東海大學管理學院

統計研究所

--- 碩士論文 ---

時間至事件資料之多階段模型分析：應用於企業信用評等

Multiple-State Models for the Analysis

of Time-to-Event Data :

Application to Corporate Rating

指導教授：沈葆聖 博士

研究生：楊織蔚

中華民國一〇四年七月

謝　誌

首先，我要感謝我的指導老師 沈葆聖教授，感謝沈老師能夠在個人事務繁忙的情況下，還能抽空指導我論文，面對非本科生、基礎薄弱的我，沈老師還是很有耐心和熱心地傳授自身所學，讓我在短短兩年間成長非常多，從沈老師身上我看到也學習到了做學問的專注與熱誠，還有將傳授他人知識當成使命的責任心，衷心感謝沈老師的幫助與教導，讓我得以順利完成本篇碩士論文。

感謝 戴政教授和 林正祥教授在百忙之中能夠撥空擔任口試委員，除了給予我論文上的指導與建議，讓我的論文能夠更加完善之外，也分享了許多人生智慧，幫助我在未來的就業職涯上能更加順遂前進。

感謝東海統計系所有教授與助教們，因為有大家的教導與幫助，熱心的傳授知識，給予成長的養分，讓我在這兩年的學習之旅，收穫豐滿，很高興能在美麗的東海大學完成我的碩士學位，未來我將會繼續精進專業，努力不懈。

感謝碩士班的同學們，這兩年一路走來，大家一起同甘共苦，一起歡笑悲傷，創造了許多美好的回憶，很感謝大家在課業上的幫助，對於我在課業上的難題與疑惑，總是熱心的幫我解決，讓我能跟上課程進度，並順利通過每次考試與每科修課課程，完成學業。

最後，感謝我的家人，因為有家人的支持，讓我能心無旁騖的衝刺學業，無後顧之憂。我衷心感謝這兩年來，陪我一路走來以及曾幫助過我的所有人，感謝各位。

學生 楊織蔚 謹識於
東海大學統計研究所
中華民國 104 年 7 月

Multiple-State Models for the Analysis of Time-to-Event Data : Application to Corporate Rating

Director: Pao-sheng Shen

Student: Chih-wei Yang

2015.07

Department of Statistics
Tunghai University
Taichung, Taiwan 40704

Abstract

A multi-state model (MSM) is a model for a continuous time stochastic process allowing individuals to transit among a finite number of states. The MSM of disease have been of use in the past in designing and evaluating cancer screening programs. Recently MSM are also used to examines the determinants of different states of financial distress. In this article, using MSM, we investigate key factors that affect corporate rating transitions. The transition intensities are modeled using Cox Markov models (CMM)/Cox semi-Markov models (CSMM). Based on a data set of 123 firms in Taiwan over the period 1995 to 2010, we investigate key factors that affect the entry of four states of corporate rating: low risk, moderate risk, high risk and default.

Keywords: Financial distress; Survival analysis; Cox proportional model; transition probability.

Contents

Chapter 1

Introduction

Over the last decade there have been many cases of firm collapses, e.g. Enron, WorldCom, Tyco and HealthSouth. The collapses of firms often entails significant direct and indirect costs to many stockholders. It can be easy to blame a firm's collapse on the economic climate or rival firms, but the main responsibility for the firm's well-being lies most with the senior management team. Specifically, risk managers need to quantify their firm's risk positions and had to include a quantitative disclosure of market risks in their financial statements for the convenience of investors. Collapse of firm can be avoided if key risk factors can be identified. Financially distressed firms usually face a continuum of financial distress states before they go bankrupt. Multi-state models (MSM) of disease have been of use in the past in designing and evaluating cancer screening programs. Recently, MSM are also used to examines the determinants of different states of financial distress, e.g. Chancharat et al. (2010) examined the determinants of multiple states of financial distress by applying a competing-risks model. They investigated the effect of financial ratios, market-based variables and company-specific variables, including company age, size and squared size on three different states of corporate financial distress: active companies; distressed external administration companies; and distressed takeover, merger or acquisition companies. Since MSM incorporate multiple states of financial distress, they provide a wider range of distress scenarios and an opportunity to examine the effect of explanatory variables across the diverse states of distress. In this article, by applying Cox Markov models (CMM) and Cox semi-Markov models (CSMM), we examine the determinants of four different states of company ratings: low-risk, moderate-risk, high-risk and default. Based on a data set of 123 firms in Taiwan over the period 1995 to 2010, we investigate key factors that affect rating transitions from 'low-risk' to 'moderate-risk', 'moderate-risk' to 'high-risk', 'low-risk' to 'high-risk' and 'high-risk' to 'default'. In Section 2, we introduce MSM, CMM, CSMM, AMM and ASMM. In Section 3, we analyze the real data set using CMM and CSMM.

Chapter 2

Multi-state models

2.1 Transition Probabilities and Intensities

Multi-state models (MSM) are models for stochastic processes which occupy one of a set of discrete states at any time. The MSM are well adapted for modeling complex event histories. They are useful in describing a process in which an individual moves through a series of states in continuous time and can provide a better understanding of the process of the failure, i.e. a better knowledge of the evolution of the disease/depressed over time. Based on MSM, one may estimate progression rates, assess the effects of risk factors and survival rates. There exists an extensive literature on MSM. Main contributions include books by Andersen et al. (1999) and Hougaard (2000). Reviews on MSM can be found in the papers by Commenges, (1999), Hougaard (1999) and Meira-Machado et al. (2009).

The complexity of MSM depend on the number of states defined, as well as the transitions allowed among these states. The simplest form of MSM is the "two-state model", with only two states, e.g. "alive" and "dead" and a single transition between the two states. A more complex model, three-state model, originate from splitting the "Alive" state into two transient states. The MSM can be described using diagrams with rectangular boxes to represent possible states and arrows between the states denoting the allowed transitions. States can be either transient or absorbing as shown in Figure 1. A state is called an absorbing state (e.g. death) if no transitions can emerge from the state.

