6-7-2023

Creating Regression Model for Non-Markov Transition Probability Using Pseudo-Observations

Michael Gray
Portland State University

Creating Regression Model For Non-Markov Transition Probability Using Pseudo-Observations

by

Michael Gray

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Mathematical Sciences

Dissertation Committee:
Jong Sung Kim, Chair
Robert Fountain
Daniel Taylor-Rodriguez
Ge Zhao

Portland State University
2023

Abstract

A multi-state model is a graphical tool widely used to illustrate a transitional relationship between states in many applications. We will study the transition probabilities of an illness-death model, which is an example of a multi-state model. We will investigate transition probabilities using a counting process approach. Aalen-Johansen estimator is the gold-standard in estimating a transition probability. However, Aalen-Johansen estimator may be biased when the Markov assumption is violated. Therefore, Aalen-Johansen estimator is an unreliable estimator when the Markov assumption is violated. Several papers have published non-parametric estimators that accommodate for non-Markov models using a counting process approach.

Furthermore, there are few existing work in creating a regression model for transition probabilities in the non-Markov setting. Our goal is to contribute to the few existing work of regression models that accommodate non-Markov behavior. In creating the regression model, we use the jackknife method, pseudo-observations. In finding parameter estimates, generalized estimation equation(GEE) will be used. An important requirement in using pseudo-observations is that we need an unbiased estimator. Aalen-Johansen estimator would be a unreliable choice since it is susceptible to bias. We propose in using Titman estimator as an alternative estimator to create the pseudo-observation for the regression model. Titman estimator is shown to be unbiased from [28]. It also can be used in time-irreversible and time-reversible models. This feature of Titman estimator allows practitioners to find the transition probability of recovering from an illness in the illness-death model.

In a simulation study, we will compare the results when creating pseudo-observations by using Titman estimator and Aalen-Johansen estimator. We will illustrate the regression model using the illness-death model when recovery is not assumed and

illness-death model when recovery is assumed. We will study when the model is "pathologically" non-Markov and the model has a frailty effect. Both cases violate the Markov assumption. Finally, we will analyze the liver cirrhosis dataset using our proposed method.

Dedicated to my dad and mom, Peter and Ayako Gray

Acknowledgments

I would like to thank my advisor, Dr. Jong Sung Kim. He has been very supportive and patient with my academics and research throughout my graduate school career. I would also like to thank other members of my dissertation committee: Dr. Robert Fountain, Dr. Daniel Taylor-Rodriguez, and Dr. Ge Zhao. I appreciate them agreeing to be in the dissertation committee and their insightful feedback for this project. I would like to thank the Fariborz Maseeh Department of Mathematics and Statistics for giving me the opportunity to become a graduate teaching assistant and funding the Ph.D. program. I would like to thank Dr. Andrew Titman from Lancaster University, for his assistance with the simulation code.

I would like to thank my parents, Peter and Ayako Gray for their support throughout my whole academic career. They have always been encouraging me to study and work hard. I would also like to thank my sister, Sharon Aguilar-Gray, and my brother-in-law, Roberto Aguilar-Cerna for their support and encouragement throughout the Ph.D. program. I would like to thank my grandfather, Dr. Gordon Gray for his generous financial gifts to help fund my college career. I would like to thank my friend, Dr. Naghmeh Daneshi for her support and encouraging me to join the Ph.D. program.

Table of Contents

List of Tables

List of Figures

1 Introduction

Multi-state model is a popular way of illustrating an individual's progression to various stages. A multi-state model is a directed graph that represents the transitional relationship between the states(nodes). Illness-death models are useful in describing medical applications, particularly a progression of a disease. The illness-death model is a three-state model that represents a "healthy", "illness" and "death" state. When recovery is not possible, it is possible for an individual to transition from being healthy to getting ill or death. It is also possible for an individual to transition from being ill to death. See Figure 1. We will denote the healthy state as state 0, the illness state as state 1, and the death state as state 2. In a multi-state model, we are interested in the transition probability of an individual moving to another state at a certain time. The transition probability of state j to state k from time s to time t can be expressed as

$$P_{jk}(s,t) = P(X(t) = k|X(s) = j) \tag{1}$$

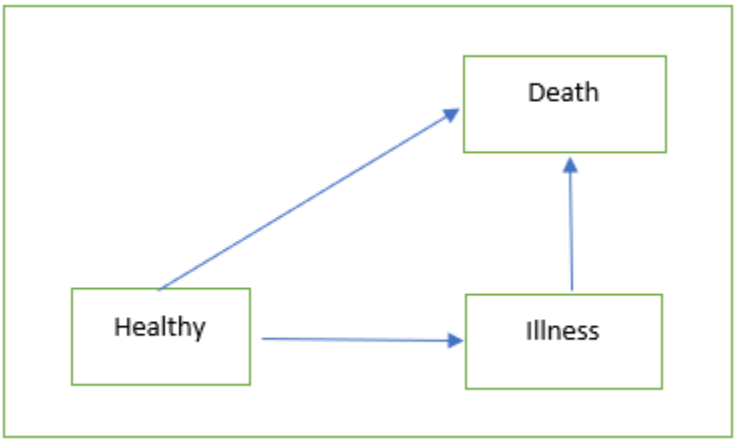

Figure 1: Illness-Death Model without Recovery

Extensive research has been done in computing the state-occupation probability and transition probability from state to state. The most well-known method was established by Aalen and Johansen [1]. [1] created a non-parametric estimator(known as the Aalen-Johansen estimator) that estimates state-occupation probability and transition probability using Nelson-Aalen estimator and the product integral. Aalen-Johansen estimator has the form,

$$\hat{\mathbf{P}}(s,t) = \prod_{(s,t)} (\mathbf{I} + d\hat{\mathbf{\Lambda}}(u)) \tag{2}$$

where $\hat{\mathbf{P}}(s,t)$ is a matrix of transition probabilities. $\mathbf{I}$ is the identity matrix. $\hat{\mathbf{\Lambda}}(u)$ is a matrix of Nelson-Aalen estimators. Let $dN_{jk}(u)$ be the number of observed events at time u. Let $Y_{jk}(u)$ be the number of individuals who are at risk of encountering the event at time $u-$.

$$\hat{\mathbf{\Lambda}}(t) = \begin{cases} \int_0^t dN_{jk}(u)/Y_{jk}(u) & j \neq k \\ -\sum_{j \neq k} \Lambda_{jk}(t) & j = k \end{cases} \tag{3}$$

Aalen-Johansen estimator is considered the gold standard of estimating transition probabilities and state-occupation probabilities. Aalen-Johansen estimator assumes the Markov assumption holds. The Markov property holds when transitioning to a different state only depends on the state that the individual is currently at. This is a strong assumption in many applications. When the Markov property is violated, the Aalen-Johansen estimator may be biased in estimating transition probabilities ([21], [2], [10], [28]). Thus, the Aalen-Johansen estimator may be a poor choice in estimating transition probabilities for non-Markov models. [9] showed that for non-Markov models, the Aalen-Johansen estimator is a consistent estimator for state-occupation

probability with the assumption of independent censoring. However, this does not hold for transition probabilities. In the past couple of decades, non-parametric estimators for transition probability in the non-Markov setting have been published ([21], [2], [10], [28]).

In survival analysis, there are few existing work in creating semi-parametric regression models for transition probability in multi-state model for non-Markov models. [5] created a direct binomial regression model for transition probabilities using inverse censoring probability weighting introduced in [25].[13] created a semi-parametric regression model using the Landmark Aalen-Johansen estimator ([23]) in inverse-probability weighted data.

Covariates are included in many studies and datasets. We believe practitioners will be interested in knowing whether the covariate has a postive or negative effect to the transition probability. In addition, practitioners will be interested in which covariate would be a significant predictor. Therefore, we would like to contribute to the few existing semi-parametric regression model. The main tools to create our regression model is the non-parametric estimator created by Titman [28] and the pseudo-observation method introduced by Andersen and others [4]. These methods will be explained in detail in the next section.

In Section 3 and Section 4, we will create a simulation study showing the pseudo-observation method using an illness-death model without recovery and illness-death model with recovery, respectively. In Section 3 and Section 4, we refer to non-Markov as having a "pathological" non-Markov behavior. In Section 5, we will discuss frailty which is a different non-Markov behavior. In Section 6, we will analyze the liver cirrhosis dataset using our proposed method. In Section 7, we will conclude the dissertation with discussions and extensions.

2 Main Methods

2.1 Titman Estimator

The concept of creating non-parametric landmark estimators for multi-state models was introduced by [10]. The concept of landmarking in estimation is considering a subsample of the complete data when computing the estimator. For the illness-death model with no recovery, [10] considered the subsample of individuals who are at risk at time s. [10] focused on the non-parametric estimator of the transition probability for the illness-death model without recovery. [28] also created a non-parametric estimator that used landmarking. However, [28]'s estimator works for general multi-state model where there can be more than three states. It can also accommodate for time-reversible models where it is possible for a multi-state model to have states that can transition in either directions. This includes the illness-death model with the possibility of recovery. Thus, the illness-death model with recovery is an illness-death model when it is also possible for individuals to transition from illness to healthy.

Let $\{X(t), t > 0\}$ be a multi-state process such that $X(t) \in \mathcal{S} = \{1, 2, ..., S\}$.There are S states in the process. [28] proposed an estimator of the transition of a set of states $\mathcal{J}$ to another set of states $\mathcal{K}$. $\mathcal{J}$ and $\mathcal{K}$ are non-empty sets and do not have to be disjoint subsets of $\mathcal{S}$. In our study, we are focusing on a more practical case of transition of two distinct states. The transition can be simplified as moving from state j to state k. $\mathcal{R}_k$ is denoted as the set of states reachable from state j, but cannot reach state k. $\mathcal{A}_k$ is defined such that it is either the singleton $\{k\}$ when k is an absorbing state, or an empty set when k is not an absorbing state. Then a competing risk process, $\{C_s(u), u \geq s\}$ can be defined as

$$C_s(u) = \begin{cases} 0 & if\ X(u) \notin \{\mathcal{R}_k \cup \mathcal{A}_k\} \\ 1 & if\ X(u) \in \mathcal{A}_k \\ 2 & if\ X(u) \in \mathcal{R}_k \end{cases} \tag{4}$$

Then $P(X(t) = k|X(s) = j)$ can be computed as

$$\begin{aligned} P(X(t) = k|X(s) = j) =& P(C_s(t) = 0, X(t) = k|X(s) = j) \\ &+ P(C_s(t) = 1, X(t) = k|X(s) = j) \\ &+ P(C_s(t) = 2, X(t) = k|X(s) = j) \\ =& P(C_s(t) = 0, X(t) = k|X(s) = j) + P(C_s(t) = 1|X(s) = j) \end{aligned} \tag{5}$$

The events, $C_s(t) = 2$ and $X(t) = k$ are disjoint since $\mathcal{R}_k$ is the set of states that state k is not reachable.This makes $P(C_s(t) = 2, X(t) = k|X(s) = j) = 0$. Considering $C_s(t) = 1$, $\mathcal{A}_k$ is non-empty when k is an absorbing state. Also when $C_s(t) = 1$, the process is at the absorbing state k. This implies that $X(t) = k$ when $C_s(t) = 1$. So, $P(C_s(t) = 1, X(t) = k|X(s) = j) = P(C_s(t) = 1|X(s) = j)$. From this,

$$\begin{aligned} &P(C_s(t) = 0, X(t) = k|X(s) = j) + P(C_s(t) = 1|X(s) = j) \\ &= P(X(t) = k|C_s(t) = 0, X(s) = j)P(C_s(t) = 0|X(s) = j) + P(C_s(t) = 1|X(s) = j) \end{aligned} \tag{6}$$

Now, let ${}_sY(u)$ be the risk indicator for the competing risk process, $C_s(u)$. $\mathcal{N}_j(s)$ be the set of individuals that fulfill the conditions: $X(s) = j$ and ${}_sY(s) = 1$. That is, the individual is at state j at time s and at risk in the competing risk process, $C_s(u)$. $\mathcal{N}_j(s)$ will represent the subsample from landmarking. That is, landmarking will be

used to subsample only individuals who are at state j at time s. Let ${}_s\bar{Y}(u) = \sum_s Y(u)$ be number of individuals who are at risk at time u for the process, $C_s(u)$. Let ${}_sN_1(t)$ be the counting process of observed events from type 1 and ${}_sN(t)$ be the counting process of observed events of all types from $\mathcal{N}_j(s)$ in the competing risks model. $P(C_s(t) = 0|X(s) = j)$ and $P(C_s(t) = 1|X(s) = j)$ can be estimated by using Aalen-Johansen estimator from individuals only in $\mathcal{N}_j(s)$. $P(X(t) = k|C_s(t) = 0, X(s) = j)$ can be estimated as the proportion of individuals who are in state k at time t among the individuals with $C_s(t) = 0$ in state j at time s and at risk at time t (i.e. ${}_sY(t) = 1$). Being at risk at time t requires the individual to not be censored in the study before or at time t. We will denote this proportion as $\hat{p}_{k|j}(t)$. Now the estimates for $P(C_s(t) = 0|X(s) = j)$, $P(C_s(t) = 1|X(s) = j)$, and $P(X(t) = k|C_s(t) = 0, X(s) = j)$ can be expressed as follows.

$$\hat{F}_0(t) = \hat{P}(C_s(t) = 0|X(s) = j) = \prod_{v\in[s,t]} \left(1 - \frac{d_sN(v)}{{}_s\bar{Y}(v)}\right) \tag{7}$$

$$\hat{F}_1(t) = \hat{P}(C_s(t) = 1|X(s) = j) = \int_s^t \prod_{v\in[s,t]} \left(1 - \frac{d_sN(v)}{{}_s\bar{Y}(v)}\right) \frac{d_sN_1(u)}{{}_s\bar{Y}(u)} \tag{8}$$

$$\begin{aligned} &\hat{p}_{k|j}(t) = \hat{P}(X(t) = k|C_s(t) = 0, X(s) = j) \\ &= \frac{\sum I(X(s) = j\ , X(t) = k\ ,\ {}_sY(s) = 1\ ,\ {}_sY(t) = 1)}{\sum I(X(s) = j\ ,\ {}_sY(s) = 1\ ,\ {}_sY(t) = 1)} \end{aligned} \tag{9}$$

From this, the proposed estimator of [28] becomes

$$\begin{aligned} &\hat{P}_{jk}(s,t) \\ &= \hat{P}(C_s(t) = 1|X(s) = j) + \hat{P}(C_s(t) = 0|X(s) = j)\hat{P}(X(t) = k|C_s(t) = 0, X(s) = j) \\ &= \hat{F}_1(t) + \hat{F}_0(t)\hat{p}_{k|j}(t) \end{aligned} \tag{10}$$

We will call (10) as Titman estimator. Using (10), Titman estimator of each transition of the three-state illness-death model without recovery can be expressed as the following. In each of the cases, the competing risk model reduces to a simple survival process.

For the transition from healthy to illness, the states $\{0,1\} \notin \{\mathcal{R}_k \cup \mathcal{A}_k\}$. $2 \in \mathcal{R}_k$. Then, the competing risk model can be simplified as

$$C_s(u) = \begin{cases} 0 & if\ X(u) \in \{0,1\} \\ 2 & if\ X(u) = 2 \end{cases}$$

Also, due to $\mathcal{A}_1 = \emptyset$, Titman estimator can be reduced to

$$\hat{P}_{01}(s,t) = \hat{P}(C_s(t) = 0 | X(s) = 0)\hat{P}(X(t) = 1 | C_s(t) = 0, X(s) = 0)$$

$$= \hat{F}_0(t)\hat{p}_{1|0}(t) \tag{11}$$

where

$$\hat{F}_0(t) = \hat{P}(C_s(t) = 0 | X(s) = 0) = \prod_{v \in [s,t]} \left(1 - \frac{d_s N(v)}{{}_s\bar{Y}(v)}\right)$$

$$\hat{p}_{1|0}(t) = \hat{P}(X(t) = 1 | C_s(t) = 0, X(s) = 0)$$
$$= \frac{\sum I(X(s) = 0\ , X(t) = 1\ ,\ {}_sY(s) = 1\ ,\ {}_sY(t) = 1)}{\sum I(X(s) = 0\ ,\ {}_sY(s) = 1\ ,\ {}_sY(t) = 1)}$$

For the transition from the healthy state to the death state, the states $\{0,1\} \notin \{\mathcal{R}_k \cup \mathcal{A}_k\}$. $2 \in \mathcal{A}_k$. In this case, $C_s(u)$ can be written as

$$C_s(u) = \begin{cases} 0 & if\ X(u) \in \{0,1\} \\ 1 & if\ X(u) = 2 \end{cases}$$

In the healthy to illness transition, $2 \in \mathcal{R}_k$, but in this case, $2 \in \mathcal{A}_k$. That is, the death state(state 2) is an absorbing state. Since state 2 is an absorbing state, $P(X(t) = 2|C_s(t) = 0, X(s) = 0) = 0$. This is because the only time $X(t) = 2$ is when $C_s(t) = 1$.This reduces the estimator to

$$\hat{P}_{02}(s,t) = \hat{P}(C_s(t) = 1|X(s) = 0) = \hat{F}_1(t) \tag{12}$$

where

$$\hat{F}_1(t) = \hat{P}(C_s(t) = 1|X(s) = 0) = \int_s^t \prod_{v \in [s,t]} \left(1 - \frac{d_s N(v)}{{}_s\bar{Y}(v)}\right) \frac{d_s N_1(u)}{{}_s\bar{Y}(u)}$$

In competing risks, $\hat{F}_1(t)$ is also known as the cumulative incidence function of type 1. We can directly compute the expression above, but when we have a simple survival process we have another option. When we have a simple survival process, $\hat{F}_1(t) = 1 - \hat{S}(t)$ where $\hat{S}(t)$ is the Kaplan-Meier estimator of surviving any cause of failure. $\hat{S}(t)$ can be represented as $\hat{F}_0(t)$. So, $\hat{F}_1(t)$ can be computed as

$$\hat{F}_1(t) = 1 - \hat{F}_0(t) = 1 - \hat{P}(C_s(t) = 0|X(s) = 0) = 1 - \prod_{v \in [s,t]} \left(1 - \frac{d_s N(v)}{{}_s\bar{Y}(v)}\right)$$

This alternative way of computing the estimator is equivalent to the estimator from [10].

