta}}$ denote the solution to (3.8) that estimates $\boldsymbol{\beta}_0$. Thus $\widehat{H}(t, \widehat{\boldsymbol{\beta}})$ estimates $H_0(t)$. Numerical solutions to equations (3.20) and (3.21) may be obtained using iterative methods. More details on the implementation of the computational algorithm will be presented in Section 3.4.3.

Note that the expectations of (3.6) and (3.7) are zero. This unbiasedness is crucial for obtaining asymptotically unbiased estimators for $\boldsymbol{\beta}_0$ and H_0. However, due to the bias-adjustment weight $\omega(t, \widetilde{T}, \Delta)$ that appears in (3.20) and (3.21), the process

$$M(t) = N(t) - \int_0^t \omega(s, \widetilde{T}, \Delta) Y(s) d\Lambda\{\boldsymbol{Z}' \boldsymbol{\beta}_0 + H_0(s)\}, \tag{3.9}$$

is no longer a martingale but a mean zero process instead. For this reason, the martingale argument given by Chen et al. (2002) to derive large sample properties needs to be modified accordingly. To identify the limiting distributions of the estimators, we define the following terms:

$$\begin{aligned}
B_1(t) &= E[\omega(t, \widetilde{T}, \Delta) Y(t) \dot{\lambda}\{\boldsymbol{Z}' \boldsymbol{\beta}_0 + H_0(t)\}], \\
B_2(t) &= E[\omega(t, \widetilde{T}, \Delta) Y(t) \lambda\{\boldsymbol{Z}' \boldsymbol{\beta}_0 + H_0(t)\}], \\
B(t, s) &= \exp\left\{ \int_s^t B_2^{-1}(u) B_1(u) dH_0(u) \right\}, \\
B_1^Z(t) &= E[\boldsymbol{Z} \omega(t, \widetilde{T}, \Delta) Y(t) \dot{\lambda}\{\boldsymbol{Z}' \boldsymbol{\beta}_0 + H_0(t)\}], \\
B_2^Z(t) &= E[\boldsymbol{Z} \omega(t, \widetilde{T}, \Delta) Y(t) \lambda\{\boldsymbol{Z}' \boldsymbol{\beta}_0 + H_0(t)\}],
\end{aligned}$$

where $\dot{\lambda}$ denotes the first derivative of λ. These terms are similar to those defined in Chen et al. (2002) that are used to simplify the expression of the limiting distribution of $\widehat{\boldsymbol{\beta}}$. In addition, we define

$$\boldsymbol{z}(t) = \frac{1}{B_2(t)} \left[B_2^Z(t) + \int_t^{\tau} \{ B_1^Z(s) - \frac{B_2^Z(s) B_1(s)}{B_2(s)} \} B(s, t) dH_0(s) \right] \tag{3.10}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{*} &= E\left[\int_0^{\tau}\{\boldsymbol{Z}-\boldsymbol{z}(t)\}\boldsymbol{Z}'Y(t)\omega(t,\widetilde{T},\Delta)\dot{\lambda}\left\{\boldsymbol{Z}'\boldsymbol{\beta}_0+H_0(t)\right\}dH_0(t)\right],\\ \boldsymbol{\Sigma}^{*} &= E\left[\int_0^{\tau}\{\boldsymbol{Z}-\boldsymbol{z}(t)\}dM(t)\right]^{\otimes 2}.\end{aligned}$$

We need to impose the following regularity conditions:

A1 For any finite K, $\lambda(x)$ is strictly positive and $\dot{\lambda}(x)$ is bounded and continuously differentiable on $(-\infty, K)$;

A2 The covariate vector $\boldsymbol{Z}$ is bounded in the sense that $\Pr\{\|\boldsymbol{Z}\| < m\} = 1$ for some constant m;

A3 The true transformation function H_0 is continuously differentiable with a strictly positive derivative on $[0, \tau]$;

A4 $E[\int_0^{\tau} \boldsymbol{Z}\omega(t,\widetilde{T},\Delta)Y(t)d\Lambda\{\boldsymbol{Z}'\boldsymbol{\beta}+H(t)\}]^2 < \infty$.

A5 Both $\boldsymbol{\Sigma}^{*}$ and $\boldsymbol{\Sigma}_{*}$ are nonsingular.

Remark: Condition A1 is a mild condition and is satisfied for distributions of ϵ in commonly encountered transformation models. Condition A2 is imposed so that modern empirical process theory can be applied without modification. Condition A4 is a mild assumption on the weight function $\omega(t,\widetilde{T},\Delta)$. For various case-cohort sampling, left-truncation, this condition can be easily verified. We can also show that the condition holds also for the length-biased sampling setup. Condition A5 is necessary since otherwise the problem becomes singular. Nonsingularity assumption on $\boldsymbol{\Sigma}^{*}$ is very mild. In fact, it basically means that the covariate vector $\boldsymbol{Z}$ does not reside in a lower dimensional hyperplane. For $\boldsymbol{\Sigma}_{*}$, however, it is in general not trivial to verify the nonsingularity with a single simple-to-verify condition. However, we find that for specific families that are commonly used for the transformation models, namely, the proportional hazards model, the proportional odds model and the normal transformation model, we can show that $\mathbf{\Sigma}_*$ is nonsingular at $\boldsymbol{\beta}_0 = 0$ due to the strictly increasing property of the corresponding hazard rate functions.

Theorem 3.4.1. *Under Conditions A1 - A4, there exists a neighborhood of* $\boldsymbol{\beta}_0$ *within which* $\widehat{\boldsymbol{\beta}}$ *exists and is unique for all large* n*. Furthermore,* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \rightarrow_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_*^{-1}\mathbf{\Sigma}^*(\mathbf{\Sigma}_*^{-1})')$ *and* $\sqrt{n}(\widehat{H}(t, \widehat{\boldsymbol{\beta}}) - H_0(t))$ *converges weakly to a Gaussian process. Consistent estimators of* $\mathbf{\Sigma}_*$ *and* $\mathbf{\Sigma}^*$ *can be obtained by substituting* $\boldsymbol{\beta}_0$ *and* H_0 *by their estimators, i.e.*

$$\begin{aligned}
\widehat{\mathbf{\Sigma}}_* &= -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i - \widehat{\boldsymbol{z}}(t)\}\,\boldsymbol{Z}_i' Y_i(t)\omega(t, \widetilde{T}_i, \Delta_i)\dot{\lambda}\left\{\boldsymbol{Z}_i'\widehat{\boldsymbol{\beta}} + \widehat{H}_0(t)\right\} d\widehat{H}_0(t) \\
\widehat{\mathbf{\Sigma}}^* &= \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^{\tau}\{\boldsymbol{Z}_i - \widehat{\boldsymbol{z}}(t)\}\, d\widehat{M}_i(t)\right]^{\otimes 2},
\end{aligned}$$

where $\widehat{M}_i(t)$ *and* $\widehat{\boldsymbol{z}}(t)$ *are similarly defined as in* (3.9) *and* (3.10) *with* $\boldsymbol{\beta}_0$ *and* H_0 *replaced by their respective estimators.*

The proof of Theorem 3.4.1 will be given in the Appendix. The limiting covariance function of $\sqrt{n}(\widehat{H}_0 - H_0)$ can be obtained through the usual asymptotic expansions and can be estimated by the same plug-in method.

3.4.3 Algorithm and Implementation

The computational algorithm closely follows that of Chen et al. (2002). First we choose an initial value $\widehat{\boldsymbol{\beta}}^{(0)}$, which can be obtained, for example, by using the maximum partial likelihood estimator and assuming the Cox proportional hazards model. With $\widehat{\boldsymbol{\beta}}^{(0)}$ being fixed, we then obtain an estimate of $H(t_1)$, where t_1 is the first observed failure time, by solving:

$$\sum_{i=1}^{n}\omega(t_1, \widetilde{T}_i, \Delta_i)Y_i(t_1)\Lambda\left\{\boldsymbol{Z}_i'\widehat{\boldsymbol{\beta}}^{(0)} + H(t_1)\right\} = 1.$$

This step is straightforward (e.g. via the Newton-Raphson algorithm) since Λ is a strictly monotone increasing function. We then estimate $H(t_k)$ by solving successively, for $k = 2, \ldots, K$,

$$\sum_{i=1}^{n} \omega(t_k, \widetilde{T}_i, \Delta_i) Y_i(t_k) \left[\Lambda\{\boldsymbol{Z}_i' \widehat{\boldsymbol{\beta}}^{(0)} + H(t_k)\} - \Lambda\{\boldsymbol{Z}_i' \widehat{\boldsymbol{\beta}}^{(0)} + H(t_{k-1})\} \right] = 1. \qquad (3.11)$$

The monotonicity of $\widehat{H}(t)$ can be seen from (3.11) that in order for the right-hand side to be one $H(t_k) > H(t_{k-1})$ must hold since $\Lambda(\cdot)$ is a monotone increasing function while both $\omega(t, \widetilde{T}_i, \Delta_i)$ and $Y_i(t)$ are non-negative. Note also that

$$\sum_{i=1}^{n} \omega(t_k, \widetilde{T}_i \Delta_i) Y_i(t_k) d\Lambda(\boldsymbol{\beta}' \boldsymbol{Z}_i + H(t_k)) \approx \sum_{i=1}^{n} \omega(t_k, \widetilde{T}_i \Delta_i) Y_i(t_k) \lambda(\boldsymbol{\beta}' \boldsymbol{Z}_i + H(t_k-)) \Delta H(t_k),$$

which gives $\widehat{H}$ to be a monotone increasing function. Denote by $\widehat{H}$ the resulting estimate, we estimate $\boldsymbol{\beta}_0$ again by solving

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \boldsymbol{Z}_i \Big([N_i(t_k) - N_i(t_{k-1})] \\ -\omega(t_k, \widetilde{T}_i, \Delta_i) Y_i(t_k) \left[\Lambda\{\boldsymbol{Z}_i' \boldsymbol{\beta} + \widehat{H}(t_k)\} - \Lambda\{\boldsymbol{Z}_i' \boldsymbol{\beta} + \widehat{H}(t_{k-1})\} \right] \Big) = 0.$$

Recall that $t_0 < t_1$ and, therefore, $N_i(t_0) = 0$ for $i = 1, \ldots, n$. Suppose $\widehat{\boldsymbol{\beta}}^{(1)}$ is the new resulting estimate, we then substitute $\widehat{\boldsymbol{\beta}}^{(0)}$ by $\widehat{\boldsymbol{\beta}}^{(1)}$ and repeat the procedure described above until convergence. Our experience indicates that convergence is usually achieved in a small number of iterations.

3.4.4 Simulations

We first specify $q(\widetilde{T}, \Delta)$ from which initial data are generated. In each subsection we describe how we resampled data with a weight proportional to the weight function ω described in Section 3. The simulation results are tabulated in Tables 1-6.

Following Chen et al. (2002), we generated data from $H(T) = -\beta_1 Z_1 - \beta_2 Z_2 + \epsilon$, with the hazard function of ϵ, $\lambda(x) = \exp(x)/\{1 + r\exp(x)\}$, where $r = 0, 0.5, 1, 1.5$

and 2. For the transformation function H, we used $H(t) = \log(t)$ for $r = 0$ and $\log(r^{-1}e^t - r^{-1})$ for other r values. Note that $r = 0$ corresponds to the proportional hazards regression while $r = 1$ corresponds to the proportional odds regression.

Covariates Z_1 and Z_2 were generated from uniform $(0, 1)$ that are independent of each other. The parameters β_1 and β_2 were chosen to be -1.0 and 1.0. Two censoring proportions were used, namely 0.1 and 0.2 for the length-biased sampling as well as 0.8 and 0.9 for various case-cohort designs. The censoring time was generated by $e^{a+0.5U}$ where U was a uniform random variable and values of a were set to attain desired censoring proportions.