For individual k, a multi-state process is a stochastic process $X_k = \{X_k(t), t \in [0, \tau]\}$ with a finite state space $E = \{1, 2, \ldots, M\}$ and with right-continuous path: $X_k(t+) = X_k(t)$. For any t, the variable $X_k(t)$ has values in E, i.e. M states. Associated with X_k is a counting process $N_{ijk}(t)$, which denotes the number of direct transition from state i to j in the interval $[0, t]$, i.e. $N_{ijk}(t) = \#\{s \leq t : X_k(s-) = i, X_k(s) = j\}$, $i \neq j$. The other process is the indicator $Y_{ik}(t) = I_{[X_k(t-)=i]}$, which denotes whether the process is in state i just before time

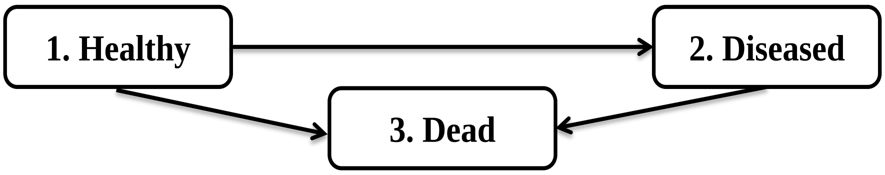

Figure 1. Illness-death model

t. Define the filtration or history process as

$$\mathcal{F}_t = \{N_{ijk}(t), Y_{ik}(u), 0 \leq u \leq t, k = 1, \ldots, n; i, j = 1, \ldots, M\}.$$

Notice that the history $\mathcal{F}_s$ of the process can also be generated by $\{X_k(u), u \leq s\ k = 1, \ldots, n\}$, i.e. $\mathcal{F}_s$ is an element of a 'filtration' and it can be understood intuitively as the trajectory of the process until time s. The law of multi-state processes can be specified by the transition probabilities

$$P_{ijk}(s, t, \mathcal{F}_s) = P(X_k(t) = j | X_k(s) = i, \mathcal{F}_{s-})\ (i = 1, \ldots, M; j = 1, \ldots, M).$$

The inference in multi-state models is traditionally performed under a Markov assumption for which past and future are independent given its present state (Aalen and Johansen (1978); Andersen and Keiding (2002)). Under Markov model, given the state at time s, the whole history before s can be forgotten:

$$P_{ijk}(s, t, \mathcal{F}_{s-}) = P(X_k(t) = j | X_k(s) = i) = P_{ijk}(s, t).$$

The observations are usually incomplete since

(1) we cannot observe the whole population of interest;

(2) we cannot observe the processes over an infinite time period.

Problem (1) can be resolved by drawing a sample from the population such that the sample is representative of the population. Problem (2) is called right censoring. The observation of the process $X_k(t)$ is stopped at a certain time C_k. If at that time, the process is in an absorbing state, e.g.'death', then the entire trajectory of the process has been observed. If not, then some transition times are right-censored.

Next, we consider transition intensities. Under a Markov assumption, if a randomly chosen individual k is in state i at time $t-$, the transition rate or intensity from i to j at time t is given by

$$d\Lambda_{ijk}(t) = P(X_k(t - +dt) = j | X_k(t-) = i) = P(X_k(t - +dt) = j | X_k(t-) = i),$$

which holds for all $X_k(u)$, $0 \leq u < t$ with $X_k(t-) = i$ and $i \neq j$. For convenience, define $d\Lambda_{iik}(t) = -\sum_{i \neq j} \Lambda_{ijk}(t)$ such that the row sums of the matrix $d\Lambda_{ijk}(t) = [d\Lambda_{ijk}]_{M \times M}$ are all equal to zero. For continuous case, we have $d\Lambda_{ij}(t) = \lambda_{ijk}(t)dt$ for all $i \neq j$, where $\lambda_{ijk}(t) = \lim_{h \to 0} h^{-1} P(X_k(t-+h) = j | X_k(t-) = i)$. Hence, $\lambda_{ijk}(t)$ is the intensity function for $i-to-j$ transition. Let $Y_{ik}(u) = I_{[X_k(u-)=i]}$. For $i \neq j; i, j \in E$, $M_{ijk}(t) = N_{ijk}(t) - \int_0^t Y_{ik}(u)\lambda_{ijk}(u)du$, are zero mean local square-integrable martingale with respect to $\mathcal{F}_t$. A state $h \in E$ is absorbing if for all $t \in [0, \tau]$, $j \neq h$, $\lambda_{hjk}(t) = 0$.

For homogenous population, $P_{ijk}(s,t) = P_{ij}(s,t)$ and $\lambda_{ijk}(t) = \lambda_{ij}(t)$ for all k. The transition probabilities can be estimated via the Aalen-Johansen (1978) estimator, which can be thought as the generalization of the Kaplan-Meier (1958) estimator for the simple mortality model (with states "alive" and "dead" and only one possible transition).

Next, we briefly describe nonparametric approach as follows. Let $\mathbf{I}$ be the identity matrix and $\boldsymbol{\Lambda}$ a matrix-valued function with element $\Lambda_{ij}(s) = \int_0^s \lambda_{ij}(u)du$, where $d\Lambda_{ii}(t) = -\sum_{j \neq i} d\Lambda_{ij}(t)$. In the discrete case, there exists a set of times $\{t_k : k = 1, 2, \dots\}$, at which transition can occur and $\mathbf{P}_k = \mathbf{I} + d\boldsymbol{\Lambda}(t_k)$ is the usual one-step probability transition matrix of a nonhomogeneous Markov chain with element $P(X(t_k) = j | X(t_k-) = i)$. Let $\mathbf{P}^{(r)}$ denote the r-step transition probability with element $P(X(t_r) = j | X(0) = i)$, $r = 1, 2, \dots,$. It is well known that

$$\mathbf{P}^{(r)} = \prod_{i=1}^{r} \mathbf{P_i} = \mathbf{P_1 P_2 ... P_r},$$

where an empty product is interpreted as $\mathbf{I}$. Notice that the order of the multiplication matters here since in general $\mathbf{P}_k$ matrices does not commute.

In the continuous case, $d\Lambda_{ij}(t) = \lambda_{ij}(t)dt$ for all i, j, where $\lambda_{ij}(t)$ is the intensity function for $i - to - j$ transition and $\lambda_{ii}(t) = -\sum_{j \neq i} \lambda_{ij}(t)$. Similar to discrete case, we can write

$$\mathbf{P}(s,t) = \prod\nolimits_{(s,t]} \big(\mathbf{I} + \boldsymbol{\Lambda}(du)\big),$$

where $\prod_{(s,t]}$ is the product integral over the interval $(s,t]$ and can be defined as the limit of a product, refining the partition $s < s_1 < \cdots < s_{p+1} = t$ of $(s,t]$:

$$\lim_{\max|s_l - s_{l-1}| \to 0} \prod_{l} \big(\mathbf{I} + \boldsymbol{\Lambda}(s_l-) - \boldsymbol{\Lambda}(s_{l-1})\big).$$

The transition probability matrix $\mathbf{P}$ can, for a Markov process, be recovered from the Kolmogorov forward equations:

$$\mathbf{P}(s,s) = \mathbf{I} \text{ and } \frac{\partial}{\partial t}\mathbf{P}(s,t) = \mathbf{P}(s,t)\lambda(t).$$