The case of transition from illness to death is similar to the healthy to death transition.

$$C_s(u) = \begin{cases} 0 & if\ X(u) \in \{0,1\} \\ 1 & if\ X(u) = 2 \end{cases}$$

However, the only difference in this case is that the landmark subsample contains individuals who are ill at time s. In the healthy to death transition, a subsample of only individuals who were healthy at time s were included. The estimator is computed as

$$\hat{P}_{12}(s,t) = \hat{P}(C_s(t) = 1 | X(s) = 1) = \hat{F}_1(t) \tag{13}$$

where

$$\hat{F}_1(t) = \hat{P}(C_s(t) = 1 | X(s) = 1) = \int_s^t \prod_{v \in [s,t]} \left(1 - \frac{d_s N(v)}{{}_s\bar{Y}(v)}\right) \frac{d_s N_1(u)}{{}_s\bar{Y}(u)}$$

or alternatively

$$\hat{F}_1(t) = 1 - \hat{F}_0(t) = 1 - \hat{P}(C_s(t) = 0 | X(s) = 1) = 1 - \prod_{v \in [s,t]} \left(1 - \frac{d_s N(v)}{{}_s\bar{Y}(v)}\right)$$

2.2 Pseudo-Observations

The other main tool that we will be using is pseudo-observation. Pseudo-observations are mainly used in jackknife methods. The pseudo-observation, denoted by $\hat{\theta}_i$ is computed for each observation i. A pseudo-observation measures the contribution of each individual observation to a summary statistic, $\hat{\theta}$. [4] introduced a regression of the pseudo-observation, $\hat{\theta}_i$ on to covariates. The slopes of the covariates are estimated using generalized estimating equations (GEE).

To further explain pseudo-observations from [4], consider $X_1, ..., X_n$ be independent and identically distributed random variables. In general, X_i can be a random vector or process ($\{X_i(t), t \geq 0\}$). Then consider the expectation, θ for some function, h. Hence,

$$\theta = E(h(X)) = \int_x h(x)f(x)dx \tag{14}$$

An unbiased estimator, $\hat{\theta} = \hat{\theta}(X)$ exists for θ. That is,

$$E_X(\hat{\theta}) = \int_x \hat{\theta}(x)f(x)dx = \theta \tag{15}$$

Let $Z_1, ..., Z_n$ be independent and identically distributed random variables of covariates. Then,

$$\begin{aligned}\theta = \int_x h(x)f(x)dx &= \int_x h(x)f(x)dx \int_z f(z|x)dz \\ &= \int_x \int_z h(x)f(x,z)dzdx \\ &= \int_x \int_z h(x)f(x|z)f(z)dzdx\end{aligned}$$

By interchanging integrals,

$$= \int_z \int_x h(x)f(x|z)f(z)dxdz$$

$$\theta = \int_z E(h(X)|Z = z)f(z)dz \tag{16}$$

Since the covariate, Z is considered random prior to obtaining data, the conditional expectation should be treated as $E(h(X)|Z)$. That is, the conditional expectation is also a random variable. From (16), the distribution of Z will be replaced with the

empirical distribution and define $\hat{\theta}$ as an estimate of the average of conditional expectations, $h(X_i|Z_i)$ from a sample. Define the random variables, $\theta_i(Z_i) = E(h(X_i)|Z_i)$ for $i = 1, ..., n$ and the average of these random variables, $\widetilde{\theta}(Z) = \frac{1}{n}\sum_j \theta_j(Z_j)$.

Now, we will verify that $\widetilde{\theta}(Z)$ is an unbiased estimator of θ with respect to the joint random variables (X, Z). Hence, we will show that

$$E_{X,Z}(\widetilde{\theta}(Z)) = E_Z(\widetilde{\theta}(Z)) = \frac{1}{n}\sum_j E_Z(\theta_j(Z_j)) = \theta$$

To show this, consider $E_Z(\widetilde{\theta}(Z))$ and $E_{X,Z}(\widetilde{\theta}(Z))$.

$$E_Z(\widetilde{\theta}(Z)) = \int_z \widetilde{\theta}(Z) f(z) dz$$

$$\begin{aligned} E_{X,Z}(\widetilde{\theta}(Z)) &= \int_z \int_x \widetilde{\theta}(Z) f(x,z) dx dz \\ &= \int_z \int_x \widetilde{\theta}(Z) f(z) f(x|z) dx dz \\ &= \int_z \widetilde{\theta}(Z) f(z) \int_x f(x|z) dx dz \\ &= \int_z \widetilde{\theta}(Z) f(z) dz \\ &= E_Z(\widetilde{\theta}(Z)) \end{aligned}$$

$$
\begin{aligned}
E_Z(\widetilde{\theta}(Z)) &= E_Z\left(\frac{1}{n}\sum_j \theta_j(Z_j)\right) \\
&= \frac{1}{n}\sum_j E_Z(\theta_j(Z_j)) \\
&= \frac{1}{n}\sum_j E_Z(E_X(h(X_i)|Z_i)) \\
&= \frac{1}{n}\sum_j E_X(h(X_i)) \\
&= \frac{1}{n}\sum_j \theta = \frac{1}{n}(n\theta) = \theta
\end{aligned}
$$

Thus, $E_{X,Z}(\widetilde{\theta}(Z)) = E_Z(\widetilde{\theta}(Z)) = \frac{1}{n}\sum_j E_Z(\theta_j(Z_j)) = \theta$. □

Let the leave-one out estimator, $\widetilde{\theta}_{-i}(Z) = \frac{1}{n-1}\sum_{j\neq i}\theta_j(Z_j)$. $\widetilde{\theta}_{-i}(Z)$ is $\widetilde{\theta}(Z)$ without the contribution of individual i. Considering $\theta_i(Z_i)$ for individual i, $\theta_i(Z_i)$ can be written as

$$
\theta_i(Z_i) = \widetilde{\theta}(Z) + (n-1)(\widetilde{\theta}(Z) - \widetilde{\theta}_{-i}(Z)) = n\widetilde{\theta}(Z) - (n-1)\widetilde{\theta}_{-i}(Z) \tag{17}
$$

Denote $\hat{\theta}_i$ as the individual contribution of individual i for the unbiased estimator, $\hat{\theta}(X)$. Similarly, $\hat{\theta}_i$ can be expressed as

$$
\hat{\theta}_i = \hat{\theta}(X) + (n-1)(\hat{\theta}(X) - \hat{\theta}_{-i}(X)) = n\hat{\theta}(X) - (n-1)\hat{\theta}_{-i}(X) \tag{18}
$$

Now, $\widetilde{\theta}(Z)$ and $\hat{\theta}(X)$ have the same expectation with respect to (X, Z). That is, $E_{X,Z}(\widetilde{\theta}(Z)) = E_{X,Z}(\hat{\theta}(X))$. To show this,

$$E_{X,Z}(\hat{\theta}(X)) = \int_x \int_z \hat{\theta}(X) f(x,z) dz dx = \int_x \int_z \hat{\theta}(X) f(x) f(z|x) dz dx$$
$$= \int_x \hat{\theta}(X) f(x) \int_z f(z|x) dz dx$$
$$= \int_x \hat{\theta}(X) f(x) dx$$
$$= E_X(\hat{\theta}(X)) = \theta$$

We already know $E_{X,Z}(\widetilde{\theta}(Z)) = \theta$ from earlier. Thus, $E_{X,Z}(\hat{\theta}(X)) = E_{X,Z}(\widetilde{\theta}(Z)) = \theta$. Note that we can also make similar arguments with $E_{X,Z}(\widetilde{\theta}_{-i}(Z))$ and $E_{X,Z}(\hat{\theta}_{-i}(X))$.

□

From these results, $\theta_i(Z_i)$ and $\hat{\theta}_i$ have the same expectation. That is, $E_{X,Z}(\theta_i(Z_i)) = E_{X,Z}(\hat{\theta}_i)$. This is critical because we will not be able to compute $\widetilde{\theta}(Z)$ from the dataset. However, the unbiased estimator, $\hat{\theta}(X)$ can be computed from the dataset. Therefore, the pseudo-observation is defined as (18) in which it measures the contribution of individual i.

[4] proposed in constructing a regression model where the pseudo-observation, $\hat{\theta}_i$ depends on the covariates, Z_i for $i = 1, ..., n$. Let $\theta_i = \theta_i(Z_i) = E(h(X_i)|Z_i)$. Then, a generalized linear model in relation of θ_i and Z_i will be

$$g(\theta_i) = \beta^T Z_i \tag{19}$$

where β is a vector of slopes of the covariates and Z_i is a vector of covariates for the ith individual including the intercept. That is, $Z_i^T = [1, Z_{i1}, ..., Z_{ip}]$ when there are p covariates. Then, the inverse link function is denoted as

$$\mu(\beta^T Z_i) = g^{-1}(\beta^T Z_i) = \theta_i \tag{20}$$

$\mu(\beta^T Z_i) = \theta_i$ is known as the mean function of the response variable.

It is also possible to have applications with multiple pseudo-observations. This is when we are interested in multiple time points, $t_1, ..., t_m$ simultaneously. $\theta = [\theta_1, ..., \theta_m]^T$. The pseudo-observations for individual i at the lth time point would be

$$\hat{\theta}_{il} = n\hat{\theta}(t_l) - (n-1)\hat{\theta}_{-i}(t_l) \qquad l = 1, ..., m \tag{21}$$

Thus, there will be m pseudo-observations.

To estimate the slopes of the covariates, β, [4] proposed in constructing a nonlinear regression where $\hat{\theta}_i$ depends on Z_i.The non-linear regression method that will be used is GEE introduced by [17].GEE are based on quasi-likelihood function. Regarding quasi-likelihood functions, consider the exponential family,

$$f(y_i^*) = \exp\left(\frac{y_i^*\eta - a(\eta)}{c(\phi)} + b(y_i^*, \phi)\right) \tag{22}$$

where η is the canonical parameter, y_i^* is the observed response variable for individual i, and ϕ is a nuisance parameter. $E(Y_i^*) = a^{'}(\eta)$ and $Var(Y_i^*) = c(\phi)a^{''}(\eta)$. We denote the response variable as Y^* to avoid confusion on the notation between the response variable and the risk indicator introduced in Section 1. In general from [31], the quasi-likelihood function, $QL(y_i^*, \mu_i)$ for the ith individual is defined as

$$QL(y_i^*, \mu_i) = \int_{y_i^*}^{\mu_i} \frac{y_i^* - t}{c(\phi)V(t)} dt \tag{23}$$

where y_i^* is the response variable and $\mu_i = E(Y_i^*)$ and $Var(Y_i^*) = c(\phi)V(\mu_i)$. $V(.)$ is

a known function. The derivative with respect to μ_i,

$$\frac{\partial QL(y_i^*, \mu_i)}{\partial \mu_i} = \frac{(y_i^* - \mu_i)}{c(\phi)V(\mu_i)} \tag{24}$$

This expression is called the quasi-score function. Ultimately, the score function with respect to β is

$$U_i(\beta) = \frac{\partial QL(y_i^*, \mu_i)}{\partial \beta} = \frac{\partial QL(y_i^*, \mu_i)}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta} = \frac{(y_i^* - \mu_i)}{c(\phi)V(\mu_i)}\frac{\partial \mu_i}{\partial \beta} \tag{25}$$

To illustrate an example of the quasi-likelihood function, consider Y_i^* that has a Poisson distribution with mean λ. Then the exponential family will be

$$f(y_i^*) = \exp\left(\frac{y_i^*\eta - a(\eta)}{c(\phi)} + b(y_i^*, \phi)\right) = \exp\left(y_i^* ln(\lambda) - \lambda - ln(y_i^*!)\right)$$

where the canonical parameter, $\eta = ln(\lambda)$, $a(\eta) = \lambda = \exp(\eta)$, $b(y_i^*, \phi) = -ln(y_i^*!)$, and $c(\phi) = 1$. $E(Y_i^*) = a^{'}(\eta) = \exp(\eta)$ and $Var(Y_i^*) = c(\phi)a^{''}(\eta) = \exp(\eta)$. Thus, $\mu_i = E(Y_i^*) = Var(Y_i^*) = \exp(\eta)$. Since $c(\phi) = 1$, this leads to $V(\mu_i) = \mu_i$. The quasi-score function is

$$\frac{\partial QL(y_i^*, \mu_i)}{\partial \mu_i} = \frac{(y_i^* - \mu_i)}{c(\phi)V(\mu_i)} = \frac{y_i^* - \mu_i}{\mu_i}$$

For the quasi-likelihood function, after we take the anti-derivative of the quasi-score function with respect to μ_i, then

$$QL(y_i^*, \mu_i) = \int \frac{\partial QL(y_i^*, \mu_i)}{\partial \mu_i} = \int \frac{y_i^* - \mu_i}{\mu_i} d\mu_i = y_i^* ln(\mu_i) - \mu_i$$

□

In our application, the exponential family and the canonical parameter will depend on which link function is chosen. For the quasi-likelihood function, θ_i is the mean function(μ_i). The pseudo-observation, $\hat{\theta}_i$ is the observed response variable that will take the role of Y_i^*. The quasi-likelihood function, $QL(\hat{\theta}_i, \theta_i)$, the quasi-score function, $\frac{\partial QL(\hat{\theta}_i,\theta_i)}{\partial \theta_i}$, and the score function with respect to β, $U_i(\beta)$ would be

$$QL(\hat{\theta}_i, \theta_i) = \int_{\hat{\theta}_i}^{\theta_i} \frac{\hat{\theta}_i - t}{c(\phi)V(t)} dt \tag{26}$$

$$\frac{\partial QL(\hat{\theta}_i, \theta_i)}{\partial \theta_i} = \frac{(\hat{\theta}_i - \theta_i)}{c(\phi)V(\theta_i)} \tag{27}$$

$$U_i(\beta) = \frac{\partial QL(\hat{\theta}_i, \theta_i)}{\partial \beta} = \frac{\partial QL(\hat{\theta}_i, \theta_i)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta} = \frac{(\hat{\theta}_i - \theta_i)}{c(\phi)V(\theta_i)} \frac{\partial \theta_i}{\partial \beta} \tag{28}$$

Suppose that we have multiple time points, $\hat{\theta}_i = [\hat{\theta}_{i1}, ..., \hat{\theta}_{im_i}]^T$ such that individual i has m_i pseudo-observations. When there are multiple time points, there will be a covariance matrix with dimensions of $m_i \times m_i$ rather than a single variance. Denote the covariance matrix for individual i, V_i. Thus,

$$U_i(\beta) = \frac{\partial \theta_i}{\partial \beta}^T V_i^{-1} (\hat{\theta}_i - \theta_i) \tag{29}$$

V_i still depends on θ_i. With p covariates, $\frac{\partial \theta_i}{\partial \beta}$ is a $m_i \times p$ matrix of partial derivatives. By adding all the individuals,

$$U(\beta) = \sum_i U_i(\beta) \tag{30}$$

Let V_i be a working covariance matrix that has dimensions of $m_i \times m_i$ where m_i is

the number of time points individual i has. V_i can be decomposed into the following.

$$V_i = c(\phi)E_i^{1/2}R_i(\gamma)E_i^{1/2} \tag{31}$$

ϕ is a scale parameter, and it is a nuisance parameter. E_i is a $m_i \times m_i$ diagonal matrix of $V(\theta_{ij})$ with j representing the diagonal element. j also represents the jth timepoint. $R_i(\gamma)$ is a "working" correlation matrix. A "working" correlation matrix is a correlation matrix that would make it possible to get consistent estimators and variance estimators, even though the covariance structure was mispecified. In the R-package, *geepack*, independence, exchangeable, and first-order autoregressive "working" correlation matrices are available ([12]). These are three common "working" correlation matrices. Independent correlation matrix is the identity matrix. An example of an exchangeable correlation matrix is

$$\begin{bmatrix} 1 & \gamma & \gamma \\ \gamma & 1 & \gamma \\ \gamma & \gamma & 1 \end{bmatrix}$$

An example of the first-order autoregressive correlation matrix is

$$\begin{bmatrix} 1 & \gamma & \gamma^2 \\ \gamma & 1 & \gamma \\ \gamma^2 & \gamma & 1 \end{bmatrix}$$

Ultimately from (29) and (30),

$$U(\beta) = \sum_i \left(\frac{\partial \theta_i}{\partial \beta}\right)^T V_i^{-1}(\hat{\theta}_i - \theta_i) = \sum_i U_i(\beta)$$

The slope estimates, $\hat{\beta}$ can be found by solving for β for $U(\beta) = 0$. Additionally, two assumptions are made.

i) $E(U(\beta)) = 0$ where β is the vector of true values of the slope.

ii) $U_i(\beta)$ $i = 1, ..., n$ are independent.