3.4.4.1 Length-biased Sampling

Given the data we generated by $q(\widetilde{T}, \Delta)$, units were resampled if $U_i \leqslant \widetilde{T}_i/\gamma$, where U_i's are from the uniform $(0, 1)$ distribution, and γ a constant larger than $\widetilde{T}_i$ for all $i = 1, ..., n$. Computation was conducted on the resampled individuals of sizes 50, 300, 500 and 1500; simulations were based on 1000 replications. Recall that the bias-adjustment weight function for length-biased sampling is $\omega(t, \widetilde{T}, \Delta) = t/\widetilde{T}$, estimating equations (3.6) and (3.7) become:

$$\sum_{i=1}^{n} \left[dN_i(t) - \frac{t}{\widetilde{T}_i} Y_i(t) d\Lambda \left\{ \boldsymbol{Z}_i' \boldsymbol{\beta} + H(t) \right\} \right] = 0, \quad 0 \leq t \leq \tau, \tag{3.12}$$

$$\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i \left[dN_i(t) - \frac{t}{\widetilde{T}_i} Y_i(t) d\Lambda \left\{ \boldsymbol{Z}_i' \boldsymbol{\beta} + H(t) \right\} \right] = 0. \tag{3.13}$$

Tables 3.1 to 3.4 summarize the simulation results. The simulation results indicate that the proposed method performs well in large samples. The parameter estimates have negligible bias, compared to standard deviations and to the biased estimates of the unadjusted method of Chen et al. (2002). The means of estimated variance are close to the empirical variance of the parameter estimates, and the 95% confidence intervals are close to nominal coverage.

We can see that for smaller sample sizes, the empirical coverage probabilities (ECP) obtained by using the unadjusted estimator of Chen et al. (2002) are very close to 1. Although the unadjusted estimator produces estimates with large biases, the estimates exhibit an inflation in their variances. This is due to the fact that the unadjusted estimator does not take the bias into account and thus overestimates the corresponding variance. Therefore, we include Table 3.4 for comparison.

3.4.4.2 Case-cohort Design

A full cohort of sample size 3000 was generated and then case-cohort samples were selected from each full cohort by selecting from cases with a probability of p such that about two thirds of the selected samples in the subcohort are controls. The average sample size for a subcohort is 1000. The parameters β_1 and β_2 were set to be -1.0 and 1.0 respectively with the censoring proportions 0.8 and 0.9. Simulations were based on 1000 replications. According to the derivations in Section 4.1, the weight function that corrects the bias is given by $\omega(t, \widetilde{T}, \Delta) = \frac{1}{\Delta + p(1-\Delta)}$. The resulting estimating equations are

$$\sum_{i=1}^{n} \left[dN_i(t) - \frac{Y_i(t) d\Lambda \left\{ \boldsymbol{Z}_i' \boldsymbol{\beta} + H(t) \right\}}{\Delta_i + p(1-\Delta_i)} \right] = 0, \quad 0 \leq t \leq \tau, \tag{3.14}$$

$$\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i \left[dN_i(t) - \frac{Y_i(t) d\Lambda \left\{ \boldsymbol{Z}_i' \boldsymbol{\beta} + H(t) \right\}}{\Delta_i + p(1-\Delta_i)} \right] = 0. \tag{3.15}$$

The performance of the proposed estimators under the case-cohort design is summarized in Table 3.5. The empirical biases were negligible and coverage probabilities were close to 0.95.

3.4.4.3 Stratified Case-cohort Design

A full cohort of sample size 3000 was generated and then case-cohort samples were selected from each full cohort by selecting from cases with a probability of $p_i =$

$1-\{1+\exp(1+Z_{1i})\}^{-1}$, and selecting among controls with a probability of $p_i = 1-\{1+\exp(-3+2Z_{1i})\}^{-1}$. The average sample size for a subcohort is 1000, with one third of the samples are cases. The parameters β_1 and β_2 were set to be -1.0 and 1.0 respectively with the censoring proportions 0.8 and 0.9. Simulations were based on 1000 replications. Since $\omega(t,\widetilde{T},\Delta) = [\Delta + p(\boldsymbol{X})(1-\Delta)]^{-1}$, the estimating equations take the same form as in the case-cohort example, but with p being replaced by $p(\boldsymbol{X})$.

We assessed the performance of the proposed estimators under the case-cohort design. Table 3.6 summarizes the performance of the estimators using the average bias, 95% coverage probability, and estimated variances. For the models, the empirical biases were negligible and coverage probabilities were close to 0.95. The estimated variances were close to the variance from the simulations.

3.4.4.4 Generalized Case-cohort Design

Table 3.7 reports the results of simulations for the generalized case-cohort design where the probability of selection in the weight function depends on the follow-up time. Similar to the stratified case-cohort design simulation, we first generated 3000 samples and then randomly chose, on average, 1000 subjects into the subcohort, using the selection probability $p(\widetilde{T}) = 1-\{1+\exp(1+\widetilde{T}\gamma)\}^{-1}$, where $\gamma = 1.2, 2$ for $p_1(\widetilde{T})$ and $p_2(\widetilde{T})$ respectively. We found that the estimates for $\boldsymbol{\beta}$ were essentially unbiased and the means of the estimated standard error are close to the empirical standard errors. The coverage probabilities were close to 0.95. Simulations were based on 1000 replications. Likewise, the derivation shown in Section 3.3.6 shows that $\omega(t,\widetilde{T},\Delta) = \left[p_1(\widetilde{T},\boldsymbol{X})\Delta + (1-\Delta)p_2(\widetilde{T},\boldsymbol{X})\right]^{-1}$, it follows that the estimating equations are

$$\sum_{i=1}^{n}\left[dN_i(t) - \frac{Y_i(t)d\Lambda\left\{\boldsymbol{Z}_i'\boldsymbol{\beta} + H(t)\right\}}{p_1(\widetilde{T}_i,\boldsymbol{X}_i)\Delta_i + (1-\Delta_i)p_2(\widetilde{T}_i,\boldsymbol{X}_i)}\right] = 0, \quad 0 \le t \le \tau, \tag{3.16}$$

$$\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{Z}_i\left[dN_i(t) - \frac{Y_i(t)d\Lambda\left\{\boldsymbol{Z}_i'\boldsymbol{\beta} + H(t)\right\}}{p_1(\widetilde{T}_i,\boldsymbol{X}_i)\Delta_i + (1-\Delta_i)p_2(\widetilde{T}_i,\boldsymbol{X}_i)}\right] = 0. \tag{3.17}$$

3.4.5 Real Data Examples

3.4.5.1 Application to Shrub Data

We applied our estimation procedure to the data on 46 shrubs used by Wang (1996), originally described in Muttlak and McDonald (1990, Table 3). Data were collected using a line-intercept sampling method for vegetation. Under the biological sampling technique, the probability a shrub was included in the sample was proportional to the width, where the width was defined to be the distance between tangents of the shrub that are parallel to the transect (Muttlak and McDonald, 1990). Two indicator covariates were used to denote the three groups of transects to which the shrubs belonged. In Wang (1996), the first covariate Z_1 was an indicator of whether the shrub belonged to transect I, and Z_2 corresponded to transect II.

For the analysis reported in Table 3.9, we defined Z_1 and Z_2 to be indicators that the shrub belonged to transect I and transect III, respectively, so that the second transect was the reference group. The recoding of the covariate was to ensure that numerically more stable estimates can be obtained compared with the counterparts estimated by using the third transect as a reference group. This is due to the fact that only six observations belonged to this category. Table 3.9 reports the fitted transformation models with $\lambda(x) = \exp(x)/\{1 + r\exp(x)\}$ for values of $r = 0$ (proportional hazards), 0.5 and 1 (proportional odds). The significant effect of β_1 does not change for different values of r. Qualitatively, the estimates for β_1 are significant and β_2 are not significant for all of the models that were fitted.

3.4.5.2 Application to Case-cohort Design - Welsh Nickel Refiners Study

Data from Appendix VIII of Breslow and Day (1987) contain complete records for 679 workers employed in a nickel refinery in South Wales before 1925. The follow-up through 1981 uncovered 56 deaths from cancer of the nasal sinus. Lin and Ying (1993) reanalyzed the mortality data on the nasal sinus cancer using the Cox model with (modified) case-cohort design. Previous studies found three significant risk factors which include AFE (age at first employment), YFE (year at first employment) and EXP (exposure level).

In Table 3.10, the first column presents the estimated parameter values obtained from the full cohort dataset via estimating equations (3.20) and (3.21). In this case, $p = 1$ for all observations. The estimates are comparable to Lin and Ying (1993). The second column displays the results from fitting the same model to data obtained from a randomly drawn, hypothetical subcohort. Such a subcohort contains all the observed failures and some censored subjects that make up two third of the size of the subcohort. We also performed an analysis on another hypothetical subcohort which was drawn from the generalized case-cohort sampling scheme discussed in Section 3.3.6. We used selection probability $p(\widetilde{T}) = 1 - \{1 + \exp(1 + \widetilde{T}\gamma)\}^{-1}$, where $\gamma = 0.012$ and 0.020 for $p_1(t)$ and $p_2(t)$, respectively. The estimated values of $\boldsymbol{\beta}$ and their standard deviations, which are summarized in the third column of Table 3.10, are consistent with the conclusion of Lin and Ying (1993). All of these studies indicate that the covariates $\log(\text{AFE} - 10)$ and $\log(\text{EXP} + 1)$ are statistically significant. Compared with the full-cohort study, the estimated standard deviation of $\widehat{\boldsymbol{\beta}}$ presented in the second and the third column of the table are slightly inflated. This is due to the fact that only a subset of the data is used for the estimation. The estimates obtained from this generalized case-cohort sampling scheme are closed to the corresponding values obtained by using a full cohort and Lin and Ying (1993). Under the generalized case-cohort setting, however, only 70% of the cases were included in the subcohort.

3.5 Accelerated Failure Time (AFT) Model

This section extends the result developed in Section 3.4 to accelerated failure time models that again provides a unified approach to analyse datasets collected under various biased sampling schemes. We show that our approach leads to estimators that are consistent and asymptotically normal and we provide consistent variance estimators. The rest of this section is organised as follows: We first specify the model setup, necessary notation as well as the estimating procedures. The asymptotic properties of the proposed estimator are studied and presented. Simulation studies under practical sample sizes and real data analyses are conducted to verify the performance of the proposed estimator. The corresponding results are reported in Table 3.7. All the technical details are relegated to Appendix.

3.5.1 Model Specification

The AFT model relates the logarithm of the failure time linearly to the concomitant covariates, say $\boldsymbol{Z} = (Z_1, \ldots, Z_p)$ in the following sense

$$\log T = -\boldsymbol{Z}'\boldsymbol{\beta} + \epsilon, \tag{3.18}$$

where $\boldsymbol{\beta}$ is a p-vector of regression coefficients and ϵ, in contrast to the transformation models, has a unknown distribution with mean 0. The parameter $\boldsymbol{\beta}$ appears to be easy to interpret for they directly refer to the level of $\log T$. The primary goal of this paper is to find semiparametric estimates of $\boldsymbol{\beta}$ under generalised biased sampling scheme. The data that are observed will consist of n iid random vectors

$$(\widetilde{T}_i, \Delta_i, \boldsymbol{Z}_i), \quad i = 1, \ldots, n.$$

For our purposes, it will be convenient to consider the counting process approach; see Gill (1980). Here and in the sequel, we shall use the notation $\widetilde{N}(t) = I(\widetilde{T} \leq t)\Delta$ to represent the counting process which jumps by one when a failure occurs. Hazard and cumulative hazard functions of T are denoted by $\widetilde{\lambda}(\cdot)$ and $\widetilde{\Lambda}(\cdot)$ respectively. $\widetilde{Y}(t) = I(\widetilde{T} \geq t)$ denotes the at-risk indicator. Moreover, it may be useful to re-express the counting process via the noise term ϵ, where $e(\boldsymbol{\beta}) = \log \widetilde{T} - \boldsymbol{Z}_i'\boldsymbol{\beta}$. Specifically, we denote $N(\boldsymbol{\beta}; t) = I(e(\boldsymbol{\beta}) \leq t)\Delta$, $Y(\boldsymbol{\beta}; u) = I(e(\boldsymbol{\beta}) \geq u)$. Hazard and cumulative hazard functions of ϵ which are unspecified under model (3.18) are denoted by $\lambda(\cdot)$ and $\Lambda(\cdot)$ respectively. Unless there exists ambiguity, the subscript $\boldsymbol{Z}$ that appear in q and $\widetilde{q}$ will be suppressed.