This can also be written as follows:

$$\mathbf{P}(s,t) = \mathbf{P}(s,s) + \int_{u \in (s,t]} \frac{\partial}{\partial u}\mathbf{P}(s,u)du = \mathbf{I} + \int_{u \in (s,t]} \frac{\partial}{\partial u}\mathbf{P}(s,u)du.$$

Since $\lambda(\mathbf{u})du = \mathbf{\Lambda}(du)$, it follows from Volterra's equation that the unique solution to the above equation is $\mathbf{P}(s,t)$. The Aalen-Johansen estimator of $\mathbf{P}(s,t)$ is obtained by plugging the matrix of Nelson-Aalen estimated matrix, i.e.

$$\hat{\mathbf{P}}(s,t) = \prod\nolimits_{(s,t]} \big(\mathbf{I} + \hat{\mathbf{\Lambda}}(du)\big),$$

where $\hat{\mathbf{\Lambda}}$ is the Nelson-Aalen matrix with element $\hat{\Lambda}_{ij}(t) = \sum_{s \le t} \hat{\Lambda}_{ij}(ds)$, where

$$\hat{\Lambda}_{ij}(t) = \int_0^t I_{[Y_i(u)>0]} \frac{dN_{ij}(u)}{Y_i(u)} = \sum_{t_l \le t} \frac{dN_{ij}(t_l)}{Y_i(t_l)},$$

where $N_{ijk}(t) = \sum_k N_{ijk}(t)$, $Y_{ik}(t) = \sum_k Y_{ik}(t)$. t_l's are the observed times.

A common model is the progressive three-state model (i.e. illness-death model) as shown in Figure 1. Explicit formulae of the Aalen-Johansen estimator for the illness-death model are as follows:

$$\hat{p}_{11}(s,t) = \prod_{s<t_{(k)} \le t} \left(1 - \frac{d_{12k} + d_{13k}}{n_{1k}}\right), \quad \hat{p}_{22}(s,t) = \prod_{s<t_{(k)} \le t} \left(1 - \frac{d_{23k}}{n_{2k}}\right),$$

$$\hat{p}_{12}(s,t) = \sum_{s \le t_{(k)} \le t} \hat{p}_{11}(s, t_{(k-1)}) \frac{d_{12k}}{n_{1k}} \hat{p}_{22}(t_{(k)}, t),$$

where $t_{(1)} < t_{(2)} < \cdots < t_{(d)}$ are the event times for transitions (e.g. disease/death) arranged in increased order, n_{1k} and n_{2k} denote the number of subjects at states 1 and 2, respectively, just prior to the event time $t_{(k)}$, and d_{ijk} is the number of transition $i \to j$ at time $t_{(k)}$. Notice that the estimator $\hat{p}_{12}(s,t)$ is a plug-in estimator obtained from the following expression:

$$p_{12}(s,t) = \int_s^t p(s,u)\lambda_{12}(u)p(u,t)dt,$$

by replacing $p_{11}(s,u) = p_{11}(s,u-)$ by $\hat{p}_{11}(s,u)$, $p_{22}(u,t)$ by $\hat{p}_{22}(u,t)$ and $\lambda_{12}(u)$ by $d\hat{\Lambda}_{12}(u)$ the increment of the Nelson-Aalen estimator $\hat{\Lambda}_{12}(u) = \sum_{t_{(k)} \le u} d_{12k}/n_{1k}$ of the cumulative disease intensity $\Lambda_{12}(t) = \int_0^t \lambda_{12}(u)du$.

2.2 Cox Markov Models (CMM)

Next, we consider heterogeneous population and for individual k there is a $p \times 1$ vector of possibly time-dependent covariates $Z_k(t) = [Z_{1k}, \ldots, Z_{pk}]^T$. One important goal in multi-state modeling is to relate the individual characteristics to the intensity rates through a possibly time-dependent covariate vector $Z_k(t)$. Several models have been used in literature. A common strategy is to decouple the whole process into various survival models, by fitting separate intensities to all permitted transitions based on some models while making appropriate adjustments to the risk set if necessary.

For individual k, let λ_{ijk} denote the intensity function for $i-to-j$ transition of individual k. Parametric or semiparametric models for λ_{ijk} can be specified, e.g. one may specify a parametric model depending on a vector of unknown parameters γ. Alternatively, one may consider the semiparametric model, e.g. Aalen's model (1980, 1989)) or Cox model (1972).

Andersen et al. (1991) developed the general theory of the "Cox Markov model" (CMM) where the intensities of the transitions from one state to the next are specified via Cox's (1972) proportional hazards regression models. Under CMM, the intensities depend only on time as measured from the origin (e.g., study entry) and not on the duration in a given state. Under CMM, given $Z_k(t)$, λ_{ijk} is written as

$$\lambda_{ijk}(t) = \lambda_{ij0}(t)\exp(Z_k(t)^T\beta_{ij}), \tag{2.1}$$

for all i, j, k with $i \neq j$ and $t > 0$, where $\lambda_{ij0}(t)$ is an unknown baseline intensity function and β_{ij} is $p\times 1$ vector of regression parameters for $i-to-j$ transition. The CMM readily fits into the multiplicative intensity framework of Cox model. Consider a right-censored sample of n individuals from model (2.1) and define the filtration or history process as

$$\mathcal{F}_t = \{N_{ijk}(t), Z_k(t), Y_{ik}(u), 0 \le u \le t, k = 1, \dots, n; i, j = 1, \dots, M\},$$

where $N_{ijk}(t)$ is the right continuous process that counts the number observed direct $i-to-j$ transition for individual k and $Y_{ik}(t)$ is the corresponding at risk process, i.e. the indicator of individual k being at risk in state i just before time t. Suppose that the censoring is independent such that for all $i \neq j$, k, $\mathcal{F}_{t-}$ and $t > 0$,

$$P(dN_{ijk}(t) = 1|\mathcal{F}_{t-}) = Y_{ik}(t)\lambda_{ijk}(t),$$

where $Y_{ik}(t) = I_{[X_k(t-)=i]}$. Let $M_{ijk}(t) = N_{ijk}(t) - \int_0^t Y_{ik}(s)\lambda_{ij0}(s)\exp(Z_k(s)^T\beta_{ij})ds$. Then, under model (2.1), $E[dM_{ijk}(t)|\mathcal{F}_{t-}] = 0$ and for $i \neq j; i, j \in E$, $M_{ijk}(t)$ are zero mean local square-integrable martingale with respect to $\mathcal{F}_t$.