Finally, Theorem 1 is an asymptotic result of $\hat{\beta}$. This theorem was stated by [4]. A sketch of the proof showing the asymptotic normality of GEE estimates in general was presented in [17].The complete proof of this result can be found in Appendix A.

Theorem 1 *$\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically normal with mean zero and covariance matrix estimated $\hat{\Sigma} = \mathcal{I}(\hat{\beta})^{-1}\widehat{var}(U(\beta))\mathcal{I}(\hat{\beta})^{-1}$ where $\mathcal{I}(\beta) = \sum_i \frac{\partial \theta_i}{\partial \beta}^T V_i^{-1} \frac{\partial \theta_i}{\partial \beta}$ and $\widehat{var}(U(\beta)) = \sum_i U_i(\hat{\beta})U_i(\hat{\beta})^T$. $\mathcal{I}(\beta)$ is a $p \times p$ information matrix.*

2.3 Applying to Multi-state models

To relate this to transition probabilities for non-Markov models, our parameter of interest is the transition probability. So, $\theta = P_{jk}(s, t)$. By choosing a link function g, the linear model would be

$$g(P_{ijk}(s, t)) = g(\theta_i) = \beta^T Z_i \tag{32}$$

To create a pseudo-observation $\hat{\theta}_i$, we would need an unbiased estimator, $\hat{\theta}(X)$ of the transition probability. For, non-Markov models, Aalen-Johansen estimator may be a poor choice due to it being susceptible to bias. Alternatively, [28] showed that Titman estimator is unbiased for Markov and non-Markov models. This leads us to believe that Titman estimator may be a suitable estimator as $\hat{\theta}(X)$.

In a dataset for n individuals, we will create a pseudo-observation, $\hat{\theta}_i$ for each individual. In a new dataset, we will use the pseudo-observation, $\hat{\theta}_i$ that we computed from the original dataset as the response variable and the covariates, Z_i for each

individual. Then to find the slope estimates, we will use the GEE method described in Section 2.2. Since we are interested in finding the transition probability for one time point t, we will only need one pseudo-observation. Instead of having a covariance matrix, we would only have one value for the variance.

3 Simulation Study

3.1 Non-Parametric Estimator

Now, we will create a simulation to study the feasibility of our proposed idea from Section 2.3. In this section, we will focus on the illness-death model without recovery. First, we want to verify that Titman estimator is unbiased in a non-Markov setting. We will use the same simulation settings as shown in [28]. In the simulation, the transition intensities are $\alpha_{01} = 0.12$, $\alpha_{02} = 0.03$, and $\alpha_{02} = 0.1$ for the Markov case. In the non-Markov case, the transition rates are $\alpha_{01} = 0.12$, $\alpha_{02} = 0.03$, and

$$\alpha_{12} = \begin{cases} 0.1 & if \;\; X(4) = 0 \\ 0.05 & if \;\; X(4) \neq 0 \end{cases} \tag{33}$$

The censoring distribution used are the uniform distribution (5,40) and exponential distribution with rate=.04. The sample sizes are $n = 200$ and $n = 500$ where n is the number of individuals in state 0 at time 0. The times, s and t are chosen so that they are the 15th and the 45th percentile of the distribution to time to death. In the Markov case, the times were $s = 3.79$ and $t = 10.501$. In the non-Markov case, $s = 4.674$ and $t = 12.791$. We compute Titman estimator and Aalen-Johansen estimator for 1,000 simulated datasets and take the average of each estimator. To find the bias, we would also need to know the true value of the transition probability for each transition. $bias = average\ of\ the\ estimator\ - true\ value$. The true transition probability for each transition is computed as below. For the Markov case,

$$
\begin{aligned}
P_{01}(s,t) &= \int_s^t P_{00}(s,u)\alpha_{01}(u)P_{11}(u,t)du \\
&= \int_s^t \exp\left(-\int_s^u (\alpha_{01}(v_1)+\alpha_{02}(v_1))dv_1\right)\alpha_{01}(u)\exp\left(-\int_u^t \alpha_{12}(v_2)dv_2\right)du
\end{aligned}
\tag{34}
$$

$$
\begin{aligned}
&P_{02}(s,t) = \int_s^t P_{00}(s,u)\alpha_{02}(u)du + \int_s^t\int_{u_1}^t P_{00}(s,u_1)\alpha_{01}(u_1)P_{11}(u_1,u_2)\alpha_{12}(u_2)du_2du_1 \\
&= \int_s^t \exp\left(-\int_s^u (\alpha_{01}(v_1)+\alpha_{02}(v_1))dv_1\right)\alpha_{02}(u)du \\
&+ \int_s^t\int_{u_1}^t \exp\left(-\int_s^{u_1} (\alpha_{01}(v_1)+\alpha_{02}(v_1))dv_1\right)\alpha_{01}(u_1) \\
&\times \exp\left(-\int_{u_1}^{u_2} \alpha_{12}(v_2)dv_2\right)\alpha_{12}(u_2)du_2du_1
\end{aligned}
\tag{35}
$$

$$
\begin{aligned}
P_{12}(s,t) &= \int_s^t P_{11}(s,u)\alpha_{12}(u)du \\
&\int_s^t \exp\left(-\int_s^u (\alpha_{12}(v))dv\right)\alpha_{12}(u)du
\end{aligned}
\tag{36}
$$

For the healthy to death transition, we need to consider two cases. The first case is when the individual transitions to the death state without a prior transition to the illness state. The other case is when the individual first transitions to the illness state before time t, and then transitions to the death state by time t.

For the non-Markov case, the healthy to illness transition and the healthy to death transition remain the same as the Markov case. Since $s = 4.674$, all individuals who are at the healthy state at 4.674, must also be at the healthy state at 4. The values of s and t still differ from the Markov case. For the illness to death transition, we

need to consider whether the individual becomes ill before time 4 or the individual becomes ill after 4, but before s. Based on when the individual got ill, it will affect α_{12}.

$$
\begin{aligned}
&P(X(t)=2|X(s)=1)\\
&=P(X(t)=2,X(4)=0|X(s)=1)+P(X(t)=2,X(4)\neq 0|X(s)=1)\\
&=P(X(t)=2|X(4)=0,X(s)=1)P(X(4)=0|X(s)=1)\\
&+P(X(t)=2|X(4)\neq 0,X(s)=1)P(X(4)\neq 0|X(s)=1)
\end{aligned}
$$

$$
\begin{aligned}
P_{12}(s,t)=&\int_s^t P_{11}(s,u)\alpha_{12}^*(u)du\ \ \frac{P_{00}(0,4)\int_4^s P_{00}(4,v)\alpha_{01}(v)P_{11}(v,s)dv}{\int_0^s P_{00}(0,v)\alpha_{01}(v)P_{11}(v,s)dv}\\
&+\int_s^t P_{11}(s,u)\alpha_{12}^\dagger(u)du\ \ \frac{\int_0^4 P_{00}(0,v)\alpha_{01}(v)P_{11}(v,s)dv}{\int_0^s P_{00}(0,v)\alpha_{01}(v)P_{11}(v,s)dv}
\end{aligned}
$$

$$
\begin{aligned}
=&\int_s^t \exp\left(-\int_s^u \alpha_{12}^*(w_1)dw_1\right)\alpha_{12}^*(w_1)du\exp\left(-\int_0^4(\alpha_{01}(w_2)+\alpha_{02}(w_2))dw_2\right)\\
&\times\frac{\int_4^s \exp\left(-\int_4^v(\alpha_{01}(w_3)+\alpha_{02}(w_3))dw_3\right)\alpha_{01}(v)\exp\left(-\int_v^s\alpha_{12}^*(w_4)dw_4\right)dv}{\int_0^s \exp\left(-\int_0^v(\alpha_{01}(w_3)+\alpha_{02}(w_3))dw_3\right)\alpha_{01}(v)\exp\left(-\int_v^s\alpha_{12}^*(w_4)dw_4\right)dv}\\
&+\int_s^t \exp\left(-\int_s^u \alpha_{12}^\dagger(w_1)dw_1\right)\alpha_{12}^\dagger(w_1)du\\
&\times\frac{\int_0^4 \exp\left(-\int_0^v(\alpha_{01}(w_3)+\alpha_{02}(w_3))dw_3\right)\alpha_{01}(v)\exp\left(-\int_v^s\alpha_{12}^\dagger(w_4)dw_4\right)dv}{\int_0^s \exp\left(-\int_0^v(\alpha_{01}(w_3)+\alpha_{02}(w_3))dw_3\right)\alpha_{01}(v)\exp\left(-\int_v^s\alpha_{12}^\dagger(w_4)dw_4\right)dv}
\end{aligned}
\tag{37}
$$

where $\alpha_{12}^*(u)=0.1$ and $\alpha_{12}^\dagger(u)=0.05$.

Table 1-Table 3 show the bias, standard deviation, and mean squared error(MSE) of Titman estimator and Aalen-Johansen(AJ) estimator for Markov and non-Markov

Estimator	n	Model	Censor	Bias	SD	MSE
Titman	200	Markov	Unif	.00042	.04224	.00178
Titman	200	Markov	Exp	-.00155	.05454	.00298
Titman	500	Markov	Unif	-.00033	.03136	.00098
Titman	500	Markov	Exp	-.00247	.03374	.00114
Titman	200	N-M	Unif	.00185	.05301	.00281
Titman	200	N-M	Exp	.0014	.05974	.00357
Titman	500	N-M	Unif	-.0016	.03422	.00117
Titman	500	N-M	Exp	.00154	.03871	.0015
AJ	200	Markov	Unif	.00004	.04061	.00165
AJ	200	Markov	Exp	.00013	.04822	.00233
AJ	500	Markov	Unif	-.00042	.0267	.00071
AJ	500	Markov	Exp	-.00009	.02982	.00089
AJ	200	N-M	Unif	.04847	.04419	.0043
AJ	200	N-M	Exp	.048	.05174	.00498
AJ	500	N-M	Unif	.04828	.0295	.0032
AJ	500	N-M	Exp	.04897	.03219	.00343

Table 1: Bias,Standard Deviation(SD), and MSE of Titman and Aalen-Johansen estimators(AJ) for 0→1(healthy to illness) of the Illness-Death Model without Recovery

Estimator	n	Model	Censor	Bias	SD	MSE
Titman	200	Markov	Unif	-.0004	.04389	.00193
Titman	200	Markov	Exp	.00053	.04768	.00227
Titman	500	Markov	Unif	-.00026	.02887	.00083
Titman	500	Markov	Exp	.00141	.03128	.00098
Titman	200	N-M	Unif	-.0025	.05141	.00265
Titman	200	N-M	Exp	-.0008	.05652	.0032
Titman	500	N-M	Unif	.00003	.03223	.00104
Titman	500	N-M	Exp	.00188	.03667	.00135
AJ	200	Markov	Unif	.00118	.03777	.00143
AJ	200	Markov	Exp	.00228	.04266	.00183
AJ	500	Markov	Unif	-.00167	.02359	.00056
AJ	500	Markov	Exp	-.00098	.02674	.00072
AJ	200	N-M	Unif	-.04772	.04194	.00404
AJ	200	N-M	Exp	-.04794	.04689	.0045
AJ	500	N-M	Unif	-.04778	.02669	.003
AJ	500	N-M	Exp	-.04737	.02974	.00313

Table 2: Bias, Standard Deviation(SD), and MSE of Titman and Aalen-Johansen estimators(AJ) for 0→2(healthy to death) of the Illness-Death Model without Recovery

Estimator	n	Model	Censor	Bias	SD	MSE
Titman	200	Markov	Unif	-.00056	.0697	.00486
Titman	200	Markov	Exp	.00076	.07716	.00595
Titman	500	Markov	Unif	.00128	.04368	.00191
Titman	500	Markov	Exp	-.00081	.04999	.0025
Titman	200	N-M	Unif	.00242	.06041	.00366
Titman	200	N-M	Exp	-.00183	.06811	.00464
Titman	500	N-M	Unif	.00033	.03858	.00149
Titman	500	N-M	Exp	-.00137	.04398	.00194
AJ	200	Markov	Unif	-.00011	.05394	.00291
AJ	200	Markov	Exp	.00084	.05887	.00347
AJ	500	Markov	Unif	-.00122	.03233	.00105
AJ	500	Markov	Exp	-.00011	.03753	.00141
AJ	200	N-M	Unif	.06885	.05321	.00757
AJ	200	N-M	Exp	.06538	.05471	.00727
AJ	500	N-M	Unif	.06848	.03228	.00573
AJ	500	N-M	Exp	.06648	.03717	.0058

Table 3: Bias, Standard Deviation(SD), and MSE of Titman and Aalen-Johansen estimators(AJ) for 1→2(illness to death) of the Illness-Death Model without Recovery

models for all three transitions. Note that in Chapter 3 and Chapter 4, non-Markov models will be labeled as "N-M" in the tables. As we expected, Titman estimator and Aalen-Johansen estimator are both unbiased in the Markov case. In the non-Markov case, Titman estimator remains unbiased, but Aalen-Johansen estimator is biased. In both cases, Aalen-Johansen estimator has a lower standard deviation than Titman estimator. Titman estimator requires individuals to satisfy the conditions to be in the set, $\mathcal{N}_j(s)$. Aalen-Johansen estimator only requires indiviudals to be at risk at time s. Therefore, Titman estimator considers a smaller set of individuals than Aalen-Johansen estimator. Due to Titman estimator considering a smaller set of individuals, the standard deviation is higher than Aalen-Johansen estimator.Comparing the MSE, Aalen-Johansen estimator has a smaller MSE for the Markov case and Titman estimator has a smaller MSE for the non-Markov case. Since bias was not an issue for either estimator for the Markov case, the lower standard deviation of

Aalen-Johansen estimator made its MSE smaller. For the non-Markov case, the bias of Aalen-Johansen estimator had a bigger effect than the smaller standard deviation when comparing it to Titman estimator. Overall, Aalen-Johansen estimator seems like a better choice for the Markov case, and Titman estimator seems like a better choice for the non-Markov case. This justifies our motivation in using Titman estimator as an alternative estimator in creating pseudo-observation in the non-Markov case.

3.2 Regression Using Pseudo-Observation

Now, suppose that we want to create a semi-parametric model. We simulate the transition intensities going from state j to state k as having the Cox proportional hazard form.

$$\alpha_{jk}(t|z) = \alpha_{jk0}\left(\exp(\beta^*_{jk}z)\right) \tag{38}$$

For the Markov case, the baseline transition intensities were $\alpha_{010} = 0.12$, $\alpha_{020} = 0.03$, and $\alpha_{120} = 0.1$. For the non-Markov case, $\alpha_{010} = 0.12$, $\alpha_{020} = 0.03$, and

$$\alpha_{120} = \begin{cases} 0.1 & if \;\; X(4) = 0 \\ 0.05 & if \;\; X(4) \neq 0 \end{cases} \tag{39}$$

The slope of the transition intensities were $\beta^*_{01} = \beta^*_{02} = \beta^*_{12} = 1$. The model has one binary covariate, $Z \sim Bernoulli(p = 0.5)$. We will compute the pseudo-observation using Titman estimator and Aalen-Johansen estimator. Then we will find the intercept estimate and slope estimate using GEE. We will use the normal link function. We will repeat this simulation 1,000 times and average the intercept estimate and slope estimate. We will compare them with the true value of the intercept and slope,

respectively. We will let $s = 2$ and $t = 6$.