In full cohort design, the regression parameters can be estimated via the weighted log-rank estimating function (Ying 1993; Jin et al. 2003)

$$U_\phi(\boldsymbol{\beta}) = \sum_{i=1}^{n} \Delta_i \phi(\boldsymbol{\beta})[\boldsymbol{Z}_i - \bar{Z}(\boldsymbol{\beta}, u)] \quad \text{or} \tag{3.19a}$$

$$U_\phi(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int \phi(\boldsymbol{\beta}; u)[\boldsymbol{Z}_i - \bar{Z}(\boldsymbol{\beta}; u)] dN(\boldsymbol{\beta}; u), \tag{3.19b}$$

where ϕ is a (data-dependent) weight function, $\bar{Z}(\boldsymbol{\beta}; u) := S^{(1)}(\boldsymbol{\beta}; u)/S^{(0)}(\boldsymbol{\beta}; u)$ with $S^{(\kappa)}(\boldsymbol{\beta}; u) = \frac{1}{n}\sum_{j=1}^{n} Y_j(\boldsymbol{\beta}; u)\boldsymbol{Z}_j^{\kappa}, \kappa = 0, 1;$ $\boldsymbol{Z}^0 = \mathbf{1}$ and $\boldsymbol{Z}^1 = \boldsymbol{Z}$. The choices of $\phi(\boldsymbol{\beta}, u) = 1$ and $\phi(\boldsymbol{\beta}, u) = 0$ correspond to log-rank and Gehan statistics respectively.

Recent literature dedicated to the inference on $\boldsymbol{\beta}$ for AFT model under various biased sampling settings include Kong and Cai (2009), Chen (2009), Mandel and Ritov (2009) and Shen, Ning and Qin (2009). Kong and Cai (2009) proposed a statistical method for analysing case-cohort data with AFT model. They developed an estimation procedure based on the convergence result of Wei, Ying and Lin (1990) that gives the following estimating equation:

$$\sum_{i=1}^{n} \int \phi(\boldsymbol{\beta}; u)[Z_i - \widetilde{Z}(\boldsymbol{\beta}; u)] dN(\boldsymbol{\beta}; u),$$

where $\widetilde{Z}(\boldsymbol{\beta}; u) = \widetilde{S}^{(1)}(\boldsymbol{\beta}; u)/\widetilde{S}^{(0)}(\boldsymbol{\beta}; u)$ with $\widetilde{S}^{(\kappa)} = n^{-1}\sum_{j=1}^{n}\left[\Delta_j + (1 - \Delta_j)\xi_j/p\right]$. Notation ξ_j is a subcohort indicator and p denotes the selection probability of subcohort which converges to a constant between 0 and 1. Chen (2009), followed by Mandel and Ritov (2009), studied the AFT model for data that are size biased. Chen (2009) only considered the uncensored case and made use of the invariance principle (see Property 1 of Chen (2009), pp 150) to provide inference procedure for the regression covariates. Shen, Ning and Qin (2009) tackled the constraint not allowing data to be right censored under the length-biased setting. They proposed the following estimating equation:

$$\sum_{i=1}^{n} q(\boldsymbol{Z}_i)\Delta_i \boldsymbol{Z}_i \frac{\log \widetilde{T}_i - \boldsymbol{Z}_i'\boldsymbol{\beta}}{\hat{w}(\widetilde{T}_i)} = 0,$$

where $\hat{w}(\widetilde{T}_i)$ is a consistent estimator of $w(\widetilde{T}_i) = \int_0^{\widetilde{T}_i} \hat{G}(u)du$ with $\hat{G}$ the Kaplan-Meier estimator for the censoring variable C.

3.5.2 Estimation Procedure

To our best knowledge, there has not yet been any unified inference procedure for the regression parameter in the AFT model under general biased sampling schemes. For simplicity, we shall consider the case with $\phi(\boldsymbol{\beta}) = 1$. We will discuss a more general case in the discussion. The method that unifies special cases of general biased sampling schemes in which the sampling probability depends on the outcome variable $(\widetilde{T}, \Delta)$ is developed upon Lemma 3.2.1.

Due to (3.3), we can obtain the following two estimating equations,

$$\sum_{i=1}^{n} \left[d\widetilde{N}_i(t) - \omega(t, \widetilde{T}_i, \Delta_i, \boldsymbol{\beta})\widetilde{Y}_i(t)d\widetilde{\Lambda}(t) \right] = 0, \quad 0 \le t \le \tau, \tag{3.20}$$

$$\sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{Z}_i \left[d\widetilde{N}_i(t) - \omega(t, \widetilde{T}_i, \Delta_i, \boldsymbol{\beta})\widetilde{Y}_i(t)d\widetilde{\Lambda}(t) \right] = 0. \tag{3.21}$$

Since there is a one-to-one correspondence between the time variable T_i and the noise term ϵ as specified in the model (3.18), we can reexpress (3.20) in via N, Y and $\lambda(\cdot)$. Note that

$$N(\boldsymbol{\beta}_0; u) = \widetilde{N}(e^{u-\boldsymbol{Z}'\boldsymbol{\beta}_0}), Y(\boldsymbol{\beta}_0; u) = \widetilde{Y}(e^{u-\boldsymbol{Z}'\boldsymbol{\beta}_0})$$

and

$$\lambda(u)du = \widetilde{\lambda}\left(e^{u-\boldsymbol{Z}'\boldsymbol{\beta}_0}\right) e^{-u-\boldsymbol{Z}'\boldsymbol{\beta}_0}du,$$

it follows that, from (3.20)

$$\begin{aligned} 0 &= \sum_{i=1}^{n}\left[d\widetilde{N}_i(e^{t-\boldsymbol{\beta}'\boldsymbol{Z}_i}) - \omega(e^{t-\boldsymbol{Z}_i'\boldsymbol{\beta}}, \widetilde{T}_i, \Delta_i)\widetilde{Y}_i(e^{t-\boldsymbol{Z}_i'\boldsymbol{\beta}})\widetilde{\lambda}(e^{t-\boldsymbol{Z}_i'\boldsymbol{\beta}})d(e^{t-\boldsymbol{Z}_i'\boldsymbol{\beta}})\right] \\ &= \sum_{i=1}^{n}\left[dN_i(\boldsymbol{\beta};t) - \omega(t, \widetilde{T}_i, \Delta_i, \boldsymbol{\beta})Y_i(\boldsymbol{\beta};t)d\Lambda(t)\right], \end{aligned} \tag{3.22}$$

which in turns gives a nonparametric estimation of the cumulative hazard rate function of ϵ:

$$d\Lambda(t) = \frac{\sum_{i=1}^{n} dN_i(\boldsymbol{\beta};t)}{\sum_{i=1}^{n}\omega(t, \widetilde{T}_i, \Delta_i, \boldsymbol{\beta})Y_i(\boldsymbol{\beta};t)}. \tag{3.23}$$

Equation (3.21) can be similarly transformed into an expression that is comprised of N_i, Y_i and $\Lambda(\cdot)$. Plugging (3.23) into the transformed version of (3.21), we obtain

$$\begin{aligned} 0 &= \sum_{i=1}^{n}\int \boldsymbol{Z}_i\left[dN_i(\boldsymbol{\beta};t) - \omega(t, \widetilde{T}_i, \Delta_i, \boldsymbol{\beta})Y_i(\boldsymbol{\beta};t)\frac{\sum_{i=1}^{n} dN_i(\boldsymbol{\beta};t)}{\sum_{i=1}^{n}\omega(t, \widetilde{T}_i, \Delta_i, \boldsymbol{\beta})Y_i(\boldsymbol{\beta};t)}\right] \\ &:= \sum_{i=1}^{n}\int\left[\boldsymbol{Z}_i - \bar{Z}_\omega(t, \boldsymbol{\beta})\right]dN_i(\boldsymbol{\beta};t), \end{aligned} \tag{3.24}$$

where

$$\bar{Z}_\omega(\boldsymbol{\beta};u) = \frac{\sum_{j=1}^{n}\boldsymbol{Z}_j\omega(t, \widetilde{T}_j, \Delta_j, \boldsymbol{\beta})Y_j(\boldsymbol{\beta};u)}{\sum_{j=1}^{n}\omega(t, \widetilde{T}_j, \Delta_j, \boldsymbol{\beta})Y_j(\boldsymbol{\beta};u)}. \tag{3.25}$$

We denote the right hand side of (3.24) as $U(\boldsymbol{\beta})$. In fact, $U(\boldsymbol{\beta})$ can also be written as

$$U(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_i(\boldsymbol{Z}_i - \boldsymbol{Z}_j)\omega(e_j(\boldsymbol{\beta}), \widetilde{T}_j, \Delta_j)I(e_j(\boldsymbol{\beta}) \geq e_i(\boldsymbol{\beta})). \tag{3.26}$$

The unbiasedness property exhibited in (3.23) and (3.24) is important in obtaining asymptotically unbiased estimator for $\boldsymbol{\beta}_0$. Note that, due the fact that the weight function ω may contain $\widetilde{T}$, which is not $\mathcal{F}_t$-measurable, the process

$$M(\boldsymbol{\beta};t) = N(\boldsymbol{\beta};t) - \int_{-\infty}^{t}\omega(t, \widetilde{T}_i, \Delta_i, \boldsymbol{\beta})Y_i(\boldsymbol{\beta};u)d\Lambda(u) \tag{3.27}$$

is not a martingale but a mean-zero process instead.

We can handle various biasing sampling schemes detailed in Section 4.1. One point that we would like to point out is, instead of directly using the examples given in Section 3.4 to tackle different biased sampling settings, since in (3.23) and (3.24), the counting process $dN(\boldsymbol{\beta};t)$ is defined with respect to the noise term, the weight function should be adjusted accordingly. We will provide an example below to illustrate this subtlety.

In the length-biased sampling case, since the density of $(\widetilde{T},\Delta)$ is proportional to $tq(t,\delta)$, according to (3.2), the weight function is given by

$$\omega(t,\widetilde{T},\Delta)=\frac{q(\widetilde{T},\Delta)\widetilde{q}(t,1)}{\widetilde{q}(\widetilde{T},\Delta)q(t,1)}=\frac{t}{\widetilde{T}}.$$

As we transform the time scale from t to $\exp\{t-\boldsymbol{\beta}'\boldsymbol{Z}\}$ so that the counting process will be converted to be expressed with respect to the noise term ϵ, the weight function thus becomes

$$\omega(u,\widetilde{T},\Delta,\boldsymbol{\beta})=\frac{e^{u-\boldsymbol{Z}'\boldsymbol{\beta}}}{e^{\log\widetilde{T}-\boldsymbol{Z}'\boldsymbol{\beta}}}=e^{u-\log\widetilde{T}}.$$

It turns out that the estimating equation for $\boldsymbol{\beta}$ becomes

$$\begin{aligned}&\sum_{i=1}^{n}\int_{-\infty}^{\tau}\boldsymbol{Z}_i\left[dN_i(\boldsymbol{\beta};u)-e^{u-\log\widetilde{T}_i}Y_i(\boldsymbol{\beta};u)d\Lambda(u)\right]\\&=\sum_{i=1}^{n}\int_{-\infty}^{\tau}\left[\boldsymbol{Z}_i-\frac{\sum_{j=1}^{n}\boldsymbol{Z}_je^{u-\log\widetilde{T}_j}Y_j(\boldsymbol{\beta};u)}{\sum_{j=1}^{n}e^{u-\log\widetilde{T}_j}Y_j(\boldsymbol{\beta};u)}\right]dN_i(\boldsymbol{\beta};u).\end{aligned}$$

For those weight functions that does not involve time, they remain the same as described in Section 3.4 as there is no time-transformation involved. Furthermore, in our model, $(\widetilde{T}_i,\Delta_i)_{i=1,\ldots,n}$ refer to the samples selected in the subcohort, which is slight different from the set up formulated in Kong and Cai (2009).

3.5.3 Asymptotic Properties

Let $\widehat{\boldsymbol{\beta}}_n$ be the root of the estimating function $U(\boldsymbol{\beta})$.