Model (2.1) can be analyzed using partial likelihood arguments based on conditional probabilities of $dN_{ijk}(x)$, $k = 1, \dots, n$ given $\{\mathcal{F}_{t-}, dN_{ij}(t), i, j \in [0, \dots, M-1], i \neq j; t > 0\}$. Given a transition $i - to - j$ occurs at some $t \in [0, \tau]$, the contributing

$$P(dN_{ijk}(t) = 1|dN_{ij.}(t) = 1, \mathcal{F}_{t-}) = \frac{Y_{ik}(t)\exp(Z_k^T(t)\beta_{ij})}{\sum_{l=1}^n Y_{il}(t)\exp(Z_l^T\beta_{ij})}.$$

The log partial likelihood is given by

$$\sum_{all\ \ i,j}\Big\{\int_0^\tau \sum_{k=1}^n Z_k^T(t)\beta_{ij} - \log\big(\sum_{l=1}^n Y_{il}(x)\exp(Z_l^T(t)\beta_{ij})dN_{ij.}(t)\big)\Big\}. \tag{2.2}$$

The parameters β_{ij}'s can be estimated by maximizing (2.2). Let $\hat{\beta}_{ij}$ denote the estimator. Since $E[dM_{ijk}(t)|\mathcal{F}_{t-}] = 0$, we have $E[dM_{ij.}(t)|\mathcal{F}_{t-}] = 0$, i.e.

$$E[dN_{ij.}(t)|\mathcal{F}_{t-}] = \lambda_{ij0}(t) \sum_{k=1}^{n} Y_{ik}(t) \exp(Z_k^T(t)\beta_{ij}).$$

By letting $dN_{ij.}(t) - \lambda_{ij0}(t) \sum_{k=1}^{n} Y_{ik}(t) \exp(Z_k^T(t)\beta_{ij}) = 0$, we obtain

$$\lambda_{ij0}(t) = \frac{dN_{ij.}(t)}{\sum_{k=1}^{n} Y_{ik}(t) \exp(Z_k^T(t)\beta_{ij})}.$$

Given $\hat{\beta}_{ij}$, the baseline cumulative incidence function $\Lambda_{ij0}(t) = \int_0^t \lambda_{ij0}(u)du$ can be obtained using the Breslow estimator (1972,1974) $\hat{\Lambda}_{ij0}(\hat{\beta}_{ij}, x) = \int_0^x d\hat{\Lambda}_{ij0}(\hat{\beta}_{ij}, u)$, where

$$d\hat{\Lambda}_{ij0}(\hat{\beta}_{ij}, t) = \frac{dN_{ij.}(t)}{\sum_{k=1}^{n} Y_{ik}(t) \exp(Z_k^T(t)\hat{\beta}_{ij})},$$

$dN_{ij.}(t) = \sum_{k=1}^{n} dN_{ijk}(t)$. Furthermore, $\Lambda_{ij}(t|Z_k)$ can be estimated by

$$\hat{\Lambda}_{ij}(t|Z_k) = \hat{\Lambda}_{ij0}(\hat{\beta}_{ij}, t) \exp(Z_k^T(t)\hat{\beta}_{ij}).$$

Given Z_k, the transition probability matrix $\mathbf{P}$ can be estimated by

$$\hat{\mathbf{P}}(s, t|Z_k) = \prod\nolimits_{(s,t]} \big(\mathbf{I} + \hat{\mathbf{\Lambda}}(du|Z_k)\big),$$

where $\mathbf{\Lambda}(u|Z_k)$ is the estimated matrix with elements $\hat{\Lambda}_{ij}(u|Z_k)$.

The asymptotic properties of the estimators $\hat{\beta}$ and $\hat{\Lambda}$ was established by Shu et al. (2007).

Remark 1:

Notice that for three-state model, given individual k and two time points $s < t$, define the transition probabilities as $p_{ij}(s, t|Z_k) = P(X_k(t) = j|X_k(s) = i)$. There are five different transition probabilities: $p_{11}(s, t|Z_k)$, $p_{12}(s, t|Z_k)$, $p_{13}(s, t|Z_k)$, $p_{22}(s, t|Z_k)$ and $p_{23}(s, t|Z_k)$.

First, the transition intensities $\lambda_{ij}(t|Z_k)$, $1 \leq i < j \leq 3$ are modelled using Cox models assuming the process to be Markovian. Based on the estimated transition intensities, the transition probabilities $p_{ij}(s, t|Z_k)$ for a given covariate Z_k can be estimated by the so-called forward Kolmogorov differential equation (Cox and Miller (1965)) as follows:

$$\hat{p}_{11}(s, t|Z_k) = \prod_{s<u\leq t} \left(1 - \sum_{j=2}^{3} d\hat{\Lambda}_{1j}(u|Z_k)\right),$$

$$\hat{p}_{22}(s,t|Z_k) = \prod_{s<u\leq t}\Big(1-\hat{\Lambda}_{23}(u|Z_k)\Big),$$

and

$$\hat{p}_{12}(s,t|Z_k) = \sum_{u\leq t}\hat{p}_{11}(s,u-|Z_k)d\hat{\Lambda}_{12}(du|Z_k)\hat{p}_{22}(u+,t|Z_k),$$

where $\hat{\Lambda}_{ij}(t|Z_k) = \hat{\Lambda}_0(t)\exp(\hat{\beta}^T Z_k)$, where $\hat{\Lambda}_0(t)$ is the breslow estimator for $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. The transition probability $p_{13}(s;t|Z_k)$ can be estimated by $1-\hat{p}_{11}(s,t|Z_k)-\hat{p}_{12}(s,t|Z_k)$.

2.3 Cox Semi-Markov Models (CSMM)

In some instances, the Markovian assumption may not be satisfied for company rating data, e.g. the transition probability to state "high-risk' after being rated as "moderate-risk" depends on the time since rated as "moderate-risk", i.e. there exists a dependence on time since entry to a state. Such dependencies can be accommodated by considering Cox semi-Markov model (CSMM), where the future of the process does not depend on the current time but rather on the duration in the current state. Under CSMM, each time the firm enters a new state time is reset to 0. For a four-stage model, the differences between CMM and CSMM (Andersen et al. 2000) resides in transition $2 \to 3$ and $3 \to 4$, in which intensities λ_{23k} and λ_{34k} are modeled as

$$\lambda_{23}(t-T_{2,k}|Z_k) = \lambda_{230}(t-T_{2,k})\exp(Z_k(t)^T\beta_{23}), \tag{2.3}$$

and

$$\lambda_{34}(t-T_{3,k}|Z_k) = \lambda_{340}(t-T_{3,k})\exp(Z_k(t)^T\beta_{34}), \tag{2.4}$$

where $T_{2,k}$ is the entry time into state 2 for individual k and $T_{3,k}$ is the entry time into state 3 for individual k.