It is difficult to calculate analytically the true transition probability, $P_{jk}(s,t)$ in the multi-state model with a covariate. [13] in their study used a large sample and created an empirical distribution to represent the true population. In our case, we will also use a large sample to find the proportion of individuals who transitioned from state j to state k between time s and time t. We sample 100,000 individuals that would represent the population. Then from the large sample, we will compute the empirical proportion that would represent the true transition probability. For example, for the healthy to illness transition, the empirical proportion will be considered the true transition probability, $P_{01}(s,t)$.

Let the indicator variable be $I(X(t) = 1|X(s) = 0)$. This represents whether the individual transitioned to the illness state by time t given that the individual is healthy at time s. The proportion will be $\frac{\sum_{i=1}^{n^*} I(X(t)=1|X(s)=0)}{n^*}$ where n^* are all of the individuals out of the 100,000 individuals who are healthy at time s. The proportions can be constructed similarly for the other two transitions. Then, we will create a pseudo-observation using the empirical proportion, and obtain the intercept estimate and slope estimate using GEE. We will compare the results of the intercept estimate and slope estimate from Titman estimator and Aalen-Johansen estimator with the results of the intercept and slope from the empirical proportion.

Table 4, Table 6, and Table 8 show the bias and the standard deviation of the intercept estimate and the slope estimate using the pseudo-observation method for healthy to illness, healthy to death, and illness to death transitions, respectively. Table 5, Table 7, and Table 9 show the bias and the standard deviation of the predicted transition probability when $Z = 1$. They also show the MSE for the intercept estimate, the slope estimate, and the predicted transition probability when $Z = 1$ for healthy to illness, healthy to death, and illness to death transitions, respectively. For the Markov case,

Estimator	n	Model	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	Markov	Unif	.0046	-.0042	.05893	.08394
Titman	200	Markov	Exp	.0057	-.0082	.06488	.09019
Titman	500	Markov	Unif	.0044	-.0055	.03772	.05332
Titman	500	Markov	Exp	.0051	-.0055	.04225	.0588
Titman	200	N-M	Unif	-.0063	.0033	.0629	.0903
Titman	200	N-M	Exp	-.0056	.0023	.0686	.10196
Titman	500	N-M	Unif	-.0049	-.0029	.03701	.0551
Titman	500	N-M	Exp	-.0086	.0003	.04417	.06011
AJ	200	Markov	Unif	.0018	-.0033	.05324	.07089
AJ	200	Markov	Exp	.0009	-.0055	.05761	.07959
AJ	500	Markov	Unif	.0019	-.0042	.03396	.04596
AJ	500	Markov	Exp	.0015	-.0032	.03571	.05042
AJ	200	N-M	Unif	.0525	-.1113	.05666	.07921
AJ	200	N-M	Exp	.0584	-.1154	.06175	.08471
AJ	500	N-M	Unif	.0547	-.1163	.03559	.05072
AJ	500	N-M	Exp	.0552	-.113	.03825	.05478

Table 4: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of 0→1(healthy to illness)

Estimator	n	Model	Censor	Bias $\hat{P}_{01}$	SD $\hat{P}_{01}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{01}$
Titman	200	Markov	Unif	.0005	.06299	.00349	.00706	.00397
Titman	200	Markov	Exp	-.0025	.06602	.00424	.0082	.00437
Titman	500	Markov	Unif	-.0011	.03813	.00144	.00287	.00146
Titman	500	Markov	Exp	-.0004	.04222	.00181	.00349	.00178
Titman	200	N-M	Unif	-.0029	.06393	.004	.00817	.0041
Titman	200	N-M	Exp	-.0033	.0744	.00474	.0104	.00555
Titman	500	N-M	Unif	-.0078	.04109	.00139	.00304	.00175
Titman	500	N-M	Exp	-.0083	.04406	.00203	.00361	.00201
AJ	200	Markov	Unif	-.0015	.05182	.00284	.00504	.00269
AJ	200	Markov	Exp	-.0046	.059434	.00332	.00637	.00355
AJ	500	Markov	Unif	-.0023	.03255	.00116	.00213	.00107
AJ	500	Markov	Exp	-.0017	.03688	.00128	.00255	.00136
AJ	200	N-M	Unif	-.0588	.05536	.00597	.01866	.00652
AJ	200	N-M	Exp	-.057	.05991	.00722	.02049	.00684
AJ	500	N-M	Unif	-.0616	.03549	.00426	.0161	.00505
AJ	500	N-M	Exp	-.0578	.03852	.00451	.01577	.00483

Table 5: Bias and Standard Deviation(SD) of $\hat{P}_{01}(s,t|Z=1)$, and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{P}_{01}(s,t|Z=1)$ using Pseudo-Observation Method of 0→1(healthy to illness)

Estimator	n	Model	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	Markov	Unif	.0023	-.0039	.04657	.08285
Titman	200	Markov	Exp	.0044	-.0063	.05182	.09175
Titman	500	Markov	Unif	.004	-.0079	.02962	.05326
Titman	500	Markov	Exp	.0042	-.0053	.03323	.0568
Titman	200	N-M	Unif	-.0019	.0007	.0441	.07764
Titman	200	N-M	Exp	-.0007	-.0016	.05037	.08255
Titman	500	N-M	Unif	-.0008	-.0014	.02923	.05154
Titman	500	N-M	Exp	.0006	-.002	.03065	.05479
AJ	200	Markov	Unif	.0064	-.0068	.04439	.07247
AJ	200	Markov	Exp	.0053	-.0011	.04805	.07591
AJ	500	Markov	Unif	.0065	-.0061	.02718	.04432
AJ	500	Markov	Exp	.0053	-.0035	.03006	.0465
AJ	200	N-M	Unif	.1117	-.2164	.04612	.06523
AJ	200	N-M	Exp	.1138	-.2175	.05121	.07214
AJ	500	N-M	Unif	.107	-.2176	.03003	.04293
AJ	500	N-M	Exp	.1074	-.2162	.03169	.04525

Table 6: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of 0→2(healthy to death)

Estimator	n	Model	Censor	Bias $\hat{P}_{02}$	SD of $\hat{P}_{02}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{02}$
Titman	200	Markov	Unif	-.0015	.06703	.00217	.00688	.0045
Titman	200	Markov	Exp	-.0019	.07302	.00271	.00846	.00534
Titman	500	Markov	Unif	-.0039	.04335	.00089	.0029	.00189
Titman	500	Markov	Exp	-.0011	.04539	.00112	.00325	.00206
Titman	200	N-M	Unif	-.0011	.05982	.00195	.00603	.00358
Titman	200	N-M	Exp	-.0023	.0663	.00254	.00682	.0044
Titman	500	N-M	Unif	-.0022	.03973	.00086	.00266	.00158
Titman	500	N-M	Exp	-.0014	.04416	.00094	.00301	.00195
AJ	200	Markov	Unif	-.0004	.05681	.00201	.0053	.00323
AJ	200	Markov	Exp	.0043	.05927	.00234	.00576	.00353
AJ	500	Markov	Unif	.0004	.03418	.00078	.002	.00117
AJ	500	Markov	Exp	.0017	.03729	.00093	.00218	.00139
AJ	200	N-M	Unif	-.1046	.04531	.0146	.05108	.01299
AJ	200	N-M	Exp	-.1037	.04998	.01557	.05251	.01325
AJ	500	N-M	Unif	-.1106	.02895	.01235	.04919	.01307
AJ	500	N-M	Exp	-.1089	.0322	.01254	.04879	.0129

Table 7: Bias and Standard Deviation(SD) of $\hat{P}_{02}(s,t|Z=1)$, and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{P}_{02}(s,t|Z=1)$ using Pseudo-Observation Method of 0→2(healthy to death)

Estimator	n	Model	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	Markov	Unif	-.0066	.0048	.0997	.13696
Titman	200	Markov	Exp	-.0042	.0108	.10518	.14381
Titman	500	Markov	Unif	-.0023	.0052	.06114	.08239
Titman	500	Markov	Exp	-.0047	.0074	.06587	.08867
Titman	200	N-M	Unif	-.0007	.0001	.07208	.11197
Titman	200	N-M	Exp	.0005	.003	.07885	.117
Titman	500	N-M	Unif	.001	.0034	.04757	.06878
Titman	500	N-M	Exp	-.0005	.0023	.05056	.07075
AJ	200	Markov	Unif	-.0108	.0118	.07287	.09762
AJ	200	Markov	Exp	-.0101	.0101	.07403	.09889
AJ	500	Markov	Unif	-.011	.0094	.0444	.05986
AJ	500	Markov	Exp	-.012	.0084	.04776	.06702
AJ	200	N-M	Unif	.1189	-.2098	.06962	.10253
AJ	200	N-M	Exp	.1143	-.2127	.07336	.10842
AJ	500	N-M	Unif	.1126	-.215	.0442	.06373
AJ	500	N-M	Exp	.1077	-.211	.04754	.06651

Table 8: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of 1→2(illness to death)

Estimator	n	Model	Censor	Bias $\hat{P}_{12}$	SD $\hat{P}_{12}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{12}$
Titman	200	Markov	Unif	-.0018	.0991	.00998	.01878	.00982
Titman	200	Markov	Exp	.0066	.10014	.01108	.0208	.01007
Titman	500	Markov	Unif	.0029	.05812	.00374	.00682	.00339
Titman	500	Markov	Exp	.0026	.06469	.00436	.00792	.00419
Titman	200	N-M	Unif	-.0007	.0928	.0052	.01254	.00861
Titman	200	N-M	Exp	.0036	.09863	.00622	.0137	.00974
Titman	500	N-M	Unif	.0043	.0559	.00226	.00474	.00314
Titman	500	N-M	Exp	.0018	.05955	.00256	.00501	.00355
AJ	200	Markov	Unif	.001	.07124	.00543	.00967	.00508
AJ	200	Markov	Exp	<.0001	.07697	.00558	.00988	.00592
AJ	500	Markov	Unif	-.0016	.04419	.00209	.00367	.00196
AJ	500	Markov	Exp	-.0035	.0487	.00243	.00456	.00238
AJ	200	N-M	Unif	-.0909	.07043	.01898	.05453	.01322
AJ	200	N-M	Exp	-.0984	.07944	.01845	.057	.01599
AJ	500	N-M	Unif	-.1023	.04459	.01463	.05029	.01245
AJ	500	N-M	Exp	-.1033	.04761	.01386	.04895	.01294

Table 9: Bias and Standard Deviation(SD) of $\hat{P}_{12}(s,t|Z=1)$, and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{P}_{12}(s,t|Z=1)$ using Pseudo-Observation Method of 1→2(illness to death)

the results are consistent to what we observed in Section 3.1. The bias is minimal when using either estimator. The standard deviation is smaller for Aalen-Johansen estimator, and this leads to a smaller MSE when using Aalen-Johansen estimator. For this reason, we believe that for the Markov case, the pseudo-observation method using Aalen-Johansen estimator is better.

For the non-Markov case, the results are what we conjectured in Section 2.3. The bias when using Titman estimator is minimal, but Aalen-Johansen estimator is quite biased. The slope estimate for Aalen-Johansen estimator is close to 0, so higher the magnitude of the true slope, the higher the bias became when using Aalen-Johansen estimator. The magnitude of healthy to death and illness to death transitions were quite high. We believe that the biased Aalen-Johansen estimator is heavily influencing the intercept and has little to no effect to the slope which is close to 0. The standard deviation is smaller when using Aalen-Johansen estimator than when using Titman estimator. Due to the large contribution of the bias in the MSE when using Aalen-Johansen estimator, the MSE is higher when using Aalen-Johansen estimator. Therefore, for the non-Markov case, we believe using Titman estimator is better.

We noticed that the standard devation in the illness to death transition is concerningly high. This is particularly the case when using Titman estimator. We believe that this is because the landmark subsample of only ill individuals is low. This can be a problem for small sample sizes. We see improvement when $n = 500$ compared to when $n = 200$. We recommend practitioners to have larger sample sizes when studying illness to death transitions.

We will study the asymptotic result of Theorem 1 using a simulation. We will show the results of the illness to death transition of the non-Markov case. The simulation was based on 1,000 datasets. Figure 2 and Figure 3 show the histograms and Normal Q-Q plots for the slope, respectively for various sample sizes. Table 10

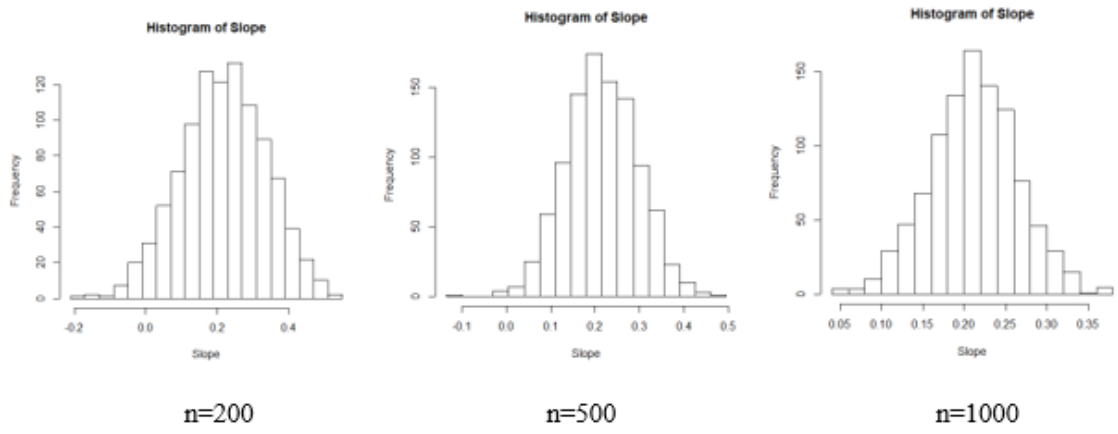

Figure 2: Histogram of the Slope Estimate using Pseudo-Observation Method with Titman Estimator for Illness-Death Model without Recovery

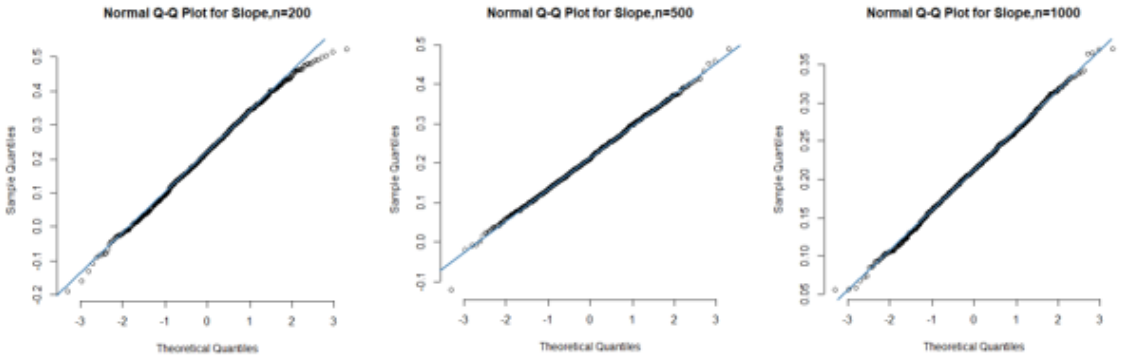

Figure 3: Normal QQ-Plot of the Slope Estimate using Pseudo-Observation Method with Titman Estimator for Illness-Death Model without Recovery

shows the results of Shapiro-Wilk test for normality of the slope for various sample sizes. We can see that for smaller n the histogram looks slightly skewed. We can see a slight departure on the upper-tail of the Normal Q-Q plot, particularly when $n = 200$. However, as n gets larger, the shape of the histogram looks closer to a normal distribution, and the departure in the Normal Q-Q plot mitigates. When n

n	P-value
200	.0983
350	.1338
500	.8293
700	.7468
1000	.7209

Table 10: Shapiro-Wilk Test for Normality of the Slope using Pseudo-Observation Method with Titman Estimator for Illness-Death Model without Recovery

gets larger, p-value is clearly insignificant for Shapiro-Wilk test. Overall, the results from the histograms, Normal Q-Q plots, and Shapiro-Wilk tests verify the asymptotic normality of Theorem 1.