Theorem 3.5.1. *Under the regularity conditions specified in the Appendix, $\widehat{\boldsymbol{\beta}}_n$ is consistent and*

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}),$$

where

$$\begin{aligned}
\boldsymbol{A}_n &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\tau}\left[\boldsymbol{Z}_i - \bar{Z}_\omega(\boldsymbol{\beta}_0; u)\right]^{\otimes 2}\left[\dot{\lambda}(u)/\lambda(u)\right] dN_i(\beta; u),\\
\boldsymbol{B}_n &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\tau}\left[\boldsymbol{Z}_i - \bar{Z}_\omega(\boldsymbol{\beta}_0; u)\right]^{\otimes 2} dN_i(\boldsymbol{\beta}; u),\\
\boldsymbol{A} &:= \lim_{n\to\infty} \boldsymbol{A}_n,\\
\boldsymbol{B} &:= \lim_{n\to\infty} \boldsymbol{B}_n
\end{aligned}$$

and $\dot{\lambda}(u) = d\lambda(u)/du$.

The proof of Theorem 3.5.1 will be given in the Appendix. It should be noted that, in general, because $U(\boldsymbol{\beta})$ is not a continuous function of $\boldsymbol{\beta}$, a unique solution to the estimating equation $U(\boldsymbol{\beta}) = 0$ may not always be plausible. As in Tsiatis (1990), one may define a solution of $\widehat{\boldsymbol{\beta}}_n$ as a zero-crossing of $U(\boldsymbol{\beta})$ such that $U(\widehat{\boldsymbol{\beta}}_n -)U(\widehat{\boldsymbol{\beta}}_n +) \leq 0$ or the miminiser of the Euclidean norm of $\|U(\boldsymbol{\beta})\|$ as in Wei, Ying and Lin (1990).

To apply the asymptotic results, one need to find a consistent estimator for $\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$. For $\boldsymbol{A}$, it involves some technical difficulty because of the unknown λ_ϵ. Tsiatis (1990) proposed to a smoothing kernel to estimate λ_ϵ. This approach can be rather unstable and need a considerably large sample size in order to yield reliable estimators. Other approaches may involve computer-intensive resampling to approximate the variance-covariance matrix; see Parzen, Wei and Ying (1994), Lin, Wei and Ying (1998) and Jin et al. (2003).

As suggested in Chen (2009), a less computer-intensive sample-based method can be used to directly estimate the variance of $n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ as discussed in Kalbfleisch and Prentice (2002); see Chen and Jewell (2001) for more justification of this method. The procedure first uses a recursive bisection algorithm to solve for $\boldsymbol{b}_j$ in $n^{1/2}U(\boldsymbol{b}_j) =$

$\boldsymbol{\alpha}_j$, for $j = 1, \ldots, p$, where $\boldsymbol{\alpha}_j$'s are the p-vectors such that $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ and $\boldsymbol{\alpha}^{\otimes 2} = \widehat{\boldsymbol{A}}$. Then a consistent variance estimator of $n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ is given by $(\boldsymbol{b}_1 - \widehat{\boldsymbol{\beta}}_n, \boldsymbol{b}_2 - \widehat{\boldsymbol{\beta}}_n, \ldots, \boldsymbol{b}_p - \widehat{\boldsymbol{\beta}}_n)^{\otimes 2}$.

3.5.4 Simulations

Simulation studies are conducted to assess the effectiveness of the proposed method on the AFT model. In our simulations, we considered the linear regression model (3.18) of $\log T = -\boldsymbol{\beta}'\boldsymbol{Z} + \epsilon$ where the random variable ϵ was assumed to follow a standard Normal distribution with the density function $(2\pi)^{-1/2}\exp\{-x^2/2\}$. Covariates Z_1 and Z_2 were generated from uniform (0,1) that are independent of each other. The parameters $\boldsymbol{\beta}_0 = (\beta_1, \beta_2)$ were chosen to be -1.0 and 1.0 respectively. The censoring time was generated by $e^{a+0.5U}$, where U is a uniform variable. Values of a were set to attain the desired censoring proportion.

In this paper, we will demonstrate simulations on three biased sampling designs, namely (i) length-biased sampling, (ii) case-cohort design and (iii) generalised case-cohort design.

For length-biased sampling, given the data generated by $q(\widetilde{T}, \Delta)$, we resampled those units with $U_i \leq \widetilde{T}_i/\gamma$, where U_i follows the uniform distribution and γ is constant which is larger than $\widetilde{T}_i$ for all $i = 1, \ldots, n$. Computation was conducted on the resampled individuals of sizes 50 and 200. Simulations were based on 500 replications.

For case-cohort design, a full cohort of sample size 3,000 was generated and then case-cohort samples were selected from each full cohort by selecting from cases with a probability of p such that about two thirds of the selected samples in the subcohort are controls. The average sample size of a subcohort is 1,000 with censoring rate 0.9, which mimics a rare-disease study. Simulations were based on 500 simulations.

For stratified case-cohort design, a full cohort of sample size 3,000 was generated and then case-cohort samples were selected from each full cohort by selecting from cases with a probability of $p_i = 1 - \{1 + \exp(1 + Z_{1i})\}^{-1}$ and controls with a probability of $p_i = 1 - \{1 + \exp\{-3 + 2Z_{1i}\}\}^{-1}$. The average size for a subcohort is 1000, with one third of samples are cases. The censoring rate is 0.9. 500 replications were created to assess the performance.

From Table 3.7, we can see that the estimates for $\boldsymbol{\beta}$ are essentially unbiased and the means of the estimated standard error are close to the empirical standard errors. The coverage probabilities are close to 0.95.

3.6 Discussion

We proposed an inferential procedure for the regression parameter and transformation function in linear transformation models and the accelerated failure time model under general biased sampling schemes. Our method unifies special cases of general biased sampling schemes in which the sampling probability depends on the outcome variable $(\widetilde{T}, \Delta)$ as in the case of the new variant of the case-cohort design.

A key ingredient in the proposed approach is the weight function which is used to make appropriate adjustment to obtain unbiased estimating equations. It is important to note that, for the method to work in practice, the weight function needs to have a manageable form. Fortunately, as demonstrated in the examples, for many important cases the weight functions are extremely simple and are readily available. This adds versatility and usefulness to the proposed method.

The proposed method reduces to various existing methods under special cases. For transformation models, in the case of length-biased data, with no censoring and $\Lambda(x) = \exp(x)$, the proposed estimator reduces to the limit of the estimator proposed by Wang (1996). When the data are assumed to be sampled via simple random sampling, the inferential procedure reduces to one that was proposed in Chen et al. (2002). For the classical case-cohort design, it is the same as Lu and Tsiatis (2006).

Zeng and Lin (2007) proposed using the nonparametric maximum likelihood estimation (NPMLE) for the family of the semiparametric transformation models. They showed that the NPMLE gives consistent and asymptotically efficient estimators. It is certainly desirable to see if the NPMLE can be used in the setting with biased sampling, so that efficient estimation can be achieved. Unfortunately, the approach does not seem to be directly applicable. A major difficulty appears to be in that the censoring distribution cannot be factored out. On the other hand, one may include a general weight function in the integrand of the estimating function to improve the efficiency. Such an improvement, however, is obtained at the cost of increasing computational complexity. The asymptotic variance of the corresponding efficient estimator does not generally have a closed form representation. A simple inference procedure is not readily available as a result.

As discussed in Ying (1993), for the censored linear regression, the rank method can be, unlike in the case with no censoring is present, more competitive since the computational and analytical edges offered by the least squares disappear. One can see a similar story in Chen (2009) and Mandel and Ritov (2009).

3.7 Appendix

This Section provides proofs of Lemma 3.2.1, Theorems 3.4.1 and 3.5.1. Note that when there is no ambiguity, E and P denote, respectively, the conditional expectation and probability given $\boldsymbol{Z}$.

A1. Proof of Lemma 3.1.

By definition, $N(t) = I(\widetilde{T} \leq t, \Delta = 1)$. Since $\widetilde{q}(t,1)$ is the sub-density of $\widetilde{T}$ on $\Delta = 1$, it follows that $E[dN(t)] = \widetilde{q}(t,1)dt$. Therefore, it suffices to show that $E[\omega(t, \widetilde{T}, \Delta)Y(t)\lambda(t)dt] = \widetilde{q}(t,1)dt$.

Recall that $\omega(t, \widetilde{T}, \Delta) = [q(\widetilde{T}, \Delta)\widetilde{q}(t, 1)]/[\widetilde{q}(\widetilde{T}, \Delta)q(t, 1)]$. We have

$$\begin{aligned}
& \frac{q(t,1)}{\widetilde{q}(t,1)} E[\omega(t, \widetilde{T}, \Delta)Y(t)] \\
= \; & E\left[\frac{q(\widetilde{T}, \Delta)}{\widetilde{q}(\widetilde{T}, \Delta)} I(\widetilde{T} \geq t)\right] \\
= \; & \int \frac{q(s,1)}{\widetilde{q}(s,1)} I(s \geq t)\widetilde{q}(s,1)ds + \int \frac{q(s,0)}{\widetilde{q}(s,0)} I(s \geq t)\widetilde{q}(s,0)ds \\
= \; & \int [q(s,1) + q(s,0)]\, I(s \geq t)ds \\
= \; & [1 - F(t)][1 - G(t)].
\end{aligned}$$

Hence, it follows that

$$\begin{aligned}
E[\omega(t, \widetilde{T}, \Delta)Y(t)]\lambda(t)dt &= [1 - F(t)][1 - G(t)]\frac{\widetilde{q}(t,1)}{q(t,1)} \frac{f(t)}{[1 - F(t)]} dt \\
&= \widetilde{q}(t,1)dt,
\end{aligned}$$

where the last equality follows from the fact that $q(t, 1) = f(t)\,[1 - G(t)]$.

A2. Proof of Theorem 3.4.1.

Following Chen et al. (2002), we divide the proof into three steps:

Step 1. Let $\widehat{H}_0(t) = \widehat{H}(t; \boldsymbol{\beta}_0)$, where $\boldsymbol{\beta}_0$ is the true parameter value. We first show that $\widehat{H}_0$ converges to H_0. Here, the proof follows closely the proof of Proposition in Lu and Ying (2004). Suppose $\widetilde{H}$ is a limit of $\widehat{H}_0$. By Helly's Lemma (van der Vaart, 2000), to show convergence of $\widehat{H}_0$ to H_0, it suffices to show that $\widetilde{H}$ must be H_0. By (3.20) and the law of large numbers, we have

$$E[N(t)] = \int_0^t E\left[Y(s)\omega(s, \widetilde{T}, \Delta)\right] \lambda\left\{\boldsymbol{Z}'\boldsymbol{\beta}_0 + \widetilde{H}(s)\right\} d\widetilde{H}(s).$$

This implies that $\widetilde{H}(\cdot)$ is differentiable and must satisfy

$$\frac{d\widetilde{H}(t)}{dt} = \frac{dE[N(t)]}{dt}\left(E\left[Y(t)\omega(t, \widetilde{T}, \Delta)\right] \lambda\left\{\boldsymbol{Z}'\boldsymbol{\beta}_0 + \widetilde{H}(t)\right\}\right)^{-1}, \qquad \text{(A1)}$$

which is a smooth function of t and $\widetilde{H}(t)$. Since (A1) is a Cauchy problem, its solution exists and is unique under local smoothness assumptions (Reinhard, 1987, Theorem

3.4.1). Note that by Lemma 3.1, H_0 satisfies (A1). Therefore, $\widetilde{H} = H_0$ and hence $\widehat{H}$ converges to H_0.

For t in a compact subset of the interior of the support of $\widetilde{T}$, we can show that the derivative of $\widehat{H}(t, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is bounded in the neighborhood of $\boldsymbol{\beta}_0$. Therefore, $\widehat{H}(t, \boldsymbol{\beta}_n) - \widehat{H}_0(t, \boldsymbol{\beta}_0) \to 0$ provided that $\boldsymbol{\beta}_n$ converges to $\boldsymbol{\beta}_0$. Since $\widehat{H}_0(t) \to H_0(t)$, it follows that $\widehat{H}(t, \widehat{\boldsymbol{\beta}}) \to H_0(t)$ provided that $\widehat{\boldsymbol{\beta}}$ is a consistent estimator.