The Semi-Markov model does not readily fit into the multiplicative intensity framework because of its renewal nature. Thus, under model (2.3), we need to circumvent this difficulty by considering a time-shifted multivariate counting process observed over a fixed interval, say $[0,\tau]$, Let $\tilde{N}_{ijk}(x)$ denote the number observed direct $i-to-j$ transitions for individual k whose transition time from i to j (i.e., the amount of time spent in i before going to j) is less than or equal to x, here x represents duration scale in state i instead of the calendar time scale which starts at zero at state 1. Similarly, define $\tilde{Y}_{ik}(x)$ be the indicator of individual k being at risk in state i just before time x - a "local" time to state i. Voelkel and Crowley (1984) showed that such formulated counting processes $\tilde{N}_{ijk}(x)$ have intensity processes $\lambda_{ijk}(x)$ in the form of a multiplicative intensity model

$$P(d\tilde{N}_{23k}(t)=1|\mathcal{S}_{t-}) = \tilde{Y}_{2k}(t)\lambda_{23k}(t-T_{2k}) = \tilde{Y}_{2k}(t)\lambda_{230}(t-T_{2k})\exp(Z_k(t)^T\beta_{23}),$$

where $\mathcal{S}_t = \{\tilde{N}_{ijk}(t), Z_k(t), \tilde{Y}_{ik}(u), 0\leq u\leq t, k=1,\ldots,n; i,j=1,\ldots,M\}$.

Models (2.3) or (2.4) can be analyzed using partial likelihood arguments based on conditional probabilities of $d\tilde{N}_{ijk}(t)$, $k = 1, \ldots, n$ given $\{\mathcal{S}_{t-}, dN_{ij}(t), i, j \in [1, \ldots, M], i \neq j; t > 0\}$. Given a transition $i - to - j$ occurs at some $x \in [0, \tau]$, the contributing

$$P(d\tilde{N}_{ijk}(t) = 1 | d\tilde{N}_{ij.}(t) = 1, \mathcal{S}_{t-}) = \frac{\tilde{Y}_{ik}(t)\exp(Z_k^T(t)\beta_{ij})}{\sum_{l=1}^n \tilde{Y}_{il}(t)\exp(Z_l^T(t)\beta_{ij})}.$$

Let $\tilde{M}_{ijk}(t) = \tilde{N}_{ijk}(t) - \tilde{Y}_{ik}(t)\lambda_{ij0}(t)\exp(Z_k(t)^T\beta_{ij})$. Then, under models (2.3)/(2.4), $E[d\tilde{M}_{ijk}(t)|\mathcal{S}_{t-}] = 0$ and for $i \neq j; i, j \in E$, $\tilde{M}_{ijk}(t)$ are zero mean local square-integrable martingale with respect to $\mathcal{S}_t$. The log partial likelihood is given by

$$\sum_{all\ \ i,j} \left\{ \int_0^\tau \sum_{k=1}^n Z_k^T(t)\beta_{ij} - \log\Big(\sum_{l=1}^n \tilde{Y}_{il}(t)\exp(Z_l^T\beta_{ij})d\tilde{N}_{ij.}(t)\Big) \right\}. \qquad (2.5)$$

The parameters β_{ij}'s can be estimated by maximizing (2.5). Let $\tilde{\beta}_{ij}$ denote the estimator. Given $\tilde{\beta}_{ij}$, the baseline cumulative incidence function $\Lambda_{ij0}(t)$ can be obtained by

$$d\tilde{\Lambda}_{ij0}(\tilde{\beta}_{ij}, t) = \frac{d\tilde{N}_{ij.}(t)}{\sum_{k=1}^n \tilde{Y}_{ik}(t)\exp(Z_k^T(t)\tilde{\beta}_{ij})},$$

where $d\tilde{N}_{ij.}(t) = \sum_{k=1}^n d\tilde{N}_{ijk}(t)$.

2.4 Aalen Markov Models (AMM)

An alternative model to Cox model is the Aalen's model (Aalen (1980, 1989, 1993), McKeague (1988); and Huffer and McKeague (1991)), which allows for time-dependent regression coefficients $\beta_{ij}(t)$ with

$$\lambda_{ij}(t|Z_k) = \lambda_{ij0}(t) + \beta_{ij}(t)^T Z_k(t). \qquad (2.6)$$

The parameter $B_{ij}(t) = \int_0^t \beta_{ij}(s)ds$ can be estimated using the following arguments:

Let $M_{ijk}(t) = N_{ijk}(t) - \int_0^t Y_{ik}(s)[\lambda_{ij0}(s) + \beta_{ij}(s)^T Z_k(s)]ds$. Let $R_i(t) = [R_{i1}(t), \ldots, R_{in}(t)]^T$ be a $n \times (p+1)$ matrix, where

$$R_{ik}(t) = [Y_{ik}(t), Y_{ik}(t)Z_{1k}(t), \ldots, Y_{ik}(t)Z_{pk}(t)]^T$$

is a $(p+1) \times 1$ vector. Let $N_{ij}(t) = [N_{ij1}(t), \ldots, N_{ijn}(t)]^T$ be a $n \times 1$ vector. Let $M_{ij}(t) = N_{ij}(t) - \int_0^t R_i(u)\beta_{ij}(u)du$ is a $n \times 1$ vector. Thus, under (2.6) $E[dM_{ij}(t)|\mathcal{F}_{t-}] = 0$ for $i \neq j; i, j \in E$, $M_{ijk}(t)$ and $M_{ij}(t)$'s are zero mean local square-integrable martingale with respect to $\mathcal{F}_{t-}$. Based on the arguments above, the following ordinary least squared (OLS) estimating function can be used for the estimation of $B_{ij}(t)$ (Aalen (1980)):

$$\hat{B}_{ij}(t) = \sum_{x \leq t} R_i^-(x) N_{ij}(dx) = (\hat{B}_{ij0}(t), \hat{B}_{ij1}(t), \ldots, \hat{B}_{ijp}(t))^T,$$

where $N_{ij}(dx) = N_{ij}(x) - N_{ij}(x-)$ and $R_i^-(x) = [R_i(x)^T R_i(x)]^{-1} R_i(x)^T$ is a generalized inverse of $R_i(x)$. To obtain a more efficient estimator, Huffer and McKeague (1991) and McKeague (1988) considered a weighted least-squared generalized inverse

$$R_{Wij}^-(u) = [R_i(u)^T W_{ij}(u) R_i(u)]^{-1} R_i(u)^T W_{ij}(u),$$

where $W_{ij}(u)$ is an $n \times n$ diagonal matrix taken to have the (h,h)th element, $W_{ij,h}(u)$, proportional to the inverse of the variance of $dM_{ij}(t)$. A kernel-smoothed estimator of $\beta_{ij}(t)$ is needed for estimating the variance of $dM_{ij}(t)$, which is given by $\lambda_{ij}(t|Z_k) = Z_k(t)^T \beta_{ij}(t)$.