4 Illness-Death Model With Recovery

4.1 Introduction

Consider the three-state illness-death model, but with the possibility of transitioning from the illness state to the healthy state. This means that it is possible for individuals to recover from the illness. See Figure 4. This multi-state model is also practical to medical practitioners since many individuals diagnosed with a disease recover. We assume individuals may get ill and recover more than one time. That is, there is no restriction in the number of times an individual becomes ill or recovers.

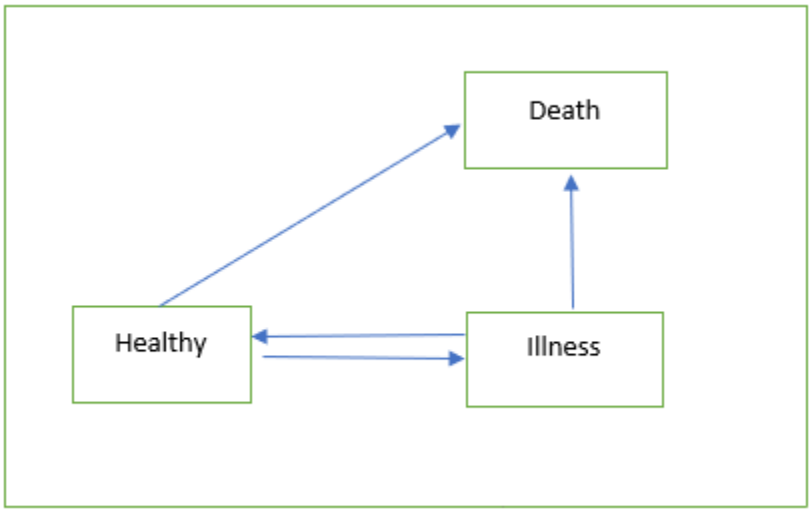

Figure 4: Illness-Death Model with Recovery

There will be four transitions. From [28], the competing risk process, $C_s(u)$ remains the same for the healthy to illness, the healthy to death, and the illness to death transitions. For the transition of illness to healthy, the states, $\{0, 1\} \notin \{\mathcal{R}_k \cup \mathcal{A}_k\}$. $2 \in \mathcal{R}_k$. Then the competing risk model can be simplified to

$$C_s(u) = \begin{cases} 0 & if\ X(u) \in \{0,1\} \\ 2 & if\ X(u) = 2 \end{cases}$$

Then the estimator will be

$$\hat{P}_{10}(s,t) = \hat{P}(C_s(t) = 0|X(s) = 1)\hat{P}(X(t) = 0|C_s(t) = 0, X(s) = 1)$$

$$= \hat{F}_0(t)\hat{p}_{0|1}(t) \tag{40}$$

where

$$\hat{F}_0(t) = \hat{P}(C_s(t) = 0|X(s) = 1) = \prod_{v\in[s,t]} \left(1 - \frac{d_s N(v)}{{}_s\bar{Y}(v)}\right)$$

$$\hat{p}_{0|1}(t) = \hat{P}(X(t) = 0|C_s(t) = 0, X(s) = 1)$$
$$= \frac{\sum I(X(s) = 1\ , X(t) = 0\ ,\ {}_sY(s) = 1\ ,\ {}_sY(t) = 1)}{\sum I(X(s) = 1\ ,\ {}_sY(s) = 1\ ,\ {}_sY(t) = 1)}$$

The landmark subsample will only consist of individuals who are ill at time s.

4.2 Non-Parametric Estimator

Similar to the illness-death model without recovery, [28] showed that Titman estimator is unbiased for the Markov case and the non-Markov case. Similar to Section 3.1, we want to verify that Titman estimator is unbiased in a non-Markov setting in the illness-death model with recovery. We will use the same transition rates that [28] used which are different from Section 3.1. For the Markov case, $\alpha_{01} = 0.5$, $\alpha_{02} = 0.02$, $\alpha_{12} = 0.1$, and $\alpha_{10} = 0.3$. For the non-Markov case, all of the transition rates remain the same as the Markov case except α_{01}.

$$\alpha_{01} = \begin{cases} 0.3 & if \;\; X(4) = 1 \\ 0.5 & if \;\; X(4) \neq 1 \end{cases} \tag{41}$$

In addition, we will let $\alpha_{01} = 0.5$ for time before time 4. s and t are the 15th and 45th percentile of the time to death distribution. In the Markov case, $s = 3.1927$ and $t = 9.6482$. In the non-Markov case, $s = 3.1922$ and $t = 9.9237$. The other settings of the simulation remain the same as the simulation from Section 3.1.

It is difficult to compute the true transition probability as we did in Section 3.1. This is because there are no restriction in the number of times an individual may get ill or recover. Thus, an individual may transition between the healthy state and the illness state arbitrary number of times. We will compute the true transition probability using matrix exponentials involving Kolmogorov's forward and backward equations discussed in [14].

Suppose that we have a transition rate matrix, A. Also, consider the transition probability matrix, $P(t)$. In our study, we have s and t. [14] assumes that $s = 0$, but s does not have be 0. In our simulations, $s > 0$. We will let t^* be the length of time between s and t. Consider $P(t^*) = \exp(At^*)$. $P(t^*)$ is the solution to the Kolmogorov's forward equation $P^{'}(t^*) = P(t^*)A$, and Kolmogorov's backward equations $P^{'}(t^*) = AP(t^*)$. The initial condition, $P(0)$ is the identity matrix. We assume that $\sum_{k=0}^{2} \alpha_{jk} = 0$ for $j = 0, 1, 2$. Thus, we have

$$\frac{dP(t^*)}{dt^*} = \frac{d\exp(At^*)}{dt^*} = \exp(At^*)A = P(t^*)A \tag{42}$$

$$\frac{dP(t^*)}{dt^*} = \frac{d\exp(At^*)}{dt^*} = A\exp(At^*) = AP(t^*) \tag{43}$$

$\exp(At^*)$ is called a matrix exponential. $\exp(At^*)$ can be expressed as an infinite

sum. That is, $P(t^*) = \exp(At^*) = \sum_{r=0}^{\infty} A^r \frac{(t^*)^r}{r!}$. By using spectral decomposition, $A = U^* D U^{*-1}$ where U^* is a matrix of eigenvectors. D is a diagonal matrix of eigenvalues. By using a property of the matrix exponential, $\exp(At^*)$ can be written as

$$\exp(At^*) = U^* \exp(Dt^*) U^{*-1} \tag{44}$$

Applying the matrix exponential to the illness-death model with recovery, we have

$$A = \begin{bmatrix} \alpha_{00} & \alpha_{01} & \alpha_{02} \\ \alpha_{10} & \alpha_{11} & \alpha_{12} \\ 0 & 0 & 0 \end{bmatrix}$$

where $\alpha_{00} = -\alpha_{01} - \alpha_{02}$ and $\alpha_{11} = -\alpha_{10} - \alpha_{12}$. Let $\xi = \sqrt{(\alpha_{00} - \alpha_{01})^2 + 4\alpha_{01}\alpha_{10}}$.

$$D = \begin{bmatrix} \frac{\alpha_{00}+\alpha_{11}+\xi}{2} & 0 & 0 \\ 0 & \frac{\alpha_{00}+\alpha_{11}-\xi}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where the eigenvalues are $\frac{\alpha_{00}+\alpha_{11}+\xi}{2}$, $\frac{\alpha_{00}+\alpha_{11}-\xi}{2}$, and 0.

$$U^* = \begin{bmatrix} 2\alpha_{01} & 2\alpha_{01} & 1 \\ -\alpha_{00} + \alpha_{11} + \xi & -\alpha_{00} + \alpha_{11} - \xi & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

In the Markov case, we can compute the transition probabilites using (44). In the non-Markov case, it is more complicated. We will need two transition rate matrices, A_1 and A_2. This is because α_{01} varies based on whether $X(4) \neq 1$ or $X(4) = 1$. Let A_1 be the transition rate matrix when $X(4) \neq 1$ and A_2 be the transition rate matrix

when $X(4) = 1$. Since we assume $\alpha_{01} = 0.5$ before time 4 and $\alpha_{01} = 0.5$ when $X(4) \neq 1$, we will use A_1 for time before time 4. We will define the transition probability matrix, $P^{(1)}(t^*)$ and $P^{(2)}(t^*)$ in matrix exponential form as $P^{(1)}(t^*) = \exp(A_1(t^*))$ and $P^{(2)}(t^*) = \exp(A_2(t^*))$. The matrix elements are denoted as $P^{(1)}_{jk}(t^*)$ and $P^{(2)}_{jk}(t^*)$ for $j = 0, 1$ and $k = 0, 1, 2$. By using matrix exponential, we will derive the formula of the transition probabilities for each transition. The numerical calculations were done in R.

$$\begin{aligned}
P_{01}(s,t) &= P(X(t) = 1|X(s) = 0) \\
&= P(X(t) = 1, X(4) = 1|X(s) = 0) + P(X(t) = 1, X(4) \neq 1|X(s) = 0) \\
&= P(X(4) = 1|X(s) = 0)P(X(t) = 1|X(4) = 1, X(s) = 0) \\
&+ P(X(4) \neq 1|X(s) = 0)P(X(t) = 1|X(4) \neq 1, X(s) = 0) \\
&= P^{(1)}_{01}(4-s)P^{(2)}_{11}(t-4) + P^{(1)}_{00}(4-s)P^{(1)}_{01}(t-4) \qquad (45)
\end{aligned}$$

$$\begin{aligned}
P_{02}(s,t) &= P(X(t) = 2|X(s) = 0) \\
&= P(X(t) = 2, X(4) = 1|X(s) = 0) + P(X(t) = 2, X(4) \neq 1|X(s) = 0) \\
&= P(X(t) = 2, X(4) = 1|X(s) = 0) \\
&+ P(X(t) = 2, (X(4) = 0 \cup X(4) = 2)|X(s) = 0) \\
&= P(X(t) = 2, X(4) = 1|X(s) = 0) \\
&+ P((X(t) = 2, X(4) = 0) \cup (X(t) = 2, X(4) = 2)|X(s) = 0)
\end{aligned}$$

Note that $X(4) = 0$ and $X(4) = 2$ are disjoint events.

$$
\begin{aligned}
&= P(X(t) = 2, X(4) = 1 | X(s) = 0) \\
&+ P((X(t) = 2, X(4) = 0 | X(s) = 0) \\
&+ P(X(t) = 2, X(4) = 2) | X(s) = 0) \\
&= P(X(4) = 1 | X(s) = 0) P(X(t) = 2 | X(4) = 1, X(s) = 0) \\
&+ P(X(4) = 0 | X(s) = 0) P(X(t) = 2 | X(4) = 0, X(s) = 0) \\
&+ P(X(4) = 2 | X(s) = 0) P(X(t) = 2 | X(4) = 2, X(s) = 0)
\end{aligned}
$$

$$
= P_{01}^{(1)}(4-s) P_{12}^{(2)}(t-4) + P_{00}^{(1)}(4-s) P_{02}^{(1)}(t-4) + P_{02}^{(1)}(4-s) \tag{46}
$$

$$
\begin{aligned}
P_{12}(s,t) &= P(X(t) = 2 | X(s) = 1) \\
&= P(X(t) = 2, X(4) = 1 | X(s) = 1) + P(X(t) = 2, X(4) \neq 1 | X(s) = 1) \\
&= P(X(t) = 2, X(4) = 1 | X(s) = 1) \\
&+ P(X(t) = 2, (X(4) = 0 \cup X(4) = 2) | X(s) = 1) \\
&= P(X(t) = 2, X(4) = 1 | X(s) = 1) \\
&+ P((X(t) = 2, X(4) = 0) \cup (X(t) = 2, X(4) = 2) | X(s) = 1)
\end{aligned}
$$

Note that $X(4) = 0$ and $X(4) = 2$ are disjoint events.

$$
\begin{aligned}
&= P(X(t) = 2, X(4) = 1 | X(s) = 1) \\
&+ P((X(t) = 2, X(4) = 0 | X(s) = 1) \\
&+ P(X(t) = 2, X(4) = 2) | X(s) = 1)
\end{aligned}
$$

$$
\begin{aligned}
&= P(X(4)=1|X(s)=1)P(X(t)=2|X(4)=1,X(s)=1) \\
&+ P(X(4)=0|X(s)=1)P(X(t)=2|X(4)=0,X(s)=1) \\
&+ P(X(4)=2|X(s)=1)P(X(t)=2|X(4)=2,X(s)=1) \\
&= P_{11}^{(1)}(4-s)P_{12}^{(2)}(t-4) + P_{10}^{(1)}(4-s)P_{02}^{(1)}(t-4) + P_{12}^{(1)}(4-s) \qquad (47)
\end{aligned}
$$

$$
\begin{aligned}
P_{10}(s,t) &= P(X(t)=0|X(s)=1) \\
&= P(X(t)=0,X(4)=1|X(s)=1) + P(X(t)=0,X(4)\neq 1|X(s)=1) \\
&= P(X(4)=1|X(s)=1)P(X(t)=0|X(4)=1,X(s)=1) \\
&+ P(X(4)\neq 1|X(s)=1)P(X(t)=0|X(4)\neq 1,X(s)=1) \\
&= P_{11}^{(1)}(4-s)P_{10}^{(2)}(t-4) + P_{10}^{(1)}(4-s)P_{00}^{(1)}(t-4) \qquad (48)
\end{aligned}
$$

Table 11-14 show the bias, standard deviation, and MSE of Titman estimator and Aalen-Johansen estimator for each transition. Titman estimator is unbiased for all transitions for Markov and non-Markov case. Aalen-Johansen estimator is unbiased for all transitions for Markov case. For non-Markov case of Aalen-Johansen estimator, the bias is under .01 in the healthy to death and the illness to death transitions. This can be close in being considered unbiased. For the healthy to illness and the illness to healthy transitions, the bias is approximately .02. Thus, Aalen-Johansen estimator has minimal bias in those transitions.This leads us to believe that we may get reasonable results for the GEE estimates of the intercept and slope when using Aalen-Johansen estimator to create the pseudo-observation in the non-Markov case.

Estimator	n	Model	Censor	Bias	SD	MSE
Titman	200	Markov	Unif	-.00087	.05889	.00347
Titman	200	Markov	Exp	.0004	.06893	.00475
Titman	500	Markov	Unif	-.00213	.0374	.0014
Titman	500	Markov	Exp	.00086	.04191	.00176
Titman	200	N-M	Unif	-.00042	.06269	.00393
Titman	200	N-M	Exp	-.00028	.06806	.00463
Titman	500	N-M	Unif	-.00032	.03862	.00149
Titman	500	N-M	Exp	-.00123	.04194	.00176
AJ	200	Markov	Unif	-.00186	.03974	.00158
AJ	200	Markov	Exp	-.00215	.04553	.00208
AJ	500	Markov	Unif	-.00174	.02421	.00059
AJ	500	Markov	Exp	.00113	.02899	.00084
AJ	200	N-M	Unif	-.02056	.03826	.00189
AJ	200	N-M	Exp	-.0204	.04455	.0024
AJ	500	N-M	Unif	-.01957	.02695	.00111
AJ	500	N-M	Exp	-.01916	.02837	.00117

Table 11: Bias, Standard Deviation(SD), and MSE of Titman and Aalen-Johansen estimators(AJ) for 0→1(healthy to illness) of the Illness-Death Model with Recovery

Estimator	n	Model	Censor	Bias	SD	MSE
Titman	200	Markov	Unif	.00035	.05328	.00284
Titman	200	Markov	Exp	-.00025	.06381	.00407
Titman	500	Markov	Unif	-.00092	.03441	.00119
Titman	500	Markov	Exp	-.00094	.03877	.0015
Titman	200	N-M	Unif	-.0022	.0583	.0034
Titman	200	N-M	Exp	-.00174	.06005	.00361
Titman	500	N-M	Unif	.0004	.03647	.00133
Titman	500	N-M	Exp	-.00134	.03882	.00151
AJ	200	Markov	Unif	.00043	.03753	.00141
AJ	200	Markov	Exp	.00197	.0407	.00166
AJ	500	Markov	Unif	.00081	.02359	.00056
AJ	500	Markov	Exp	-.00154	.02614	.00069
AJ	200	N-M	Unif	-.00732	.03707	.00143
AJ	200	N-M	Exp	-.00844	.04038	.0017
AJ	500	N-M	Unif	-.00706	.02429	.00064
AJ	500	N-M	Exp	-.00903	.02632	.00077

Table 12: Bias, Standard Deviation(SD), and MSE of Titman and Aalen-Johansen estimators(AJ) for 0→2(healthy to death) of the Illness-Death Model with Recovery