We next show the consistency of $\widehat{\boldsymbol{\beta}}$. Let $\dot{U}(\boldsymbol{\beta})$ denote the derivative of $U(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. Applying the uniform law of large numbers (Pollard, 1990), we can show, for $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$, that $n^{-1}U(\boldsymbol{\beta})$ converges uniformly to a nonrandom limiting function $u(\boldsymbol{\beta})$ and that $n^{-1}\dot{U}(\boldsymbol{\beta})$ converges uniformly to $\dot{u}(\boldsymbol{\beta})$. Thus, $n^{-1}\dot{U}(\boldsymbol{\beta})$ is nonsingular in a neighborhood of $\boldsymbol{\beta}_0$, provided that $\dot{u}(\boldsymbol{\beta}_0) = -\Sigma_*$, which is to be shown in the next step. Since $u(\boldsymbol{\beta}_0) = 0$, it follows that there exists a neighborhood of β_0 such that $\widehat{\boldsymbol{\beta}}$ exists and is unique and that $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$.

Step 2. We next show that $n^{-1}\dot{U}(\beta_0)$ converges to $-\Sigma_*$. Let $a > 0$ and b be constants and define

$$\begin{aligned} \lambda^*\{H_0(t)\} &= B(t, a) \\ \Lambda^*(x) &= \int_b^x \lambda^*(s)ds, \end{aligned}$$

for $t > 0$ and $x \in (-\infty, \infty)$. Here, a and b are chosen such that the integrals are finite. By the definition of $B(t, s)$, we easily see that

$$\lambda^*\{H_0(s)\}/\lambda^*\{H_0(t)\} = B(s, t). \tag{A2}$$

Similarly, by the definition of $B_1(t)$, we can get

$$d\lambda^*\{H_0(t)\} = \lambda^*\{H_0(t)\}dB_1(t).$$

From these and mimicking Steps A2 and A3 of Chen et al. (2002, p. 666), we get

$$\frac{\partial}{\partial \boldsymbol{\beta}} d\widehat{H}(t, \boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = -B_2^{-1}(t)\left[B_1^Z(t) + B_1(t)\frac{\partial}{\partial \boldsymbol{\beta}}\widehat{H}(t, \boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + o_p(1)\right] dH_0(t), \tag{A3}$$

$$\left.\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{H}(t,\boldsymbol{\beta})\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = -\int_0^t \frac{B(s,t)}{B_2(s)} B_1^Z(s) dH_0(s) + o_p(1). \tag{A4}$$

Finally,

$$\begin{aligned} -\dot{U}(\beta_0) &= \sum_{i=1}^n \frac{\partial}{\partial\boldsymbol{\beta}} \int_0^\tau \boldsymbol{Z}_i \omega(t,\widetilde{T}_i,\Delta_i) Y_i(t) \lambda\{\boldsymbol{Z}_i'\boldsymbol{\beta} + \widehat{H}(t,\boldsymbol{\beta})\} d\widehat{H}(t,\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\ &= \sum_{i=1}^n \int_0^\tau \left[\boldsymbol{Z}_i - \frac{B_2^Z(t)}{B_2(t)}\right] \omega(t,\widetilde{T}_i,\Delta_i) Y_i(t) \dot{\lambda}\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + H_0(t)\} \boldsymbol{Z}_i' dH_0(t) \\ &\quad + n \int_0^\tau \left[B_1^Z - \frac{B_1(t) B_2^Z(t)}{B_2(t)}\right] \frac{\partial}{\partial\boldsymbol{\beta}} \widehat{H}(t,\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} dH_0(t) + o_p(1), \end{aligned}$$

where the last equality follows from (A3) and definitions of B_1, B_2, B_1^Z and B_2^Z. Combining this with (A4) and rearranging terms, we get

$$n^{-1}\dot{U}(\beta_0) = -n^{-1} \sum_{i=1}^n \int_0^\tau (\boldsymbol{Z}_i - \boldsymbol{z}(t))\, \boldsymbol{Z}_i' \omega(t,\widetilde{T}_i,\Delta_i) Y_i(t) \dot{\lambda}\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + H_0(t)\} dH_0(t) + o_p(1),$$

which converges to $-\boldsymbol{\Sigma}_*$.

Step 3. Finally, we show the asymptotic normality of $U(\boldsymbol{\beta}_0)$. Write

$$U(\boldsymbol{\beta}_0) = \sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i dM_i(t) - \sum_{i=1}^n \int_0^\tau \boldsymbol{Z}_i \omega(t,\widetilde{T}_i,\Delta_i) Y_i(t) d[\Lambda\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + \widehat{H}_0(t)\} - \Lambda\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + H_0(t)\}]. \tag{A5}$$

Again by following the derivation of Chen et al. (2002, p. 667), we can show that the last term in (A5) is equal to

$$\sum_{i=1}^n \int_0^\tau \left[\frac{B_2^Z(t)}{B_2(t)} + \frac{\lambda^*\{H_0(t)\}}{B_2(t)} \int_0^\tau \left(\frac{B_1^Z(s)}{\lambda^*\{H_0(s)\}} - \frac{B_1(s) B_2^Z(s)}{B_2(s)\lambda^*\{H_0(s)\}}\right) dH_0(s)\right] dM_i(t) + o_p(n^{1/2}).$$

Combining this with (A.2), (A.5) and the definition of $\boldsymbol{z}(t)$, we get

$$U(\boldsymbol{\beta}_0) = \sum_{i=1}^n \int_0^\tau \{\boldsymbol{Z}_i - \boldsymbol{z}(t)\}\, dM_i(t) + o_p(n^{1/2}),$$

which is a sum of independent zero-mean random vectors. Thus the classical central limit theorem implies that $n^{-1/2}U(\boldsymbol{\beta}_0)$ converges to $\mathcal{N}(0,\Sigma^*)$. From this and the result of Step 2, we have $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_*^{-1}\boldsymbol{\Sigma}^*(\boldsymbol{\Sigma}_*^{-1})')$. To show the weak convergence of $\sqrt{n}\{\widehat{H}(t,\widehat{\boldsymbol{\beta}}) - H_0(t)\}$, observe that

$$\begin{aligned}\sqrt{n}\{\widehat{H}(t,\widehat{\boldsymbol{\beta}}) - H_0(t)\} &= \sqrt{n}\{\widehat{H}(t,\widehat{\boldsymbol{\beta}}) - \widehat{H}(t,\boldsymbol{\beta}_0)\} - \sqrt{n}\{\widehat{H}(t,\boldsymbol{\beta}_0) - H_0(t)\} \\ &= \sqrt{n}\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{H}(t,\boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) - \sqrt{n}\{\widehat{H}(t,\boldsymbol{\beta}_0) - H_0(t)\} \\ &\quad + o_p(1). \end{aligned} \tag{A6}$$

By (A4), the first term on the right hand side of (A6) equals

$$\frac{1}{\sqrt{n}}A(t)\boldsymbol{\Sigma}_*^{-1}\sum_{i=1}^n\int_0^\tau\{\boldsymbol{Z}_i - \boldsymbol{z}(s)\}\,dM_i(s) + o_p(1), \tag{A7}$$

where $A(t) = -\int_0^t \frac{B(s,t)}{B_2(s)}B_1^Z(s)dH_0(s)$. To tackle the second term, observe that

$$\begin{aligned}\sum_{i=1}^n dM_i(t) &= \sum_{i=1}^n dN_i(t) - \sum_{i=1}^n \omega(t,\widetilde{T}_i,\Delta_i)Y_i(t)d\Lambda\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + H_0(t)\} \\ &= \sum_{i=1}^n \omega(t,\widetilde{T}_i,\Delta_i)Y_i(t)d\Lambda\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + \widehat{H}(t,\boldsymbol{\beta}_0)\} \\ &\quad - \sum_{i=1}^n \omega(t,\widetilde{T}_i,\Delta_i)Y_i(t)d\Lambda\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + H_0(t)\} \\ &= (1+o_p(1))\sum_{i=1}^n \omega(t,\widetilde{T}_i,\Delta_i)Y_i(t) \\ &\quad \times d\left(\lambda\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + H_0(t)\}\left[\widehat{H}(t,\boldsymbol{\beta}_0)H_0(t)\right]\right). \end{aligned} \tag{A8}$$

Let $J_n(t) = \frac{\sum_{i=1}^n \omega(t,\widetilde{T}_i,\Delta_i)Y_i(t)\dot{\lambda}\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + H_0(t)\}}{\sum_{i=1}^n \omega(t,\widetilde{T}_i,\Delta_i)Y_i(t)\lambda\{\boldsymbol{Z}_i'\boldsymbol{\beta}_0 + H_0(t)\}}$ and $J(t) = \lim_{n\to\infty} J_n(t)$. Then (A8) can be used to show that

$$\begin{aligned}&\sum_{i=1}^n \frac{dM_i(t)}{\sum_{j=1}^n \omega(t,\widetilde{T}_j,\Delta_j)Y_j(t)\lambda\{\boldsymbol{Z}_j'\boldsymbol{\beta}_0 + H_0(t)\}} \\ =\ &(1+o_p(1))\left\{J_n(t)\left[\widehat{H}(t,\boldsymbol{\beta}_0) - H_0(t)\right]dH_0(t) + d\left(\widehat{H}(t,\boldsymbol{\beta}_0) - H_0(t)\right)\right\}.\end{aligned}$$

Therefore,

$$\begin{aligned}e^{\int_0^t J(s)ds}\left(\widehat{H}(t,\boldsymbol{\beta}_0) - H_0(t)\right) &= \sum_{i=1}^n\int_0^t \frac{e^{\int_0^s J(u)du}dM_i(s)}{\sum_{j=1}^n \omega(s,\widetilde{T}_j,\Delta_j)Y_j(s)\lambda(\boldsymbol{Z}_j'\boldsymbol{\beta}_0 + H_0(s))} \\ &\quad + o_p(n^{-1/2}).\end{aligned}$$

It follows that

$$\widehat{H}(t, \boldsymbol{\beta}_0) - H_0(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{B(s,t)}{B_2(s)} dM_i(s) + o_p(n^{-1/2}). \tag{A9}$$

Combining (A8) and (A9), we have

$$\begin{aligned} \sqrt{n}\{\widehat{H}(t, \widehat{\boldsymbol{\beta}}) - H_0(t)\} = & \frac{1}{\sqrt{n}} A(t) \boldsymbol{\Sigma}_*^{-1} \sum_{i=1}^{n} \int_0^\tau \{\boldsymbol{Z}_i - \boldsymbol{z}(s)\} \, dM_i(s) \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \frac{B(s,t)}{B_2(s)} dM_i(s) + o_p(1). \end{aligned} \tag{A10}$$

Both the first and the second terms on the right hand side of (A10) are sums of independently and identically mean-zero terms. Observe that $| \int_0^t \omega(t, \widetilde{T}, \Delta) Y(t) d\Lambda\{\boldsymbol{Z}'\boldsymbol{\beta} + H_0(t)\}|$ is bounded above by $| \int_0^\tau \omega(t, \widetilde{T}, \Delta) d\Lambda\{\boldsymbol{Z}'\boldsymbol{\beta} + H_0(t)\}|$ and by Condition A4, this envelope function has a finite second moment. Since $B(s,t)/B_2(s)$ is bounded for all s and t, the second term on the right hand side of (A10) has also a finite second moment. By the multivariate central limit theorem, $\sqrt{n}\{\widehat{H}(t, \widehat{\boldsymbol{\beta}}) - H_0(t)\}$ converges in finite dimensional distribution to a mean-zero Gaussian process. Similar to Bilias et al. (1997), $\boldsymbol{z}(t)$ is of bounded variations, all the major terms on the right hand side of (A10) can be written as differences between two monotone functions in t. Since $M_i(t)$ is also a difference of two monotone functions in t, it follows that, due to the fact that monotone functions have pseudodimension one, $\sqrt{n}\{\widehat{H}(t, \widehat{\boldsymbol{\beta}}) - H_0(t)\}$ is manageable in the sense of Pollard (1990). As a result, we can claim that the process $\sqrt{n}\{\widehat{H}(t, \widehat{\boldsymbol{\beta}}) - H_0(t)\}$ is tight and hence converges weakly to a Gaussian process (Pollard, 1990).

A3. Proof of Theorem 3.5.1.

With reference to the proof Theorem 3.5.1. We assume the following regularity conditions that are similar to those in Ying (1993):

1. The covariates are uniformly bounded, and without loss of generality, we may assume that $\sup_i \|\boldsymbol{Z}_i\| \leq 1$.