2.5 Aalen Semi-Markov Models (ASMM)

Similar to CSMM, for a four-stage model, the differences between AMM and ASMM resides in transition $2 \to 3$ and $3 \to 4$, in which intensities λ_{23k} and λ_{34k} are modeled as

$$\lambda_{23}(t - T_{12,k}|Z_k) = \lambda_{230}(t - T_{12,k}) + \beta_{23}(t - T_{12,k})^T Z_k(t), \tag{2.7}$$

and

$$\lambda_{34}(t - T_{23,k}|Z_k) = \lambda_{340}(t - T_{23,k}) + \beta_{34}(t - T_{23,k})^T Z_k(t). \tag{2.8}$$

Similar to CSMM, models (2.7) or (2.8) can be analyzed using process $\tilde{N}ijk(t)$ and $\tilde{Y}_{ik}(t)$.

Chapter 3

Corporate Rating Data

3.1 Analysis Under CMM

We consider the four-state model depicted in Figure 2. We assume that all companies are in State 1 (low risk) at time $t = 0$, and that they may either visit State 2 (moderate risk) at some time point; or not, going directly to the State 3 (high risk); or not, going to directly to the State 4 (default). The reverse transition can occur.

The data set consists of 123 semiconductor firms which were evaluated between 1995 to 2010. Thirty-six covariates were considered for inclusion in the model by using forward selection methods from SAS software packages.

The p-values for entry of covariates were set at 0.1. The covariates considered are shown in Table 1.

Table 1. The description of covariates (Aver.: Average; Rev.: Revnue)

Var.	Variable Names	Description
A1	Current Ratio	Current Assets/Current Liabilities×100%
A2	Quick Ratio	Liquid capital /Current Liabilities×100%
A3	Interest Expense Ratio	Interest Expense /Operating Income ×100%
A4	Debt/Equity Ratio	Total Liabilities /Shareholders' Equity ×100%
A5	Debt/Asset Ratio	Total Liabilities /Total Assets×100%
A6	Equity/Asset Ratio	Shareholders' Equity /Total Assets×100%
A7	P. C./Fixed Assets Ratio	Permanent Capital /Fixed Assets ×100%
A8	Debt /Equity Ratio	(Long&Short-term Debt) /Shareholders' Equity ×100%
A9	C.L./Equity Ratio	Contingent Liabilities /Shareholders' Equity ×100%
A10	O. P./Capital Stock Ratio	Operating Profit /Capital Stock ×100%
A11	P.B.T./C. S. Ratio	Profit Before Tax /Capital Stock ×100%
B1	I.R./Equity Ratio	Inventory and Receivables /Shareholders' Equity×100%
B2	Total Assets Turnover	Revenue /Aver. Total Assets
B3	Receivables Turnover	Revenue /Aver. Receivables
B4	Days' S. in A. R. Ratio	Aver. Receivables /Net Rev. ×Days
B5	Inventory Turnover	Operating Costs /Aver. Inventory
B6	Average Days in Sales	Aver. Inventory /Operating Costs×Days
B7	Fixed Asset Turnover	Revenue /Average Fixed Assets
B8	Equity Turnover	Revenue /Aver. Shareholders' Equity
B9	Days Payable Outstanding	Aver. Payables /Operating Costs×Days
B10	Net Operating Cycle	Days Inventory Outstanding+ Days Sales Outstanding+Days Payables Outstanding
B11	Receivables /Rev. Ratio	Receivables /Rev.×100%
B12	Inventory /Rev. Ratio	Inventory /Rev.×100%
C1	ROA(EBITDA)	EBITDA /Aver. Total Assets ×100%
C2	ROA(EBIT)	[Net Income+Interest Expense*(1-Tax Rate)] /Aver. Total Assets ×100%
C3	Return On Assets (EBPT)	EBPT /Aver. Total Assets×100%
C4	Return On Equity (BTAX)	Income before Tax /Aver. Total Equity×100%
C5	Return On Equity	Recurring Income /Average Total Equity ×100%
C6	Gross Profit Margin	Gross Profit /Net Rev.×100%
C7	Gross Profit Rate (Real Estate)	Gross Profit(Real Estate) /Net Rev. ×100%
C8	Operating Margin	Operating Income /Net Rev.×100%
C9	Earning before Tax Margin	Income before Tax /Net Revenue×100%
C10	Earning after Tax Margin	Income after Tax Margin /Net Rev.×100%
C11	Cost-plus Ratio	(Selling Price-Production Cost) /Production Cost×100%
C12	N. Non-operating In./N. Rev.	Net Non-operating Income /Net Rev. ×100%
C13	R. Income(after Tax)/N. Rev.	Recurring Income(after Tax) /Net Rev. ×100%

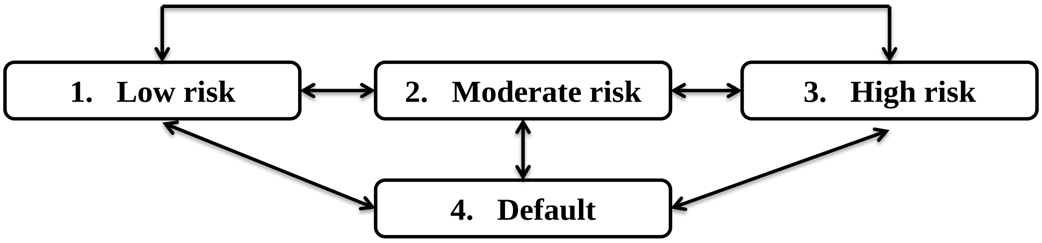

Figure 2. Schematic depiction of multi-state model

3.1.1 $1 - to - 2$ transition

In the section, for $1 - to - 2$ transition (i.e. low risk-to-moderate risk transition), 33 semiconductor firms, is used to fit Cox regression model for downgrading prediction of low risk-to-moderate risk transition. There are 23 uncensored observations and 10 right-censored observations. Based on the forward selection method (slentry=0.1) in SAS software, only two covariates A10 (Operating Profit /Capital Stock Ratio) and A3 (Interest Expense Ratio) enter the model with p-values of 0.0002 and 0.0114, respectively. The corresponding estimated parameters for A10 and A3 are equal to -0.086 and 0.022. It indicates that the A10 is negatively correlated while the A3 is positively correlated with low risk-to-moderate risk transition. Hence, the higher "Operating Profit /Capital Stock Ratio" is the lower hazard of transition from rating low risk to moderate risk. On the other hand, the higher "Interest Expense Ratio" is the higher hazard of this transition.