Estimator	n	Model	Censor	Bias	SD	MSE
Titman	200	Markov	Unif	.00015	.05088	.00259
Titman	200	Markov	Exp	.00081	.05408	.00293
Titman	500	Markov	Unif	.00145	.03339	.00112
Titman	500	Markov	Exp	-.00199	.03657	.00134
Titman	200	N-M	Unif	.00099	.05137	.00264
Titman	200	N-M	Exp	.00003	.05542	.00307
Titman	500	N-M	Unif	-.00055	.03226	.00104
Titman	500	N-M	Exp	-.00002	.03646	.00133
AJ	200	Markov	Unif	.00077	.04038	.00163
AJ	200	Markov	Exp	.00249	.04264	.00182
AJ	500	Markov	Unif	.00058	.02466	.00061
AJ	500	Markov	Exp	<.00001	.02778	.00077
AJ	200	N-M	Unif	.00719	.04026	.00167
AJ	200	N-M	Exp	.00647	.04503	.00207
AJ	500	N-M	Unif	.00637	.02539	.00069
AJ	500	N-M	Exp	.00603	.02675	.00075

Table 13: Bias, Standard Deviation(SD), and MSE of Titman and Aalen-Johansen estimators(AJ) for 1→2(illness to death) of the Illness-Death Model with Recovery

Estimator	n	Model	Censor	Bias	SD	MSE
Titman	200	Markov	Unif	.00249	.04947	.00245
Titman	200	Markov	Exp	.00242	.05391	.00291
Titman	500	Markov	Unif	-.00119	.02965	.00088
Titman	500	Markov	Exp	-.00077	.03573	.00128
Titman	200	N-M	Unif	-.00233	.05169	.00268
Titman	200	N-M	Exp	.00031	.0589	.00347
Titman	500	N-M	Unif	-.00013	.03205	.00103
Titman	500	N-M	Exp	.00155	.03614	.00131
AJ	200	Markov	Unif	.00249	.0336	.00114
AJ	200	Markov	Exp	.00101	.03803	.00145
AJ	500	Markov	Unif	-.00069	.02178	.00048
AJ	500	Markov	Exp	.00107	.02415	.00058
AJ	200	N-M	Unif	-.02182	.03689	.00184
AJ	200	N-M	Exp	-.02385	.04102	.00225
AJ	500	N-M	Unif	-.0222	.02353	.00105
AJ	500	N-M	Exp	-.02283	.02723	.00126

Table 14: Bias, Standard Deviation(SD), and MSE of Titman and Aalen-Johansen estimators(AJ) for 1→0(illness to healthy) of the Illness-Death Model with Recovery

Similar to the illness-death model without recovery, the standard deviation of Aalen-Johansen estimator is lower than Titman estimator. This is due to Titman estimator having a smaller set of individuals due to landmarking. While Aalen-Johansen estimator has mild bias in some transitions, MSE of Aalen-Johansen estimator is lower than MSE of Titman estimator. This is due to the lower standard deviation of Aalen-Johansen estimator. Thus, Aalen-Johansen estimator has the potential of being the superior estimator to use in the pseudo-observation method when creating a semi-parametric model.

4.3 Regression Using Pseudo-Observation

Now, suppose that we want to create a semi-parametric model. Similar to Section 3.2, we simulate the transition intensities going from state j to state k as having the Cox proportional hazard form. We will have one binary covariate, $Z \sim Bernoulli(p = 0.5)$. For the Markov case, the baseline transition intensities for the transitions were $\alpha_{010} = 0.5$, $\alpha_{020} = 0.02$, $\alpha_{120} = 0.1$, and $\alpha_{100} = 0.3$. For the non-Markov case, $\alpha_{020} = 0.02$, $\alpha_{120} = 0.1$, $\alpha_{100} = 0.3$, but

$$\alpha_{010} = \begin{cases} 0.3 & if \;\; X(4) = 1 \\ 0.5 & if \;\; X(4) \neq 1 \end{cases} \tag{49}$$

The rest of the simulation settings are similar to when we did the semi-parametric study in Section 3.2. This includes letting $s = 2$ and $t = 6$. We will create the pseudo-observation using Titman estimator and Aalen-Johansen estimator. By using GEE, we will create a regression model of the pseudo-observation on the covariate.

Due to difficulty in calculating the true transition probability analytically, we will compute the empirical proportion of individuals who transitioned from state

j to state k between time s and time t. We will sample 100,000 individuals that would represent the true population.The empirical proportion will represent the true transition probability. Ultimately, we will create the pseudo-observation using the empirical proportion and find the intercept and slope estimates using GEE. We will compare the GEE estimates when using Titman estimator, Aalen-Johansen estimator, to the GEE estimates of the empirical proportion.

Estimator	n	Model	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	Markov	Unif	.0006	-.0014	.08043	.10524
Titman	200	Markov	Exp	.0047	.0006	.08535	.11833
Titman	500	Markov	Unif	.0051	-.004	.05143	.06956
Titman	500	Markov	Exp	.0042	-.0036	.05486	.07488
Titman	200	N-M	Unif	-.0037	.0088	.07935	.10446
Titman	200	N-M	Exp	-.0039	.0102	.0878	.11786
Titman	500	N-M	Unif	-.002	.0087	.05122	.0667
Titman	500	N-M	Exp	-.0057	.0108	.05772	.076
AJ	200	Markov	Unif	.0118	-.0178	.05713	.07634
AJ	200	Markov	Exp	.0107	-.0192	.06171	.08436
AJ	500	Markov	Unif	.0105	-.016	.03763	.05185
AJ	500	Markov	Exp	.0094	-.0166	.04107	.05246
AJ	200	N-M	Unif	-.0075	.0079	.05736	.07727
AJ	200	N-M	Exp	-.0055	.0017	.06278	.08375
AJ	500	N-M	Unif	-.0038	.0025	.0351	.04785
AJ	500	N-M	Exp	-.0074	.0066	.04021	.05342

Table 15: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of 0→1(healthy to illness)

Table 15, Table 17, Table 19, and Table 21 show the bias and the standard deviation of the intercept and slope estimates using the pseudo-observation method for the healthy to illness, the healthy to death, the illness to death, and the illness to healthy transitions, respectively. Table 16, Table 18, Table 20, and Table 22 show the bias and the standard deviation of the predicted transition probability when $Z = 1$. They also show the MSE of the intercept estimate,the slope estimate, and the predicted

Estimator	n	Model	Censor	Bias $\hat{P}_{01}$	SD $\hat{P}_{01}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{01}$
Titman	200	Markov	Unif	-.0008	.07388	.00647	.01108	.00546
Titman	200	Markov	Exp	.0053	.08331	.00731	.014	.00697
Titman	500	Markov	Unif	.0011	.04664	.00267	.00486	.00218
Titman	500	Markov	Exp	.0007	.05203	.00303	.00562	.00271
Titman	200	N-M	Unif	.005	.0739	.00631	.01099	.00549
Titman	200	N-M	Exp	.0064	.08322	.00772	.014	.00697
Titman	500	N-M	Unif	.0067	.0472	.00263	.00453	.00227
Titman	500	N-M	Exp	.0052	.05105	.00336	.00589	.00263
AJ	200	Markov	Unif	-.006	.05256	.0034	.00615	.0028
AJ	200	Markov	Exp	-.0085	.05634	.00392	.00749	.00325
AJ	500	Markov	Unif	-.0055	.03398	.00153	.00294	.00119
AJ	500	Markov	Exp	-.0073	.03482	.00178	.00303	.00127
AJ	200	N-M	Unif	.0004	.05081	.00335	.00603	.00258
AJ	200	N-M	Exp	-.0038	.05552	.00397	.00702	.0031
AJ	500	N-M	Unif	-.0013	.03271	.00125	.0023	.00107
AJ	500	N-M	Exp	-.0008	.03641	.00167	.0029	.00133

Table 16: Bias and Standard Deviation(SD) of $\hat{P}_{01}(s,t|Z=1)$, and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{P}_{01}(s,t|Z=1)$ using Pseudo-Observation Method of 0→1(healthy to illness)

Estimator	n	Model	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	Markov	Unif	-.0026	.004	.06335	.10558
Titman	200	Markov	Exp	-.0005	.0025	.0664	.11381
Titman	500	Markov	Unif	.0009	.0042	.04102	.06738
Titman	500	Markov	Exp	-.0014	.0069	.04444	.0748
Titman	200	N-M	Unif	.0013	-.0078	.061	.10155
Titman	200	N-M	Exp	.0054	-.011	.07088	.11096
Titman	500	N-M	Unif	.003	-.011	.04019	.0658
Titman	500	N-M	Exp	.0015	-.0097	.04293	.073
AJ	200	Markov	Unif	.0066	.0175	.0464	.07318
AJ	200	Markov	Exp	.0075	.015	.04973	.07903
AJ	500	Markov	Unif	.0083	.0184	.02915	.04758
AJ	500	Markov	Exp	.0081	.0191	.03196	.05099
AJ	200	N-M	Unif	.0126	-.0014	.0458	.07398
AJ	200	N-M	Exp	.0092	.0046	.04945	.08145
AJ	500	N-M	Unif	.0081	.0025	.02822	.04605
AJ	500	N-M	Exp	.0092	.0015	.03153	.05157

Table 17: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of 0→2(healthy to death)

Estimator	n	Model	Censor	Bias $\hat{P}_{02}$	SD $\hat{P}_{02}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{02}$
Titman	200	Markov	Unif	.0014	.08379	.00402	.01116	.00702
Titman	200	Markov	Exp	.002	.09264	.00441	.01296	.00859
Titman	500	Markov	Unif	.005	.05158	.00168	.00456	.00269
Titman	500	Markov	Exp	.0056	.05759	.00198	.00564	.00335
Titman	200	N-M	Unif	-.0065	.08089	.00372	.01037	.00659
Titman	200	N-M	Exp	-.0056	.09051	.00505	.01243	.00822
Titman	500	N-M	Unif	-.008	.05148	.00162	.00445	.00271
Titman	500	N-M	Exp	-.0082	.05775	.00185	.00542	.0034
AJ	200	Markov	Unif	.0244	.05541	.0022	.00566	.00367
AJ	200	Markov	Exp	.0225	.06237	.00253	.00647	.0044
AJ	500	Markov	Unif	.0266	.03675	.00092	.0026	.00206
AJ	500	Markov	Exp	.0272	.03781	.00109	.00297	.00217
AJ	200	N-M	Unif	.0112	.05709	.00226	.00548	.00339
AJ	200	N-M	Exp	.0138	.06321	.00253	.00666	.00419
AJ	500	N-M	Unif	.0106	.03585	.00086	.00213	.0014
AJ	500	N-M	Exp	.0106	.0395	.00108	.00266	.00167

Table 18: Bias and Standard Deviation(SD) of $\hat{P}_{02}(s,t|Z=1)$, and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{P}_{02}(s,t|Z=1)$ using Pseudo-Observation Method of 0→2(healthy to death)

Estimator	n	Model	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	Markov	Unif	-.0012	.0003	.06947	.1006
Titman	200	Markov	Exp	-.0005	-.0019	.07392	.11344
Titman	500	Markov	Unif	-.0023	.002	.04305	.06412
Titman	500	Markov	Exp	-.0004	-.0033	.04642	.0713
Titman	200	N-M	Unif	.0024	.0012	.06782	.10161
Titman	200	N-M	Exp	.0054	.0005	.07634	.11196
Titman	500	N-M	Unif	.0025	.0039	.04349	.064
Titman	500	N-M	Exp	.0037	-.004	.04675	.0696
AJ	200	Markov	Unif	-.0112	-.0074	.04987	.07608
AJ	200	Markov	Exp	-.0112	-.009	.05373	.08225
AJ	500	Markov	Unif	-.0091	-.0083	.03169	.04819
AJ	500	Markov	Exp	-.0085	-.0073	.03468	.05333
AJ	200	N-M	Unif	-.0032	-.0114	.05074	.07605
AJ	200	N-M	Exp	-.0055	-.004	.05674	.0823
AJ	500	N-M	Unif	-.0056	-.0061	.03152	.04642
AJ	500	N-M	Exp	-.0065	-.0067	.03496	.05294

Table 19: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of 1→2(illness to death)

Estimator	n	Model	Censor	Bias $\hat{P}_{12}$	SD $\hat{P}_{12}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{12}$
Titman	200	Markov	Unif	-.0009	.07711	.00483	.01012	.00595
Titman	200	Markov	Exp	-.0024	.08359	.00546	.01287	.00699
Titman	500	Markov	Unif	-.0003	.04745	.00186	.00412	.00225
Titman	500	Markov	Exp	-.0037	.05364	.00216	.0051	.00289
Titman	200	N-M	Unif	.0036	.07685	.00461	.01033	.00592
Titman	200	N-M	Exp	.0059	.08449	.00586	.01254	.00717
Titman	500	N-M	Unif	.0065	.04781	.0019	.00411	.00233
Titman	500	N-M	Exp	-.0003	.0523	.0022	.00486	.00274
AJ	200	Markov	Unif	-.0186	.05553	.00261	.00584	.00343
AJ	200	Markov	Exp	-.0202	.06419	.00301	.00685	.00453
AJ	500	Markov	Unif	-.0174	.03631	.00109	.00239	.00162
AJ	500	Markov	Exp	-.0157	.03832	.00128	.0029	.00172
AJ	200	N-M	Unif	-.0146	.05733	.00259	.00591	.0035
AJ	200	N-M	Exp	-.0095	.05977	.00325	.00679	.00366
AJ	500	N-M	Unif	-.0117	.03489	.00103	.00219	.00135
AJ	500	N-M	Exp	-.0131	.03946	.00127	.00285	.00173

Table 20: Bias and Standard Deviation(SD) of $\hat{P}_{12}(s,t|Z=1)$, and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{P}_{12}(s,t|Z=1)$ using Pseudo-Observation Method of 1→2(illness to death)

Estimator	n	Model	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	Markov	Unif	.0014	-.0002	.07036	.09142
Titman	200	Markov	Exp	.0009	-.0006	.07285	.09733
Titman	500	Markov	Unif	-.0006	.0031	.04304	.0574
Titman	500	Markov	Exp	.0009	-.0013	.04652	.06072
Titman	200	N-M	Unif	.0056	-.0074	.07156	.0952
Titman	200	N-M	Exp	.0037	-.0055	.08336	.10916
Titman	500	N-M	Unif	.0073	-.0087	.04325	.05926
Titman	500	N-M	Exp	.0057	-.0076	.04934	.06579
AJ	200	Markov	Unif	.0155	-.0037	.04959	.06424
AJ	200	Markov	Exp	.0121	.0004	.05485	.07183
AJ	500	Markov	Unif	.015	-.0058	.03116	.04164
AJ	500	Markov	Exp	.016	-.0062	.03373	.04322
AJ	200	N-M	Unif	.0138	-.0088	.04837	.06613
AJ	200	N-M	Exp	.0143	-.0019	.05464	.07125
AJ	500	N-M	Unif	.0142	-.0072	.03099	.04208
AJ	500	N-M	Exp	.0165	-.0103	.03514	.04602

Table 21: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of 1→0(illness to healthy)

Estimator	n	Model	Censor	Bias $\hat{P}_{10}$	SD $\hat{P}_{10}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{10}$
Titman	200	Markov	Unif	.0011	.06038	.00495	.00836	.00364
Titman	200	Markov	Exp	.0003	.06442	.00531	.00947	.00415
Titman	500	Markov	Unif	.0025	.03784	.00185	.0033	.00144
Titman	500	Markov	Exp	-.0004	.04077	.00216	.00369	.00166
Titman	200	N-M	Unif	-.0017	.06346	.00515	.00912	.00403
Titman	200	N-M	Exp	-.0018	.07151	.00696	.01195	.00512
Titman	500	N-M	Unif	-.00143	.03883	.00192	.00359	.00151
Titman	500	N-M	Exp	-.00193	.04416	.00247	.00439	.00195
AJ	200	Markov	Unif	.0117	.04252	.0027	.00414	.00195
AJ	200	Markov	Exp	.0125	.04874	.00316	.00516	.00253
AJ	500	Markov	Unif	.0092	.02694	.0012	.00177	.00081
AJ	500	Markov	Exp	.0098	.02894	.00139	.00191	.00093
AJ	200	N-M	Unif	.005	.04566	.00253	.00445	.00211
AJ	200	N-M	Exp	.0124	.05162	.00319	.00508	.00282
AJ	500	N-M	Unif	.007	.02998	.00116	.00182	.00095
AJ	500	N-M	Exp	.0062	.03139	.00151	.00222	.00102

Table 22: Bias and Standard Deviation(SD) of $\hat{P}_{10}(s,t|Z=1)$, and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{P}_{10}(s,t|Z=1)$ using Pseudo-Observation Method of 1→0(illness to healthy)

transition probability when $Z = 1$ for each transition. Since in Section 4.2, Titman estimator was unbiased and Aalen-Johansen estimator had at most mild bias, the regression model using the pseudo-observation worked using either estimator. This includes using Aalen-Johansen estimator in the non-Markov case. The standard deviation when using Aalen-Johansen estimator is smaller than the standard deviation when using Titman estimator. This leads to the MSE when using Aalen-Johansen estimator being smaller than the MSE using Titman estimator. Therefore, when creating a regression model using a pseudo-observation, Aalen-Johansen estimator works better in this case for Markov and non-Markov cases. This leads us to believe that Aalen-Johansen estimator does not perform poorly in every non-Markov case. This simulation shows that as long as the estimator is unbiased, pseudo-observation method is a valid method to create a regression model. In addition, this simulation

shows that the pseudo-observation method works for estimators with mild bias.