2. The error density f_ϵ and its derivative f'_ϵ are bounded, satisfying that

$$\int (f'_\epsilon(t)/f(t))^2 f(t)dt < \infty.$$

3. The C_i have uniformly bounded densities g_i, that is, there exists B_c such that $|g_i(t)| \leq B_c$ for all t and i.

4. $\sup_i E|\min\{\epsilon_i, C_i\}|^{\theta_0}$, for some $\theta_0 > 0$.

5. $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$ are nonnegative definite.

Remark: In the case where $\omega(\cdot)$ is non-increasing with respect to t, for example, in the case-cohort or stratified case-cohort designs, Assumption 5 holds automatically as shown as follows: Following the notation used in Ying (1993), we let

$$\Gamma_{n,k}(t) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i^{\otimes k}\omega(t, \widetilde{T}_i, \Delta_i, \boldsymbol{\beta})\bar{G}(e^{t-\boldsymbol{Z}_i'\boldsymbol{\beta}}), \quad k = 0, 1, 2,$$

where $\boldsymbol{Z}_i^0 = \mathbf{1}$, $\boldsymbol{Z}_i^1 = \boldsymbol{Z}_i$ and $\boldsymbol{Z}_i^{\otimes 2} = \boldsymbol{Z}_i\boldsymbol{Z}_i'$, and $R_n(t) = \Gamma_{n,2}(t) - \Gamma_{n,1}(t)\Gamma'_{n,1}(t)/\Gamma_{n,0}(t)$. Since

$$R_n(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n}\boldsymbol{Z}_j\omega(t, \widetilde{T}_j, \Delta_j, \boldsymbol{\beta})\bar{G}(e^{t-\boldsymbol{Z}_j'\boldsymbol{\beta}})}{\sum_{j=1}^{n}\omega(t, \widetilde{T}_j, \Delta_j, \boldsymbol{\beta})\bar{G}(e^{t-\boldsymbol{Z}_j'\boldsymbol{\beta}})}\right]^{\otimes 2} \bar{G}(e^{t-\boldsymbol{Z}_j'\boldsymbol{\beta}}) \geq 0,$$

hence $R_n(t)$ are nonnegative definite. In addition, for $t_1 \leq t_2$,

$$\begin{aligned} R_n(t_1) &\geq \frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n}\boldsymbol{Z}_j\omega(t, \widetilde{T}_j, \Delta_j, \boldsymbol{\beta})\bar{G}(e^{t_1-\boldsymbol{Z}_j'\boldsymbol{\beta}})}{\sum_{j=1}^{n}\omega(t, \widetilde{T}_j, \Delta_j, \boldsymbol{\beta})\bar{G}(e^{t_1-\boldsymbol{Z}_j'\boldsymbol{\beta}})}\right]^{\otimes 2} \bar{G}(e^{t_1-\boldsymbol{Z}_j'\boldsymbol{\beta}}) \\ &\geq \frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n}\boldsymbol{Z}_j\omega(t, \widetilde{T}_j, \Delta_j, \boldsymbol{\beta})\bar{G}(e^{t_2-\boldsymbol{Z}_j'\boldsymbol{\beta}})}{\sum_{j=1}^{n}\omega(t, \widetilde{T}_j, \Delta_j, \boldsymbol{\beta})\bar{G}(e^{t_2-\boldsymbol{Z}_j'\boldsymbol{\beta}})}\right]^{\otimes 2} \bar{G}(e^{t_2-\boldsymbol{Z}_j'\boldsymbol{\beta}}) \\ &= R_n(t_2), \end{aligned}$$

which concludes that $R_n(t)$ is non-increasing. It follows that $\boldsymbol{A}_n$ is nonnegative-definite because

$$A_n = \int R_n\left[f'(t) + \frac{f^2(t)}{\bar{F}(t)}\right] = \int R_n(t)\frac{f^2(t)}{\bar{F}(t)}dt + \int f(t)d(-R_n(t)).$$

By Theorem 1 of Ying (1993), $U(\boldsymbol{\beta})$ is thus asymptotically linear in the following sense

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\leq d_n} \left\{ \frac{\|U(\boldsymbol{\beta}) - U(\boldsymbol{\beta}_0) - \boldsymbol{A}_n n(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|}{n^{1/2} + n\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|} \right\} = o_p(1)$$

as $d_n > 0$ and $d_n \xrightarrow{\mathcal{P}} 0$. Assume that all the eigenvalues of $\boldsymbol{A}_n$ are bounded away from zero for sufficiently large n. Then the eigenvalues of $\boldsymbol{B}_n$ will also be bounded away from zero. By Corollary 1 in Ying (1993), there exists a closed neighbourhood containing $\boldsymbol{\beta}_0$ as an interior point such that $\widehat{\boldsymbol{\beta}}_n$ is strongly consistent and $n^{1/2}\boldsymbol{A}_n^{-1/2}\boldsymbol{B}_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{I}_{p\times p})$. Furthermore, since $\boldsymbol{A}_n \xrightarrow{\mathcal{P}} \boldsymbol{A}$ and $\boldsymbol{B}_n \xrightarrow{\mathcal{P}} \boldsymbol{B}$, by Slusky Theorem, we conclude that

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}),$$

which concludes the proof.

Table 3.1: Estimates and standard errors for beta in transformation models with a sample size of 50

			β_1				β_2			
Estimator	CP	r	Bias	Var	$\widehat{\text{Var}}$	ECP	Bias	Var	$\widehat{\text{Var}}$	ECP
Proposed	0.10	0.0	−0.040	0.491	0.677	0.987	0.020	0.480	0.454	0.938
Chen			−0.388	0.462	0.645	0.777	0.367	0.487	0.630	0.815
Proposed	0.20		−0.040	0.512	0.842	0.994	0.049	0.544	0.535	0.946
Chen			−0.327	0.502	0.642	0.834	0.341	0.512	0.682	0.812
Proposed	0.10	0.5	0.010	0.841	1.013	0.975	−0.011	0.827	0.796	0.926
Chen			0.042	0.594	0.824	0.828	−0.036	0.635	0.811	0.867
Proposed	0.20		0.015	0.846	1.211	0.985	−0.064	0.834	0.919	0.956
Chen			0.058	0.624	0.814	0.864	−0.100	0.654	0.809	0.880
Proposed	0.10	1.0	0.097	1.397	1.348	0.945	−0.062	1.302	1.727	0.948
Chen			0.341	0.750	0.973	0.830	−0.294	0.797	0.993	0.855
Proposed	0.20		0.063	1.264	1.563	0.968	−0.025	1.279	1.832	0.970
Chen			0.278	0.757	0.982	0.854	−0.248	0.798	0.980	0.868
Proposed	0.10	1.5	0.167	1.806	2.177	0.929	−0.123	1.860	3.373	0.923
Chen			0.477	0.914	1.203	0.805	−0.439	0.959	1.211	0.856
Proposed	0.20		0.061	1.915	2.857	0.913	−0.002	1.867	5.064	0.944
Chen			0.393	0.904	1.236	0.830	−0.351	0.948	1.203	0.858
Proposed	0.10	2.0	0.289	2.494	4.611	0.904	−0.205	2.447	5.640	0.907
Chen			0.582	1.070	1.501	0.794	−0.574	1.116	1.424	0.850
Proposed	0.20		0.287	2.375	5.988	0.925	−0.182	2.387	2.644	0.919
Chen			0.596	1.060	1.362	0.829	−0.546	1.107	1.409	0.853

Note: Bias, Var, $\widehat{\text{Var}}$ and ECP are defined as the difference between the estimated and the true parameter values, the asymptotic variance estimated, the variance of the simulated estimated parameter values as well as the empirical coverage probability respectively.

Table 3.2: Estimates and standard errors for beta in transformation models with a sample size of 300

			β_1				β_2			
Estimator	CP	r	Bias	Var	$\widehat{\text{Var}}$	ECP	Bias	Var	$\widehat{\text{Var}}$	ECP
Proposed	0.10	0.0	−0.006	0.160	0.188	0.978	0.007	0.161	0.163	0.955
Chen			−0.163	0.206	0.212	0.857	0.163	0.207	0.213	0.866
Proposed	0.20		−0.003	0.183	0.239	0.978	−0.004	0.188	0.192	0.955
Chen			-0.129	0.218	0.230	0.894	0.121	0.219	0.236	0.901
Proposed	0.10	0.5	0.016	0.325	0.353	0.963	−0.011	0.324	0.330	0.959
Chen			0.066	0.282	0.292	0.937	−0.061	0.284	0.296	0.936
Proposed	0.20		−0.007	0.347	0.404	0.965	0.000	0.345	0.372	0.963
Chen			0.047	0.285	0.309	0.926	−0.053	0.287	0.308	0.931
Proposed	0.10	1.0	0.002	0.537	0.541	0.944	0.005	0.588	0.637	0.969
Chen			0.178	0.362	0.373	0.913	−0.165	0.364	0.375	0.915
Proposed	0.20		0.001	0.539	0.589	0.957	0.000	0.552	0.708	0.967
Chen			0.166	0.355	0.369	0.923	−0.159	0.358	0.380	0.912
Proposed	0.10	1.5	0.003	0.844	0.819	0.922	0.005	0.864	0.979	0.943
Chen			0.250	0.445	0.464	0.906	−0.231	0.449	0.482	0.897
Proposed	0.20		−0.003	0.826	0.862	0.926	0.012	0.818	0.887	0.955
Chen			0.222	0.430	0.452	0.901	−0.220	0.433	0.461	0.904
Proposed	0.10	2.0	0.062	1.182	1.427	0.916	0.031	1.393	2.223	0.913
Chen			0.304	0.524	0.539	0.900	−0.279	0.529	0.581	0.897
Proposed	0.20		−0.016	1.146	1.306	0.902	−0.052	1.276	2.064	0.919
Chen			0.265	0.505	0.524	0.900	−0.299	0.509	0.553	0.891

Note: Bias, Var, $\widehat{\text{Var}}$ and ECP are defined as the difference between the estimated and the true parameter values, the asymptotic variance estimated, the variance of the simulated estimated parameter values as well as the empirical coverage probability respectively.

Table 3.3: Estimates and Standard Errors for beta in Transformation Models with a Sample Size of 500

			β_1				β_2			
Estimator	CP	r	Bias	Var	$\widehat{\text{Var}}$	ECP	Bias	Var	$\widehat{\text{Var}}$	ECP
Proposed	0.10	0.0	0.003	0.126	0.144	0.973	0.002	0.124	0.125	0.953
Chen			−0.152	0.166	0.078	0.476	0.158	0.163	0.072	0.428
Proposed	0.20		0.002	0.140	0.182	0.987	0.004	0.136	0.148	0.971
Chen			−0.120	0.171	0.175	0.880	0.130	0.171	0.170	0.878
Proposed	0.10	0.5	0.016	0.254	0.272	0.967	−0.001	0.257	0.257	0.958
Chen			0.070	0.223	0.230	0.930	−0.057	0.223	0.232	0.940
Proposed	0.20		0.002	0.289	0.314	0.973	−0.001	0.261	0.289	0.976
Chen			0.057	0.224	0.231	0.943	−0.057	0.225	0.223	0.923
Proposed	0.10	1.0	0.018	0.422	0.415	0.939	0.007	0.428	0.496	0.965
Chen			0.182	0.286	0.289	0.906	−0.165	0.288	0.291	0.910
Proposed	0.20		0.014	0.442	0.454	0.959	−0.015	0.478	0.550	0.972
Chen			0.171	0.280	0.298	0.873	−0.174	0.282	0.292	0.893
Proposed	0.10	1.5	0.008	0.685	0.656	0.934	0.032	0.801	0.868	0.955
Chen			0.244	0.353	0.375	0.870	−0.224	0.356	0.374	0.893
Proposed	0.20		−0.001	0.672	0.642	0.935	−0.030	0.742	0.894	0.960
Chen			0.231	0.340	0.367	0.878	−0.246	0.342	0.353	0.886
Proposed	0.10	2.0	0.022	0.958	1.008	0.922	0.037	0.921	1.529	0.946
Chen			0.284	0.417	0.445	0.870	−0.267	0.419	0.429	0.903
Proposed	0.20		0.029	0.891	0.986	0.928	0.001	0.879	0.939	0.948
Chen			0.288	0.400	0.413	0.867	−0.272	0.402	0.408	0.892

Note: Bias, Var, $\widehat{\text{Var}}$ and ECP are defined as the difference between the estimated and the true parameter values, the asymptotic variance estimated, the variance of the simulated estimated parameter values as well as the empirical coverage probability respectively.