3.1.2 $2 - to - 3$ transition

In the section, 86 semiconductor firms is used to fit Cox regression model for downgrading prediction of moderate risk-to-high risk transition. There are 28 uncensored observations and 58 right-censored observations. Five covariates, namely, A9 (Contingent Liabilities /Equity Ratio), B10 (Net Operating Cycle), A5 (Debt /Asset Ratio), C1 (Return On Assets(EBITDA)) and B5 (Inventory Turnover) enter the model with p-values of 0.036, 0.0008, 0.002, 0.036 and 0.048, respectively. The corresponding estimated parameters for the five covariates are equal to 0.062, 0.02, 0.046, -0.482 and 0.046. It indicates that covariate C1 is negatively correlated with moderate risk-to-high risk transition while four covariates (A9, B10, A5 and B5) are positively correlated with this transition. The results show that the increase in the C1 variable (Return On Assets(EBITDA)) will decrease this transition probability, while the increase in A9/B10/A5/B5 will increase this transition probability.

3.1.3 $1 - to - 3$ transition

For $1 - to - 3$ transition, no covariate is selected into model. One explanation for this result is that only few observations with $1 - to - 3$ transition are available.

3.1.4 $3 - to - 4$ transition

In the section, for $3 - to - 4$ transition (i.e. high risk-to-default transition), 33 semiconductor firms is used to fix Cox regression model for the prediction of Low risk-to-Moderate risk transition. There are 27 uncensored observations and 6 right-censored observations. Forward selection procedure (slentry=0.1) in SAS software is conducted to select important covariates. Only three covariates C4 (Return On Equity (BTAX)), A11 (Profit Before Tax /Capital Stock Ratio) and A9 (Contingent Liabilities /Equity Ratio) enter the model with p-values of 0.004, 0.015 and 0.027, respectively. The corresponding estimated parameters for three covariates are equal to -0.173, 0.231 and 0.166. It indicates that the C4 is negatively correlated with high risk-to-default transition while A9 and A11 are positively correlated with this transition.

3.2 Analysis Under CSMM

3.2.1 $1 - to - 2$ transition

In the section, for $1 - to - 2$ transition (i.e. low risk-to-moderate risk transition), twenty-five semiconductor firms is used to fix Cox regression model for the prediction of Low risk-to-Moderate risk transition. There are 15 uncensored observations and 10 right-censored observations. Forward selection procedure (slentry=0.1) in SAS software is conducted to select important covariates. Only two covariates C2 (Return On Assets(EBIT $\times$ (1-Tax Rate)) and C9 (Earning before tax margin) enter the model with p-values of 0.0176 and 0.0003, respectively, i.e. there is a significant relationship between the two covariates and downgrading time in the semiconductor industry. The corresponding estimated parameters for C2 and C9 are equal to 0.472 and -0.109. It indicates that the C2 is positively correlated with Low risk-to-Moderate risk transition while the C9 is negatively correlated with this transition. Hence, the higher Return On Assets the higher hazard of transition from rating Low risk to Moderate risk. On the other hand, the higher Earning before tax margin the lower hazard of transition from rating Low risk to Moderate risk.

3.2.2 $2 - to - 3$ transition

In the section, 83 semiconductor firms is used to fix Cox regression model for the downgrading prediction of moderate risk-to-high risk transition. There are 27 uncensored observations and 56 right-censored observations. Based on forward selection procedure (slentry=0.1), five covariates C2(Return On Assets(EBIT×(1-Tax Rate))), A3(Interest Expense Ratio), B3(Receivables Turnover), B5(Inventory Turnover) and B10(Net Operating Cycle) enter the model with p-values of less than 0.0001, 0.0001, less than 0.0001, 0.0023 and 0.1038, respectively. The corresponding estimated parameters for five covariates are equal to -0.228, -0.004, 0.01, 0.406 and 0.027, which indicates that two covariates (C2 and A3) are negatively correlated while the three covariates (B10, B3 and B5) are positively correlated with this transition. The results show that the increase in the C2/A3 variables will decrease the transition probability from moderate risk to high risk, On the other hand, the increase in the B3/B5/B10 variables will increase this transition probability.

3.2.3 $1 - to - 3$ transition

For $1 - to - 3$ transition, no covariate is selected into model. One explanation for this result is that only few observations with $1 - to - 3$ transition are available.

3.2.4 $3 - to - 4$ transition

In the section, 26 semiconductor firms is used to fix Cox regression model for the downgrading prediction of high risk-to-default transition. There are 6 uncensored observations and 20 right-censored observations. Based on forward selection procedure (slentry=0.1), only one covariate C2 (Return On Assets(EBIT×(1-Tax Rate))) is entered in the model with p-values of 0.0027. The corresponding estimated parameters for C2 is equal to -0.124. It indicates that the C2 is negatively correlated with high risk-to-default transition, i.e. the increase in the C2 variable will decrease the transition probability from high risk to default.

Remark 2:

From the above analysis, we see that the covariates selected in the model under CMM differ from that under CSMM. This is expected due to different assumption between the two models.

3.3 Assessing Model Adequacy

Next, based on martingale residuals, we check whether CMM or CSMM is a better choice. Under CMM, for $i \to j$ transition, let

$$\hat{M}_{ijk}(t) = N_{ijk}(t) - \int_0^t Y_{ik}(u) \exp(Z_k(u)^T \hat{\beta}_{ij}) d\hat{\Lambda}_{ij0}(\hat{\beta}_{ij}, u).$$

Define $\hat{M}_{ijk} = \hat{M}_{ijk}(\infty)$. Analogous to the properties of the martingales, these residuals satisfy $\sum_k \hat{M}_{ijk} = 0$ and for large samples, $Cov(\hat{M}_{ijk}, \hat{M}_{ijl}) \simeq 0$ for all $l \neq k$.

Similarly, under CSMM, for $i \to j$ transition, let

$$\hat{\tilde{M}}_{ijk}(t) = \tilde{N}_{ijk}(t) - \int_0^t \tilde{Y}_{ik}(u) \exp(Z_k(u)^T \tilde{\beta}_{ij}) d\tilde{\Lambda}_{ij0}(\tilde{\beta}_{ij}, u)$$

Define $\hat{\tilde{M}}_{ijk} = \hat{\tilde{M}}_{ijk}(\infty)$.

The martingale residual gives a measure of the difference between the observed and fitted value as expected from the model. The plot of the martingale residuals (i.e. $\hat{M}_{ijk}$ for CMM and $\hat{\tilde{M}}_{ijk}$) versus risk score (i.e. $Z_k(u)^T \hat{\beta}_{ij}$ for CMM and $Z_k(u)^T \hat{\beta}_{ij}$ for CSMM) can be used to detect extreme values. However, martingale residuals are highly skewed. The deviance residuals defined by a transformation of the martingale has a more normally shaped than martingale.