Figure 5 and Figure 6 show histograms and normal QQ-plots of the slope estimates, respectively. Table 23 shows p-values of Shapiro-Wilk test for various n. Slope estimates are from the illness to healthy transition from the non-Markov case. Similar

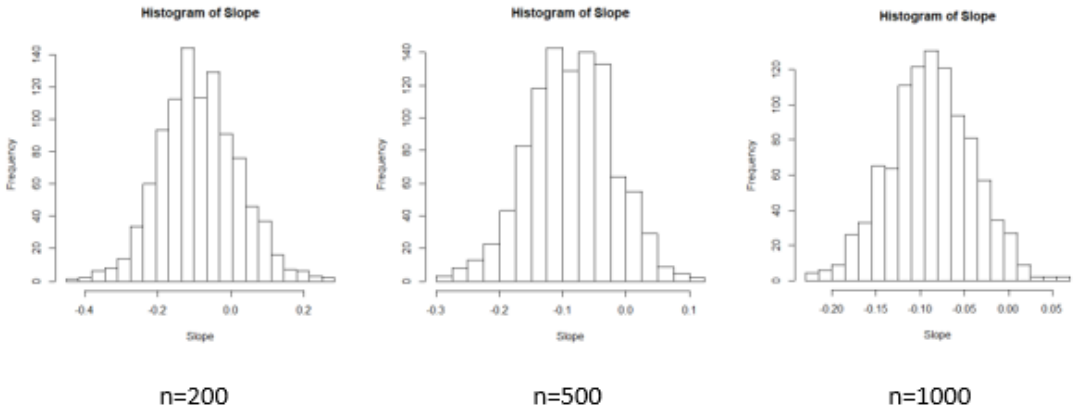

Figure 5: Histogram of the Slope Estimate using Pseudo-Observation Method with Titman Estimator for Illness-Death Model with Recovery

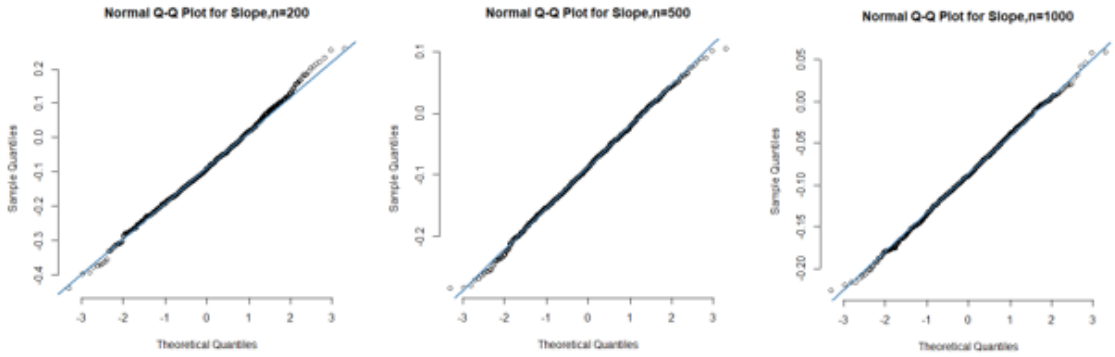

Figure 6: Normal QQ-Plot of the Slope Estimate using Pseudo-Observation Method With Titman Estimator for Illness-Death Model with Recovery

n	P-value
200	.1554
350	.7275
500	.6521
700	.2579
1000	.8624

Table 23: Shapiro-Wilk Test for Normality of the Slope using Pseudo-Observation Method with Titman Estimator for Illness-Death Model with Recovery

to Section 3.2, we simulated 1,000 datasets. We can see that there is slight skewness in the histogram and departure from normality in the QQ-plot for $n = 200$. For the most part, skewness and departure from normality disappear in the histogram and QQ-plot, respectively when n gets larger. Shapiro-Wilk test also show insignificant p-values for all n in Table 23. These results support the asymptotic normality of Theorem 1.

5 Frailty

5.1 Introduction

In the previous two sections, the non-Markov effect was pathological where a transition depended on a condition from a past state. In this section, we will study another type of non-Markov effect called frailty. [6] summarize the motivation of frailty and modeling a frailty distribution in survival models. Frailty is an unobserved heterogeneous effect in a survival model or multi-state model. A common application of frailty can be found when individuals are assigned to various groups for a treatment or getting treated at various institutions. Another application is when an unobserved covariate influences a sub-group of the study. From the attribute of the unobserved covariate, individuals are considered more frail based on the frailty distribution. For example, the individual may be more likely to die sooner in survival models([6]). In the illness-death model, an individual may transition sooner to a certain state depending on the unobserved covariate. This also includes becoming ill and recovering. The frailty random variable must be non-negative. Commonly used frailty distributions are gamma distribution and log-normal distribution.The unobserved heterogeneity of frailty violates the Markov assumption because the transition would not solely depend on the current state.

5.2 Non-Parametric Estimator

Now, we will study the bias and standard deviation of Titman estimator and Aalen-Johansen estimator with the presence of frailty. We will create a simulation study for illness-death model without recovery and illness-death model with recovery. In this section, we will solely study the non-parametric estimator with no consideration of covariates. Similar to the previous two sections, we will cre-

ate a similar simulation study as [28]. We will let the frailty random variable, $W \sim Gamma(\alpha = 1/2, \beta = 2)$.

For the illness-death model without recovery, the transition rates are $\alpha_{01} = 0.12$, $\alpha_{02} = 0.03$, and $\alpha_{12} = 0.1$. We will have a frailty effect multiplied to the healthy to illness and the illness to death transitions, such that $\alpha_{01}W$ and $\alpha_{12}W$. Furthermore, $\alpha_{00} = -\alpha_{01}W - \alpha_{02}$ and $\alpha_{11} = -\alpha_{12}W$. To find the true transition probability, we will use matrix exponentials as we used it in Section 4. By using a property of the matrix exponential, the transition probability matrix, $P(t^*) = \exp(At^*)$ where A represents the transition rate matrix. t^* is the length of time between s and t. Then,

$$A = \begin{bmatrix} -0.12w - 0.03 & 0.12w & 0.03 \\ 0 & -0.1w & 0.1w \\ 0 & 0 & 0 \end{bmatrix}$$

Since the transition rate matrix, A has the gamma frailty random variable, the exponential matrix depends on the frailty variable. We will denote the transition probability as $P(t^*, w)$. The matrix element of $P(t^*, w)$ is represented as $P_{ij}(t^*, w)$ $i = 0, 1$ and $j = 0, 1, 2$. Similar to the previous two sections, s and t are chosen so that they are the 15th and the 45th percentile of the distribution to time to death. From this, $s = 3.162$ and $t = 11.223$. We will show how to find the true transition probability for each transition using matrix exponentials. Since the transition probability depends on the frailty variable, we need to find the joint distribution of the transition and the gamma frailty. Then we will find the marginal distribution of the transition by integrating the gamma frailty variable from the joint distribution. Numerical calculations were done in R. For the healthy to illness transition,

$$P(X(t)=1, X(s)=0) = P(X(s)=0)P(X(t)=1|X(s)=0) = P_{00}(s,w)P_{01}(t-s,w).$$

$B_{11} = \int_0^\infty P_{00}(s,w)P_{01}(t-s,w)f(w)dw$ where $f(w)$ is a gamma density with $\alpha = 1/2$ and $\beta = 2$. Since $P(X(s)=0) = P_{00}(s,w)$, $B_{21} = \int_0^\infty P_{00}(s,w)f(w)dw$. We can find

$$P(X(t)=1|X(s)=0) = \frac{P(X(t)=1, X(s)=0)}{P(X(s)=0)} = \frac{B_{11}}{B_{21}}.$$

For the healthy to death transition,

$$P(X(t)=2, X(s)=0) = P(X(s)=0)P(X(t)=2|X(s)=0) = P_{00}(s,w)P_{02}(t-s).$$

$P_{02}(t^*)$ is the only non-zero element in the transition rate matrix that does not depend on the frailty. $B_{12} = \int_0^\infty P_{00}(s,w)P_{02}(t-s)f(w)dw$. Similar to the healthy to illness transition, $P(X(s)=0) = P_{00}(s,w)$. We can find

$$P(X(t)=2|X(s)=0) = \frac{P(X(t)=2, X(s)=0)}{P(X(s)=0)} = \frac{B_{12}}{B_{21}}.$$

For the illness to death transition,

$$P(X(t)=2, X(s)=1) = P(X(s)=1)P(X(t)=2|X(s)=1) = P_{01}(s,w)P_{12}(t-s,w).$$

$B_{13} = \int_0^\infty P_{01}(s,w)P_{12}(t-s,w)f(w)dw$. Since $P(X(s)=1) = P_{01}(s,w)$,
$B_{22} = \int_0^\infty P_{01}(s,w)f(w)dw$. We can find

$$P(X(t)=2|X(s)=1) = \frac{P(X(t)=2, X(s)=1)}{P(X(s)=1)} = \frac{B_{13}}{B_{22}}.$$

Estimator	n	Transition	Censor	Bias	SD	MSE
Titman	200	0→1	Unif	.0003	.03593	.00129
Titman	200	0→1	Exp	.0006	.03955	.00156
Titman	500	0→1	Unif	.0009	.02205	.00049
Titman	500	0→1	Exp	-.0008	.0258	.00067
AJ	200	0→1	Unif	-.0023	.03158	.001
AJ	200	0→1	Exp	-.0022	.03433	.00118
AJ	500	0→1	Unif	-.0017	.01961	.00039
AJ	500	0→1	Exp	-.0018	.02212	.00049
Titman	200	0→2	Unif	.0009	.04041	.00163
Titman	200	0→2	Exp	.0014	.04481	.00201
Titman	500	0→2	Unif	.0004	.02636	.0007
Titman	500	0→2	Exp	-.0007	.02841	.00081
AJ	200	0→2	Unif	.0033	.03869	.00151
AJ	200	0→2	Exp	.0022	.0415	.00173
AJ	500	0→2	Unif	.0026	.02331	.00055
AJ	500	0→2	Exp	.0035	.02664	.00072
Titman	200	1→2	Unif	.0003	.08642	.00747
Titman	200	1→2	Exp	.0057	.09498	.00905
Titman	500	1→2	Unif	.0032	.05401	.00293
Titman	500	1→2	Exp	.0012	.0637	.00406
AJ	200	1→2	Unif	-.0052	.06505	.00426
AJ	200	1→2	Exp	-.0026	.06879	.00474
AJ	500	1→2	Unif	-.0071	.04183	.0018
AJ	500	1→2	Exp	-.0085	.04539	.00213

Table 24: Bias, Standard Deviation(SD), and MSE of Titman and Aalen-Johansen estimators(AJ) for All Transitions of the Illness-Death Model without Recovery with Frailty Effect

For the illness-death model with recovery, the transition rates are $\alpha_{01} = 0.5$, $\alpha_{02} = 0.02$, $\alpha_{12} = 0.1$, and $\alpha_{10} = 0.3$. We will have $W \sim Gamma(\alpha = 1/2, \beta = 2)$ multiplied to the healthy to illness transition intensity and illness to healthy transition intensity such that we have $\alpha_{01}W$ and $\alpha_{10}W$. Furthermore, $\alpha_{00} = -\alpha_{01}W - \alpha_{02}$ and $\alpha_{11} = -\alpha_{10}W - \alpha_{01}$. Using matrix exponentials, we find the transition probability matrix, $P(t^*) = \exp(At^*)$. The transtion rate matrix A is

$$A = \begin{bmatrix} -0.5w - 0.02 & 0.5w & 0.02 \\ 0.3w & -0.3w - 0.1 & 0.1 \\ 0 & 0 & 0 \end{bmatrix}$$

$s = 3.9121$ and $t = 12.4167$ which are the 15th and the 45th percentile of the distribution to time to death, respectively. The formula to find the true transition probability using matrix exponential remains the same for the healthy to illness and healthy to death transitions. The only difference in the illness to death transition is that illness to death transition rate does not depend on frailty. That is, $P_{12}(t-s)$ instead of $P_{12}(t-s,w)$. For the illness to healthy transition,

$$P(X(t)=0, X(s)=1) = P(X(s)=1)P(X(t)=0|X(s)=1) = P_{01}(s,w)P_{10}(t-s,w).$$

$B_{14} = \int_0^\infty P_{01}(s,w)P_{10}(t-s,w)f(w)dw$. Since $P(X(s)=1) = P_{01}(s,w)$,
$B_{22} = \int_0^\infty P_{01}(s,w)f(w)dw$. From this, we can find

$$P(X(t)=0|X(s)=1) = \frac{P(X(t)=0, X(s)=1)}{P(X(s)=1)} = \frac{B_{14}}{B_{22}}.$$

Table 24 and Table 25 show the bias, standard deviation, and MSE of the illness-death model without recovery and the illness-death model with recovery, respectively. In both models, Titman estimator is unbiased, but the standard deviation is higher than Aalen-Johansen estimator. Aalen-Johansen estimator seems to be unbiased for the illness-death model without recovery, but heavily biased for the illness-death model with recovery. The transition with the smallest bias was healthy to death transition which had at least a bias of .04. The transition with the most bias was

Estimator	n	Transition	Censor	Bias	SD	MSE
Titman	200	0→1	Unif	-.0009	.04368	.00191
Titman	200	0→1	Exp	.0029	.04939	.00245
Titman	500	0→1	Unif	-.0005	.02835	.0008
Titman	500	0→1	Exp	-.0003	.0319	.00102
AJ	200	0→1	Unif	.0499	.03713	.00387
AJ	200	0→1	Exp	.0503	.04399	.00447
AJ	500	0→1	Unif	.0482	.02372	.00289
AJ	500	0→1	Exp	.0486	.02689	.00309
Titman	200	0→2	Unif	.0007	.04411	.00195
Titman	200	0→2	Exp	-.001	.05208	.00271
Titman	500	0→2	Unif	-.0019	.02889	.00084
Titman	500	0→2	Exp	.0003	.03164	.001
AJ	200	0→2	Unif	.0422	.0386	.00327
AJ	200	0→2	Exp	.0428	.04438	.0038
AJ	500	0→2	Unif	.043	.02485	.00247
AJ	500	0→2	Exp	.0401	.02844	.00242
Titman	200	1→2	Unif	-.0014	.06758	.00457
Titman	200	1→2	Exp	<-.0001	.07943	.00631
Titman	500	1→2	Unif	.0011	.04129	.00171
Titman	500	1→2	Exp	<-.0001	.04762	.00227
AJ	200	1→2	Unif	-.0737	.04284	.00727
AJ	200	1→2	Exp	-.0723	.04802	.00753
AJ	500	1→2	Unif	-.0742	.02661	.00621
AJ	500	1→2	Exp	-.0758	.03052	.00668
Titman	200	1→0	Unif	.0042	.05749	.00332
Titman	200	1→0	Exp	-.0005	.06508	.00424
Titman	500	1→0	Unif	.0032	.03695	.00138
Titman	500	1→0	Exp	.0024	.04007	.00161
AJ	200	1→0	Unif	.1612	.04002	.02759
AJ	200	1→0	Exp	.1573	.04394	.02667
AJ	500	1→0	Unif	.1603	.02409	.02628
AJ	500	1→0	Exp	.1607	.02732	.02657

Table 25: Bias, Standard Deviation(SD), and MSE of Titman and Aalen-Johansen estimators(AJ) for All Transitions of the Illness-Death Model with Recovery with Frailty Effect

illness to healthy which had a bias of approximately .16.