Table 3.4: Estimates and standard errors for beta in transformation models with an average sample size of 1500

			β_1				β_2			
Estimator	CP	r	Bias	Var	$\widehat{\text{Var}}$	ECP	Bias	Var	$\widehat{\text{Var}}$	ECP
Proposed	0.10	0.0	0.005	0.049	0.048	0.961	−0.001	0.052	0.053	0.951
Chen			0.010	0.073	0.287	0.505	0.004	0.073	0.289	0.493
Proposed	0.20		−0.013	0.092	0.094	0.958	−0.006	0.095	0.093	0.954
Chen			−0.058	0.090	0.215	0.690	0.056	0.090	0.215	0.691
Proposed	0.10	0.5	0.006	0.025	0.026	0.952	0.007	0.025	0.027	0.951
Chen			0.225	0.082	0.237	0.588	−0.225	0.083	0.236	0.581
Proposed	0.20		0.002	0.029	0.029	0.941	0.007	0.029	0.028	0.966
Chen			0.183	0.095	0.229	0.667	−0.181	0.095	0.229	0.670
Proposed	0.10	1.0	0.018	0.056	0.064	0.954	0.007	0.056	0.060	0.957
Chen			0.305	0.100	0.208	0.498	−0.306	0.100	0.204	0.488
Proposed	0.20		0.009	0.059	0.048	0.942	0.004	0.059	0.066	0.953
Chen			0.278	0.109	0.209	0.544	−0.285	0.109	0.209	0.532
Proposed	0.10	1.5	0.029	0.137	0.119	0.963	0.004	0.137	0.135	0.948
Chen			0.368	0.110	0.188	0.414	−0.369	0.111	0.201	0.406
Proposed	0.20		0.014	0.102	0.103	0.947	0.013	0.103	0.087	0.944
Chen			0.329	0.123	0.208	0.499	−0.344	0.123	0.206	0.462
Proposed	0.10	2.0	0.040	0.346	0.300	0.959	0.012	0.346	0.370	0.967
Chen			0.393	0.132	0.213	0.412	−0.382	0.132	0.215	0.410
Proposed	0.20		0.016	0.164	0.191	0.936	0.035	0.164	0.154	0.966
Chen			0.363	0.147	0.216	0.497	−0.377	0.147	0.205	0.468

Note: Bias, Var, $\widehat{\text{Var}}$ and ECP are defined as the difference between the estimated and the true parameter values, the asymptotic variance estimated, the variance of the simulated estimated parameter values as well as the empirical coverage probability respectively.

Table 3.5: Estimates and standard errors for beta in transformation models under case-cohort sampling scheme

		β_1				β_2			
CP	r	Bias	Var	$\widehat{\text{Var}}$	ECP	Bias	Var	$\widehat{\text{Var}}$	ECP
0.90	0.0	0.001	0.224	0.235	0.963	−0.004	0.231	0.235	0.958
0.80		0.001	0.155	0.162	0.962	0.000	0.156	0.162	0.958
0.90	0.5	0.001	0.231	0.241	0.959	−0.004	0.240	0.246	0.956
0.80		0.001	0.165	0.170	0.962	0.000	0.166	0.175	0.963
0.90	1.0	0.000	0.248	0.254	0.961	−0.001	0.256	0.263	0.959
0.80		0.001	0.174	0.178	0.958	0.000	0.176	0.187	0.964
0.90	1.5	0.003	0.252	0.257	0.958	−0.001	0.261	0.271	0.960
0.80		0.002	0.183	0.185	0.953	−0.001	0.185	0.200	0.969
0.90	2.0	0.003	0.257	0.260	0.954	−0.001	0.267	0.279	0.968
0.80		0.001	0.193	0.193	0.949	0.000	0.195	0.213	0.959

Note: Bias, Var, $\widehat{\text{Var}}$ and ECP are defined as the difference between the estimated and the true parameter values, the asymptotic variance estimated, the variance of the simulated estimated parameter values as well as the empirical coverage probability respectively.

Table 3.6: Estimates and standard errors for beta in transformation models under stratified case-cohort sampling scheme

		β_1				β_2			
CP	r	Bias	Var	$\widehat{\text{Var}}$	ECP	Bias	Var	$\widehat{\text{Var}}$	ECP
0.90	0.0	−0.014	0.263	0.267	0.956	0.002	0.261	0.265	0.956
0.80		0.013	0.212	0.213	0.947	0.005	0.204	0.210	0.961
0.90	0.5	−0.011	0.264	0.265	0.954	0.005	0.262	0.267	0.961
0.80		−0.005	0.243	0.222	0.936	−0.005	0.233	0.224	0.956
0.90	1.0	0.015	0.275	0.274	0.956	0.005	0.272	0.281	0.958
0.80		−0.012	0.240	0.234	0.946	0.003	0.226	0.241	0.968
0.90	1.5	−0.012	0.281	0.278	0.953	0.005	0.276	0.289	0.959
0.80		0.011	0.251	0.243	0.939	0.002	0.238	0.256	0.970
0.90	2.0	−0.013	0.287	0.281	0.945	0.005	0.284	0.297	0.964
0.80		0.013	0.264	0.252	0.936	0.003	0.249	0.272	0.973

Note: Bias, Var, $\widehat{\text{Var}}$ and ECP are defined as the difference between the estimated and the true parameter values, the asymptotic variance estimated, the variance of the simulated estimated parameter values as well as the empirical coverage probability respectively.

Table 3.7: Estimates and standard errors for beta in transformation models under generalized case-cohort sampling scheme

		β_1				β_2			
CP	r	Bias	Var	$\widehat{\text{Var}}$	ECP	Bias	Var	$\widehat{\text{Var}}$	ECP
0.90	0.0	0.007	0.260	0.262	0.947	0.005	0.259	0.263	0.952
0.80		0.008	0.206	0.209	0.954	−0.003	0.200	0.209	0.962
0.90	0.5	0.010	0.228	0.230	0.945	−0.004	0.230	0.235	0.956
0.80		0.028	0.212	0.213	0.953	−0.022	0.206	0.217	0.958
0.90	1.0	0.012	0.252	0.254	0.944	−0.006	0.257	0.264	0.960
0.80		0.044	0.187	0.189	0.947	−0.044	0.184	0.198	0.957
0.90	1.5	0.023	0.259	0.258	0.943	−0.016	0.265	0.273	0.956
0.80		0.062	0.202	0.202	0.935	−0.057	0.198	0.215	0.942
0.90	2.0	0.028	0.266	0.262	0.942	−0.018	0.272	0.282	0.962
0.80		0.081	0.237	0.232	0.930	−0.077	0.230	0.248	0.950

Note: Bias, Var, $\widehat{\text{Var}}$ and ECP are defined as the difference between the estimated and the true parameter values, the asymptotic variance estimated, the variance of the simulated estimated parameter values as well as the empirical coverage probability respectively.

Table 3.8: Estimates and standard errors for beta in accelerated failure time models under various sampling schemes

			β_1				β_2			
Type	CP	n	Bias	Var	$\widehat{\text{Var}}$	ECP	Bias	Var	$\widehat{\text{Var}}$	ECP
LB	0.10	100	0.012	0.317	0.406	0.842	0.016	0.313	0.398	0.854
		300	0.008	0.119	0.122	0.944	0.003	0.114	0.129	0.955
CC	0.90	100	0.028	0.364	0.420	0.892	0.017	0.365	0.464	0.836
		300	0.015	0.210	0.204	0.938	0.011	0.219	0.209	0.953
SSC	0.90	100	0.240	0.557	0.626	0.844	0.078	0.416	0.386	0.968
		300	0.131	0.379	0.430	0.886	0.012	0.221	0.206	0.948

Note: Bias, Var, $\widehat{\text{Var}}$ and ECP are defined as the difference between the estimated and the true parameter values, the asymptotic variance estimated, the variance of the simulated estimated parameter values as well as the empirical coverage probability respectively. LB, CC and SCC denote length-biased sampling, case-cohort sampling and stratified case-cohort sampling design respectively.

Table 3.9: Estimates and Standard Errors for in Shrub Data Set

	β_1			β_2		
r	Est	SE	P-value	Est	SE	P-value
0.0	0.7655	0.3387	0.0238	-0.0752	0.3273	0.8183
0.5	2.8583	1.0537	0.0067	0.7516	0.7531	0.3183
1.0	4.2925	2.1608	0.0470	1.0118	1.2033	0.4004

Table 3.10: Cox regression analysis of time from the first employment to the nasal sinus cancer death for the Welsh nickel refiner study

Parameter	Full-cohort	Case-cohort	Generalized Case-cohort
$\log(AFE - 10)$			
Est.	2.2091	1.8426	2.1804
S.E	0.4097	0.4405	0.4323
P-value	$3.48e-08$	$3.44e-05$	$4.57e-07$
$(YFE - 1915)/10$			
Est.	0.0768	0.4801	0.0963
S.E	0.2925	0.3824	0.3418
P-value	0.6036	0.209	0.7781
$(YFE - 1915)^2/100$			
Est.	-1.2951	-1.2025	-1.4334
S.E	0.5104	0.6846	0.5913
P-value	0.006	0.079	0.0153
$\log(EXP + 1)$			
Est.	0.7883	1.1610	0.7654
S.E	0.1629	0.1934	0.1838
P-value	$6.519e-07$	$1.94e-09$	$3.123e-05$

Chapter 4

Conclusion and Future Directions

4.1 Concluding Remarks

The preceding chapters cover several semiparametric inference problems on financial data and data collected under biased sampling schemes.

We first proposed the use of derivative prices in combination with the characteristic function as constraints in estimation via the empirical likelihood method. The consistency and the asymptotics of the estimates from the proposed method are proved. The inclusion of derivative prices, which are vanilla European calls with the same underlying asset, improves the estimation efficiency and provides a more responsive estimate to the model parameters. Some small scale scenario analyses have been conducted for the cases where there is a change point after which some model parameters' values have changed. Constraints due to the derivative prices included offers users a more realistic estimate whereas the methodology suggested by Chan et al. (2009) may not be sensitive enough to detect such a change. As mentioned in Chapter 2, the idea of using derivative prices as one of the estimating equations is not restricted only to call prices.

We also studied two semiparametric models for data that are collected from various biased sampling schemes. The unified approach proposed allows us to assemble many commonly encountered biased sampling designs under one framework. We also proposed new variants of case-cohort design that may be applicable in clinical trials. The estimating equations constructed rely heavily on the weight function which is able to make appropriate adjustment. As the derivations in Chapter 3 suggests, the weight functions for many important cases, enjoy a manageable form. Simulations, numerical studies include verify the consistency and the asymptotic properties of the estimators derived for the two models.

4.2 Future Directions

In Chapter 2, we first introduced a single-period model and then further extended it to a multiple-period model in order to make the inclusion of derivatives with multiple maturities possible. However, when we are dealing with derivatives with different maturities, one has to divide the data into different intervals so as to match the return periods with the maturities. Such a setup reduces the sample size for the constraints that are defined based on derivatives that have long maturities. Jing, Yuan and Zhou (2009) introduced a so-called jacklife empirical likelihood (JEL) method. The new method is extremely simple to use in practice. In particular, the JEL is shown to be very effective in handling one and two-sample U-statistics. For Lévy processes, due to the fact that the increments are independent, we can in fact express the derivative price with multiple-period maturities as a U-statistic. By leveraging the effective JEL method, we may obtain more efficient estimates. Furthermore, Chen, Peng and Yu (2012) studied an empirical likelihood approach for both parameter estimation and model specication testing based on the conditional characteristic function for processes with either continuous or discontinuous sample paths on Markov models. The corresponding extension that combines the conditional characteristic function and the derivative prices should be challenging yet worthwhile.

In connection to the possible extension of the semiparametric models for survival data elaborated in Chapter 3, one may investigate the extension to partially linear transformation models (Lu and Zhang 2010) and semiparametric accelerated failure time model with non-linear component (Xue et al. 2006). Another potential challenge is the inclusion of time-varying covariates in the transformation models or accelerated failure time models under general biased sampling, similar to that for traditional survival data in Cheng, Wei and Ying (1995). A significant amount of research effort would be necessary to establish the large sample properties and computational algorithms in all these aforementioned settings.