The deviance residual, denoted by $\hat{D}_{ijk}$ for CMM and $\hat{\tilde{D}}_{ijk}$ for CSMM are given by

$$\hat{D}_{ijk} = \text{sign}[\hat{M}_{ijk}][-2(\hat{M}_{ijk} + \delta_k \log(\delta_k - \hat{M}_{ijk}))]^{1/2}$$

and

$$\hat{\tilde{D}}_{ijk} = \text{sign}[\hat{\tilde{M}}_{ijk}][-2(\hat{\tilde{M}}_{ijk} + \delta_k \log(\delta_k - \hat{\tilde{M}}_{ijk}))]^{1/2},$$

respectively.

The deviance residual has a value of zero when the martingale is zero. The logarithm tends to inflate the value of deviance residual when the martingale is close to one and shrink large negative values. When no influential observations exit and censoring is light, the plot of deviance residual versus the risk score will appear as a normal distribution. When censoring is heavy (i.e. many observations with $\delta_k = 0$), a lot of points near zero distort the normal distribution. Possible influential observations will have large absolute values of deviance residuals. Furthermore, large positive values of martingale residual indicate that the estimated intensities are lower than the true intensities, i.e. the observed transition actually occur before the model predicts it. On the contrary, large negative values of martingale residual indicate that the observed transition actually occur after the model predicts it. Both cases lead to large values of deviance residuals.

Figures 3 and 4 show the plots of deviance residual versus the risk score for CMM and CSMM, respectively. For CMM, among 86 observations, there are four points (2.19, 2.37, 2.43 and 2.54) with values of deviance residuals lager than 2.0. For CSMM, among 83 observations, there are three points (2.09, 2.17. 2.47) with values of deviance residuals lager than 2.0. Hence, the CSMM seems to fit better than the CMM, i.e. the transition probability to state high-risk after being rated as "moderate-risk" may depend on the time since rated as "moderate-risk".

Chapter 4

Conclusions

In this article, we have demonstrated the use of CMM model in analyzing corporate rating data. Under CMM model, It is found that key factors that affect rating transitions from ‘low risk’ to ‘moderate risk’ are A3 and A10 while the key factors that affect rating transitions from ‘moderate risk’ to ‘high risk’ are A5, A9, B5, B10 and C1. The key factors that affect rating transitions from ‘high risk’ to ‘default’ are A9, A11 and C4. Our transition model has baseline intensities of regression parameters specific to each transition type. By comparing the MSM with the Cox model with time-dependent covariates, we have demonstrated that the MSM can provide new insights while confirming some of the results obtained from the Cox model.

In this article, it is assumed that the onset time of a state can be observed exactly. However, in practice, the corporate is usually rated at a given number of visits, leading to interval-censored observations. Incorporation of interval-censored data into the proportional hazards model does not enable canceling of the baseline hazard function, i.e. the partial likelihood approach is not available for interval-censored data. Pan (2000) used a multiple imputation procedure to fill-in failure times for the interval censored events and then applied the standard partial likelihood analysis. Heller (2011) proposed an alternative methodology, which is based on estimating equations and uses an inverse probability weight to select event time pairs where the ordering is unambiguous. Sun et al. (2015) proposed two simple estimation approaches that do not need estimation of the baseline cumulative hazard function. Further research is required to extend CMM/CSMM model to interval censored data.

References

[**1**] Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations, *Scandinavian Journal of Statistics*, **5**, 141-150.

[**2**] Aalen, O. O. (1980). A model for non-parametric regression analysis of counting process. In:Klonecki, W., Kozek, A., Rosiski, J. (Eds). Lecture Notes on Mathematical Statistics and Probability, Vol. 2. Springer, New York, pp. 1-25.

[**3**] Aalen, O. O. (1989). A linear regression model for analysis of life times. *Statistics in Medicine*, **8**, 907-925.

[**4**] Aalen, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine*, **17**, 1569-1588.

[**5**] Andersen P. K., Hansen L. S., Keiding N. (1991) Non- and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous Markov process. *Scandinavian Journal of Statistics*, **18**, 153-167.

[**6**] Andersen P. K., Borgan, *O*., Gill, R. D. & Keiding, N. (1993). Statistical models based on counting processes. New York: Springer-Verlag.

[**7**] Andersen P. K., Esbjerg S., S*o*rensen TI (2000). Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine*, **19**, 587-599.

[**8**] Andersen, P.K. and Keiding, N. (2002). Multi-state models for event history analysis, *Statistical Methods in Medical Research*, **11**, 91-115.

[**9**] Breslow, N. E. (1972). Discussion following "Regression models and life tables" by D. R. Cox. *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.

[**10**] Breslow, N. E. and Crowley, J. (1974). A large-sample study of the life table and product limit estimates under random censorship. *Annals of Statistics*, **2**, 437-454.

[**11**] Chancharat, N., Tian, G.; Davy, P., McCrae, M., and Lodh, S. (2010). Multiple states of financially distressed companies : tests using a competing-risks model, *Australasian Accounting, Business and Finance Journal*, **4**, 27-44.

[**12**] Commenges D. (1999). Multi-state models in epidemiology. *Lifetime Data Analysis*, 315-327.

[**13**] Cox D. R., Miller HD (1965). The Theory of Stochastic Processes. Chapman and Hal, London.

[**14**] Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of Royal Statistical Society, Ser.B*, **34**, 187-220.

[**15**] Heller, G. (2011). Proportional hazards regression with interval censored data using an inverse probability weight. *Lifetime Data Analysis*, **17**, 373-385.

[**16**] Hougaard P. (1999). Multi-state models: a review. *Lifetime Data Analysis*, 239-264.

[**17**] Huffer, F. W. and Mckeague, F. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, **86**, 114-129.

[**18**] Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observation. *Journal of the American Statistical Association*, **53**, 457-481.

[**19**] McKeagues, I. W. (1988). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. *in Statistical Inference from Stochastic Processes*, page 139-152.

[**20**] Meira-Machado, L.; de de Uña-Álvarez, J.; Cadarso-Suárez, C. and Andersen, P.K. (2009). Multi-state models for the analysis of time to event data, *Statistical Methods in Medical Research*, **18**, 195-222.

[**21**] Pan, W. (2000). A multiple imputation approach to cox regression with interval-censored data. *Biometrics*, **56**, 199-203.

[**22**] Shu, Y., Klein, J. P. and Zhang, M.-J. (2007). Asymptotic theory for the Cox semi-Markov illness-death model. *Lifetime Data Analysis*, **13**, 91-113.

[**23**] Sun, J., Feng, Y. and Zhao, H. (2015). Simple estimation procedures for regression analysis of interval-censored failure time data under the proportional hazards model. *Lifetime Data Analysis*, **21**, 138-155.

[**24**] Voelkel J. G. and Crowley J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations. *The Annals of Statistics*, **12**, 142-160.