Bibliography

Aalen, O. O. (1975). Statistical inference for a family of counting processes. Ph.D. thesis, Univ. of California, Berkeley.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). Statistical models based on counting processes. New York: Springer.

Applebaum, D. (2004). Lévy Processes and Stochastic Calculus. Cambridge University Press.

Asgharian, B. (2004). Deposition of hygroscopic particles in the lung and its application to aerosol drug delivery. Respiratory drug delivery IX, Vol. 2, (Dalby, R. N., Byron, P. R., Peart, J., Suman, J. D., and Farr, S. J., editors).

Asgharian, M., M'Lan, C. E. and Wolfson, D. B. (2002). Length-biased sampling with right censoring: An unconditional approach. *J. Am. Statist. Assoc.*, **95**, 888–902.

Asgharian, M. and Wolfson, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *Ann. Statist.*, **33**, 2109–31.

Bhattacharya, P. K., Chernoff, H. and Yang, S. S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.*, **11**, 505–14.

Bhattacharyya, B. B., Franklin, L.A. and Richardson, G. D. (1988). A comparison of nonparametric unweighted and length biased density estimation of fibres. *Comm. Statist., Theory and Methods*, **17**, 3629–44.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. (1993). Efficient and adaptive estimation for semiparametric models. Baltimore, MD: Johns Hopkins University Press.

Blumenthal, S. (1967). Proportional sampling in life length studies. *Technometrics*, **9**, 205–18.

Borgan, O., Langholz, B., Samuelsen, S. O., Goldstein, L. and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Anal.*, **6**, 39–58.

Breslow N. E. and Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand. J. Stat*, **34**, 86–102.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–36.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–36.

Carr, P. and Madan, D. B. (1999). Option valuation using the fast Fourier transform. *J. Comput. Finance* **2**, 61–73.

Carrasco, M. and Florens, J. P. (2000). Generalization of GMM to a continuum of moment conditions. *Econ. Theory* **2**, 797–834.

Chan, N. H., Chen, S. X., Peng, L. and Yu, Cindy L. (2009). Empirical likelihood methods based on characteristic functions with applications to Lévy processes. *J. Am. Statist. Assoc.*, **88**, 1621–30.

Chen, K. (2001). Generalized case-cohort sampling. *J. R. Statist. Soc. B* **63**, 791–809.

Chen, Y. Q. (2009). Semiparametric regression in size-biased sampling. *Biometrics* **66**, 149–58.

Chen, Y. Q. and Jewell, S. P. (2001). On a general class of hazards regression models. *Biometrika* **88**, 678–702.

Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659–68.

Chen, K. and Lo, S.-H. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika* **86**, 755–64.

Chen, S. X., Peng, L. and Yu, C. (2012). Parameter estimation and model testing for continuous-time Markov processes via conditional characteristic functions. *Bernoulli*, to appear.

Chen, Y. Q. and Wang, M.-C. (2000). Analysis of accelerated hazards models. *J. Am. Statist. Assoc.*, **95**, 619–27.

CHENG, S. C., WEI, L. J. AND YING, Z. (1999). Analysis of transformation models with censored data. *Biometrika* **82**, 835–45.

CLAYTON, D. G. AND CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model. *J. R. Statist. Soc. A* **148**, 82–108.

CONT, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **1**, 223–36.

CONT, R. AND TANKOV, P. (2004). Financial modelling with jump processes. Chapman and Hall /CRC.

COX, D. R. (1969). Some Sampling Problems in Technology in *New Developments in Survey Sampling, eds. Johnson and Smith*. New York: Wiley.

COX, D. R. (1972). Regression models and life-tables (with Discussion). *J. R. Statist. Soc. B* **34**, 187–220.

COX, D. R. AND OAKES, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.

CUZICK, J. (1988). Rank regression. *Ann. Statist.*, **16**, 1369–89.

DE UÑA ÁLVAREZ, J. (2004). Nonparametric estimation under length-biased sampling and Type I censoring: a moment based approach. *Ann. Inst. Statist. Math.* **56**, 667–81.

FEUERVERGER, A. AND MCDUNNOUGH, P. (1981a). On the efficiency of empirical characteristic funtion procedures. *J. R. Statist. Soc. B*, **43**, 20–7.

FEUERVERGER, A. AND MCDUNNOUGH, P. (1981b). On some Fourier methods for inference. *J. Am. Statist. Assoc.*, **76**, 379–87.

FEUERVERGER, A. AND MUREIKA, R. (1977). The empirical characteristic function and its applications. *Ann. Statist.*, **5**, 88–97.

FLEMMING, T. R. AND HARRINGTON, D. P. (1991). *Counting processes and survival analysis*. New York: Wiley.

GILL, R. D. (1980). Censoring and stochastic integrals *Mathematical Center Tract 124. Mathematische Centrum. Amsterdam.*

GROSS, S. T. (1996). Weighted estimation in linear regression for truncated survival data *Scand. J. Stat.* **23**, 179–93.

Helsen, K. and Schmittlein, D. C. (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Science* **11**, 395–414.

Huang, C.-Y. and Qin, J. (2011). Nonparametric estimation for length-biased and right-censored data. *Biometrika* **98**, 177–86.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–53.

Jing, B.-Y., Yuan, J., and Zhou, W. (2009). Jacklife empirical likelihood. *J. Am. Statist. Assoc.*, **104**, 1224–32.

Jones, M. C. (1991). Kernel density estimation for length biased data. *Biometrika* **78**, 511–19.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data, 2nd ed.* New York: Wiley.

Keiding, N. and Gill, R. D. (1990). Random truncation models and Markov processes. *Ann. Statist.*, **18**, 582–602.

Kiefer, N. M. (1988). Economic duration data and hazard functions. *J. Econ. Lit.* **26**, 646–79.

Kim, J. P., Lu, W., Sit, T. and Ying, Z. (2012). A unified approach to semiparametric transformation models under generalized biased sampling schemes. Manuscript.

Kong, L. and Cai, J. (2009). Case-cohort analysis with accelerated failure time model. *Biometrics* **65**, 135–42.

Kong, L., Cai, J., and Sen, P. K. (2004). Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. *Biometrika* **91**, 305–19.

Kou, S. G. (2002). A jump diffusion model for option pricing. *Manag. Sci.* **48**, 1086-1101.

Kou, S. G., Petrella, G. and Wang, H. (2005). Pricing path-dependent options with jump risk via Laplace transforms. *Kyoto Economic Review* **74**, 1–23.

Kou, S. G. and Wang, H. (2004). Option pricing under a double exponential jump diffusion model. *Manag. Sci.* **50**, 1178–92.

Kulich M. and Lin, D. Y. (2004). Improving the efficiency of relative risk estimation in case-cohort studies. *J. Am. Statist. Assoc.*, **99**, 832–44.

Lagakos, S. W., Barraj, L. M. and De Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with applications to AIDS. *Biometrika* **75**, 515–23.

Lai, T. L. and Ying, Z. (1991a). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.* **19**, 1370–402.

Lai, T. L. and Ying, Z. (1991b). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531–56.

Lin, D. Y., Wei, L. J. and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika*, **85**, 605–18.

Lin, D. Y. and Ying, Z. (1993). Cox regression with missing covariates. *J. Am. Statist. Assoc.*, **88**, 1341–9.

Lu, W. and Tsiatis, A. A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika*, **93**, 207–14.

Lu, W. and Zhang, H. H. (2010). On Estimation of partially linear transformation models. *J. Am. Statist. Assoc.*, **105**, 683–91.

Lin, D. Y. and Ying, Z. (1993). Cox regression with missing covariates. *J. Am. Statist. Assoc.*, **88**, 1341–9.

Luo, X. and Tsai, W. Y. (2009). Nonparametric estimation for right-censored length-biased data: a pseudo-partial likelihood approach. *Biometrika*, **96**, 873–86.

Mandel, M. and Ritov, Y. (2009). The accelerated failure time model under biased sampling. *Biometrics* **66**, 1306–8.

McFadden, J. A. (1962). On the lengths of intervals in a stationary point process. *J. R. Statist. Soc. B* **24**, 364–82.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *J. Finan. Econ.* **3**, 125–44.

Muttlak, H. A. and McDonald, L. L. (1990). Ranked set sampling with size-biased probability of selection. *Biometrics* **46**, 435–46.

Naik, V. and Lee, M. (1990). General equilibrium pricing of options on the market portfolio with discontinuous returns. *Rev. Finan. Stud.* **3**, 493–521.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–49.

Owen, A. (2001). *Empirical likelihood.* Chapman and Hall /CRC.

Paik, J. (2009). Contributions to the analysis of survival data subject to nonstandard sampling schemes. Ph.D. thesis, Columbia Univ.

Parzen, M. I., Wei, L. J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–50.

Prentice R. L. (1994). Linear rank tests with right censored data. *Biometrika* **69**, 674–6.

Prentice, R. L. (1986). A Case-cohort design for epidemiological cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–25.

Prentice, R. L. (1986). Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics* **66**, 382–92.

Reid, N. (1994). A conversation with Sir David Cox. *Stat. Sci.* **9**, 439–55.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303–28.

Samuelsen, S. O., Ånestad, H., and Skrondal, A. (2007). Stratified case-cohort analysis of general cohort sampling designs. *Scand. J. Stat*, **34**, 103–19.

Schoutens, W. (2002). *Lévy processes in finance - Pricing financial derivatives.* New York: Wiley.

Self, S. G., and Prentice, R. L. (1990). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16**, 64–81.

Segal, I. E. (1976). Mathematical cosmology and extragalactic astronomy, Volume 68. Academic Press.

Shen, Y., Ning, J. and Qin, J. (2009). Analyzing length-biased data with semi-parametric transformation and accelerated failure time models. *J. Am. Statist. Assoc.* **104**, 1192–202.

Shen, Y., Ning, J. and Qin, J. (2012). Likelihood approaches for the invariant density ratio model with biased-sampling data. *Biometrika.* **99**, 363–78.

SHREVE, S. E. (2004). Stochastic calculus for finance II: Continuous-time models. New York:Springer.

STOKEY, N. L. AND LUCAS, R. E. (1989). Recursive methods in economics dynamics. Harvard University Press, Cambridge MA.

STUTZER, M. (1996). A simple nonparametric approach to derivative security valuation. *Biometrika.* **51**, 1633–52.

SUEISHI, N. (2005). Quasi-likelihood estimation of stable distributions by method of scoring. mimeo, Kyoto University.

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354–72.

TSAI, W.-Y., JEWELL, N. P. AND WANG, M.-C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74**, 883–6.

TSUI, K. L., JEWELL, N. P. AND WU, C. F. J. (1988). A nonparametric approach to the truncated regression problem. *J. Am. Statist. Assoc.* **83**, 785–92.

TURNBULL, B. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Statist. Soc. B* **38**, 290–5.

VARDI, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist* **10**, 616–20.

VARDI, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13**, 178–203.

WALKER, T. B. AND HALEY, M. R. (2010). Alternative tilts for nonparametric option pricing. *J. Futures Markets* **30**, 983–1006.

WANG, M.-C. (1987). Product limit estimates: A generalized maximum likelihood study. *Comm. Statist. A* **16**, 3117–32.

WANG, M.-C. (1996). Hazards regression analysis for length-biased data. *Biometrika* **83**, 343–54.

WEI, L. J., YING, Z. AND LIN, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–51.

WANG, M.-C., BROOKMEYER, R. AND JEWELL, N. P. (1993). Statistical models for prevalent cohort data. *Biometrics* **49**, 1–11.

Wang, M.-C., Jewell, N. P. and Tsai, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.* **14**, 1597–605.

Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–51.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163–77.

Xue, H., Lam, K. F., Cowling, B. J., and de Wolf, F. (2006). Semi-parametric accelerated failure time regression analysis with application to interval-censored HIV-AIDS data. *Stat. Med.* **25**, 3850–63.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76–99.

Zeng, D. and Lin, D.Y. (2007). Maximum likelihood estimation in semiparametric models with censored data (with discussion). *J. R. Statist. Soc. B* **69**, 507–64.