For the illness-death model without recovery, the lower standard deviation of Aalen-Johansen estimator caused MSE to be lower than MSE of Titman estimator. However, for the illness-death model with recovery,the heavy bias of Aalen-Johansen estimator caused MSE of Titman estimator to be lower than MSE of Aalen-Johansen estimator. Evaluating the estimators' performances, we believe that using either estimator in creating pseudo-observation will lead to unbiased results in the semi-parametric regression model for the illness-death model without recovery. However, Aalen-Johansen estimator would be the better estimator to use due to its smaller standard deviation. In the illness-death model with recovery, we believe Titman estimator is the better estimator to use when creating pseudo-observation in the semi-parametric regression. The bias of Aalen-Johansen estimator violates the requirement of having an unbiased estimator in creating a pseudo-observation.

5.3 Regression Using Pseudo-Observation

Now, suppose that we want to create a semi-parametric model. Similar to the semi-parametric subsections in the past two sections, we simulate the transition intensities going from state j to state k as having the Cox proportional hazard form. See (38). We will have one binary covariate, $Z \sim Bernoulli(p = 0.5)$. For the illness-death model without recovery, the baseline transition intensities are $\alpha_{010} = 0.12$, $\alpha_{020} = 0.03$, and $\alpha_{120} = 0.1$. We will have a frailty random variable, $W \sim Gamma(\alpha = 1/2, \beta = 2)$ multiplied to the healthy to illness baseline transition intensity and illness to death baseline transition intensity such that $\alpha_{010}W$ and $\alpha_{120}W$. For the illness-death model with recovery, the baseline transition intensities are $\alpha_{010} = 0.5$, $\alpha_{020} = 0.02$, $\alpha_{120} = 0.1$, and $\alpha_{100} = 0.3$. The frailty random variable,

W is multiplied to the healthy to illness baseline transition intensity and illness to healthy baseline transition intensity such that $\alpha_{010}W$ and $\alpha_{100}W$. The other settings are the same as the settings of the semi-parametric regression simulations of the previous two sections. We also are letting $s = 2$ and $t = 6$. We create the pseudo-observation using Titman estimator and Aalen-Johansen estimator. Then by using GEE, we create a regression model of the pseudo-observation on the covariate.

Due to difficulty of computing the true transition probability analytically when having a covariate, we will compute the empirical proportion of the transition probability from state j to state k. It consists of 100,000 observations. Similar to the last two sections, we will create the pseudo-observation using the empirical proportion. By using GEE, we compute the intercept and slope estimates. We compare the GEE results using Titman estimator and Aalen-Johansen estimator to the GEE results of the empirical proportion.

Estimator	n	Trans.	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	0→1	Unif	-.0031	.0062	.0438	.06407
Titman	200	0→1	Exp	-.0024	.0029	.04875	.06977
Titman	500	0→1	Unif	-.0016	.0048	.02769	.03948
Titman	500	0→1	Exp	-.0014	.0029	.03028	.04349
AJ	200	0→1	Unif	-.0026	.0024	.03852	.05467
AJ	200	0→1	Exp	-.0042	.0025	.04442	.06153
AJ	500	0→1	Unif	-.0038	.0028	.02589	.03638
AJ	500	0→1	Exp	-.0033	.004	.02869	.03999
Titman	200	0→2	Unif	-.0002	.0021	.04571	.07074
Titman	200	0→2	Exp	.0007	-.0025	.04959	.081
Titman	500	0→2	Unif	.0015	.0005	.02767	.04639
Titman	500	0→2	Exp	.0025	.0008	.03073	.05078
AJ	200	0→2	Unif	.0046	.0012	.04161	.07217
AJ	200	0→2	Exp	.0072	-.0015	.04702	.07654
AJ	500	0→2	Unif	.0009	.0028	.02578	.04313
AJ	500	0→2	Exp	.0036	.0001	.02907	.04649

Table 26: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of Illness-Death Model without Recovery for 0→1 and 0→2 Transitions

Estimator	n	Trans.	Censor	Bias $\hat{P}_{jk}$	SD $\hat{P}_{jk}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{jk}$
Titman	200	0→1	Unif	.0032	.04653	.00193	.00414	.00218
Titman	200	0→1	Exp	.0005	.05108	.00238	.00488	.00261
Titman	500	0→1	Unif	.0032	.02942	.00077	.00158	.00088
Titman	500	0→1	Exp	.0015	.032	.00092	.0019	.00103
AJ	200	0→1	Unif	-.0002	.04046	.00149	.003	.00164
AJ	200	0→1	Exp	-.0017	.04385	.00199	.00379	.00193
AJ	500	0→1	Unif	-.0011	.02559	.00069	.00133	.00066
AJ	500	0→1	Exp	.0007	.02812	.00083	.00162	.00079
Titman	200	0→2	Unif	.0019	.05523	.00209	.00501	.00305
Titman	200	0→2	Exp	-.0019	.06315	.00246	.00657	.00399
Titman	500	0→2	Unif	.0019	.03667	.00077	.00215	.00135
Titman	500	0→2	Exp	.0033	.03988	.00095	.00258	.0016
AJ	200	0→2	Unif	.0058	.05515	.00175	.00521	.00308
AJ	200	0→2	Exp	.0057	.05755	.00226	.00586	.00335
AJ	500	0→2	Unif	.0037	.03481	.00067	.00187	.00123
AJ	500	0→2	Exp	.0046	.03683	.00086	.00216	.00138

Table 27: Bias and Standard Deviation(SD) of $\hat{P}_{jk}(s,t|Z=1)$; j=0, k=1,2 and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{P}_{jk}(s,t|Z=1)$; j=0, k=1,2 using Pseudo-Observation Method of Illness-Death Model without Recovery for 0→1 and 0→2 Transitions

Estimator	n	Trans.	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	1→2	Unif	.0003	-.0072	.12888	.17918
Titman	200	1→2	Exp	-.0018	-.0091	.13356	.19403
Titman	500	1→2	Unif	-.0026	-.0099	.0835	.11526
Titman	500	1→2	Exp	.0029	-.0121	.08578	.12007
AJ	200	1→2	Unif	-.0084	-.0065	.09359	.1334
AJ	200	1→2	Exp	-.012	-.006	.10259	.1473
AJ	500	1→2	Unif	-.0107	-.008	.0606	.08453
AJ	500	1→2	Exp	-.012	-.0099	.06772	.09312

Table 28: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of Illness-Death Model without Recovery for 1→2 Transition

Estimator	n	Trans.	Censor	Bias $\hat{P}_{12}$	SD $\hat{P}_{12}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{12}$
Titman	200	1→2	Unif	-.0069	.12408	.01661	.03216	.01544
Titman	200	1→2	Exp	-.0109	.14	.01784	.03773	.01972
Titman	500	1→2	Unif	-.0125	.07973	.00698	.01338	.00651
Titman	500	1→2	Exp	-.0093	.08388	.00737	.01456	.00712
AJ	200	1→2	Unif	-.0149	.09677	.00883	.01784	.00959
AJ	200	1→2	Exp	-.018	.10166	.01067	.02173	.01066
AJ	500	1→2	Unif	-.0187	.05926	.00379	.00721	.00386
AJ	500	1→2	Exp	-.0219	.06492	.00473	.00877	.00469

Table 29: Bias and Standard Deviation(SD) of $\hat{P}_{12}(s,t|Z=1)$ and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{P}_{12}(s,t|Z=1)$ using Pseudo-Observation Method of Illness-Death Model without Recovery for 1→2 Transition

Estimator	n	Trans.	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	0→1	Unif	.0007	.003	.05997	.0779
Titman	200	0→1	Exp	-.0001	-.0029	.06542	.08656
Titman	500	0→1	Unif	-.0006	-.0015	.03589	.04976
Titman	500	0→1	Exp	-.0027	.0014	.04031	.05459
AJ	200	0→1	Unif	.0744	-.0361	.05281	.07058
AJ	200	0→1	Exp	.0737	-.0334	.05664	.07596
AJ	500	0→1	Unif	.0727	-.0318	.03266	.04397
AJ	500	0→1	Exp	.0734	-.0319	.03458	.04739
Titman	200	0→2	Unif	.0024	-.0013	.04688	.08178
Titman	200	0→2	Exp	.0001	-.0038	.0491	.0885
Titman	500	0→2	Unif	.0009	-.0025	.02949	.04982
Titman	500	0→2	Exp	<.0001	-.002	.03063	.05454
AJ	200	0→2	Unif	.018	.0326	.04039	.07029
AJ	200	0→2	Exp	.0186	.0343	.04396	.07442
AJ	500	0→2	Unif	.0196	.0342	.02592	.04471
AJ	500	0→2	Exp	.0192	.0356	.02794	.04718

Table 30: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of Illness-Death Model with Recovery for 0→1 and 0→2 Transitions

Estimator	n	Trans.	Censor	Bias $\hat{P}_{jk}$	SD $\hat{P}_{jk}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{jk}$
Titman	200	0→1	Unif	.0038	.05299	.0036	.00608	.00282
Titman	200	0→1	Exp	-.003	.05789	.00428	.0075	.00336
Titman	500	0→1	Unif	-.0021	.03464	.00129	.00248	.0012
Titman	500	0→1	Exp	-.0014	.03747	.00163	.00298	.00141
AJ	200	0→1	Unif	.0383	.04793	.00832	.00629	.00376
AJ	200	0→1	Exp	.0403	.05189	.00864	.00689	.00432
AJ	500	0→1	Unif	.0409	.03036	.00635	.00295	.0026
AJ	500	0→1	Exp	.0415	.03348	.00658	.00326	.00284
Titman	200	0→2	Unif	.0011	.06355	.0022	.00669	.00404
Titman	200	0→2	Exp	-.0037	.07149	.00241	.00785	.00513
Titman	500	0→2	Unif	-.0016	.04028	.00087	.00249	.00163
Titman	500	0→2	Exp	-.0019	.04572	.00094	.00298	.00209
AJ	200	0→2	Unif	.0506	.05448	.00196	.006	.00553
AJ	200	0→2	Exp	.0529	.05824	.00228	.00672	.00619
AJ	500	0→2	Unif	.0537	.03443	.00106	.00317	.00407
AJ	500	0→2	Exp	.0548	.0365	.00115	.00349	.00434

Table 31: Bias and Standard Deviation(SD) of $\hat{P}_{jk}(s,t|Z=1)$; j=0, k=1,2 and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{P}_{jk}(s,t|Z=1)$; j=0, k=1,2 using Pseudo-Observation Method of Illness-Death Model with Recovery for 0→1 and 0→2 Transitions

Table 26 and Table 28 show the bias and standard deviation of the intercept and slope for each transition from the illness-death model without recovery using pseudo-observation method. Table 30 and Table 32 show the same measurements, but for the illness-death model with recovery. Table 27 and Table 29 show the bias and standard deviation of the predicted transition probability when $Z = 1$ when using pseudo-observation method for the illness-death model without recovery. Those tables also show the MSE for the intercept, slope, and predicted transition probability when $Z = 1$. Table 31 and Table 33 show the same measurements, but for the illness-death model with recovery.

The results are consistent with the results from the non-parametric estimator analysis. For the illness-death model without recovery, the intercept and slope estimates are unbiased using either estimator to create the pseudo-observation. Due to the

smaller standard deviation when using Aalen-Johansen estimator, MSE when using Aalen-Johansen estimator is smaller. This leads to Aalen-Johansen estimator being the preferred estimator to use. The standard deviation of the slope of the illness to death transition is large, especially when using Titman estimator. This is because from landmarking, there are drastically less individuals who are ill at time s.

For the illness-death model with recovery, using Aalen-Johansen estimator leads to biased results. Aalen-Johansen estimator was somewhat competitive in the intercept of the healthy to death transition. In this case, the bias was mild making it have a smaller MSE when $n = 200$. However, Titman estimator had a smaller MSE when $n = 500$. Regarding the illness to death and illness to healthy transitions, we see the same issue of landmarking that we saw in the illness-death model without recovery where far less individuals are ill at time s . We see this more of an issue

Estimator	n	Trans.	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	200	1→2	Unif	.0016	.0003	.08367	.12433
Titman	200	1→2	Exp	.005	-.0091	.0911	.1332
Titman	500	1→2	Unif	.0055	-.0064	.05486	.0805
Titman	500	1→2	Exp	.0021	-.004	.05891	.08889
AJ	200	1→2	Unif	-.0409	-.0576	.05193	.07766
AJ	200	1→2	Exp	-.0422	-.0556	.05343	.08775
AJ	500	1→2	Unif	-.0423	-.0554	.03212	.04849
AJ	500	1→2	Exp	-.0417	-.0554	.03335	.0527
Titman	200	1→0	Unif	-.006	.0056	.08116	.10814
Titman	200	1→0	Exp	-.0031	.0045	.09379	.12516
Titman	500	1→0	Unif	-.0058	.0046	.05235	.07029
Titman	500	1→0	Exp	-.0059	.0052	.05748	.07397
AJ	200	1→0	Unif	.2019	-.0395	.05233	.07308
AJ	200	1→0	Exp	.2048	-.047	.05538	.07799
AJ	500	1→0	Unif	.2057	-.047	.03252	.04536
AJ	500	1→0	Exp	.2056	-.0461	.03653	.05105

Table 32: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of Illness-Death Model with Recovery for 1→2 and 1→0 Transitions

Estimator	n	Trans.	Censor	Bias $\hat{P}_{jk}$	SD $\hat{P}_{jk}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{jk}$
Titman	200	1→2	Unif	.0019	.09411	.007	.01546	.00886
Titman	200	1→2	Exp	-.0041	.10218	.00832	.01783	.01046
Titman	500	1→2	Unif	-.0009	.0597	.00304	.00652	.00357
Titman	500	1→2	Exp	-.002	.06457	.00348	.00792	.00417
AJ	200	1→2	Unif	-.0985	.05829	.00437	.00935	.0131
AJ	200	1→2	Exp	-.0978	.06602	.00464	.01079	.01392
AJ	500	1→2	Unif	-.0977	.03809	.00282	.00542	.011
AJ	500	1→2	Exp	-.0971	.03949	.00285	.00585	.01099
Titman	200	1→0	Unif	-.0005	.07055	.00662	.01173	.00498
Titman	200	1→0	Exp	.0014	.08194	.00881	.01569	.00672
Titman	500	1→0	Unif	-.0012	.04507	.00277	.00496	.00203
Titman	500	1→0	Exp	-.0007	.04884	.00334	.0055	.00239
AJ	200	1→0	Unif	.1624	.05275	.0435	.0069	.02916
AJ	200	1→0	Exp	.1578	.05422	.04501	.00829	.02784
AJ	500	1→0	Unif	.1587	.03202	.04337	.00427	.02621
AJ	500	1→0	Exp	.1596	.03595	.04361	.00473	.02677

Table 33: Bias and Standard Deviation(SD) of $\hat{P}_{jk}(s,t|Z=1)$; j=1, k=0,2 and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{P}_{jk}(s,t|Z=1)$; j=1, k=0,2 using Pseudo-Observation Method of Illness-Death Model with Recovery for 1→2 and 1→0 Transitions

Estimator	n	Trans.	Censor	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	SD $\hat{\beta}_0$	SD $\hat{\beta}_1$
Titman	750	1→2	Unif	.0065	-.0073	.04267	.06502
Titman	750	1→2	Exp	.0061	-.0087	.04606	.06862
AJ	750	1→2	Unif	-.0423	-.0567	.02572	.03839
AJ	750	1→2	Exp	-.0428	-.0584	.02772	.04085
Titman	750	1→0	Unif	-.0049	.0046	.04086	.05439
Titman	750	1→0	Exp	-.005	.0057	.04517	.06046
AJ	750	1→0	Unif	.2056	-.0457	.02657	.03642
AJ	750	1→0	Exp	.205	-.0463	.0289	.03995

Table 34: Bias and Standard Deviation (SD) of $\hat{\beta}_0$ and $\hat{\beta}_1$ using Pseudo-Observation Method of Illness-Death Model with Recovery from Additional Simulations for 1→2 and 1→0 Transitions

Estimator	n	Trans.	Censor	Bias $\hat{P}_{jk}$	SD $\hat{P}_{jk}$	MSE $\hat{\beta}_0$	MSE $\hat{\beta}_1$	MSE $\hat{P}_{jk}$
Titman	750	1→2	Unif	-.0008	.04862	.00186	.00428	.00237
Titman	750	1→2	Exp	-.0026	.05089	.00215	.00478	.0026
AJ	750	1→2	Unif	-.099	.029	.00245	.00469	.01064
AJ	750	1→2	Exp	-.1011	.03212	.0026	.00508	.01125
Titman	750	1→0	Unif	-.0002	.03722	.00169	.00298	.00139
Titman	750	1→0	Exp	.0007	.04107	.00207	.00369	.00169
AJ	750	1→0	Unif	.1599	.02618	.04298	.00342	.02625
AJ	750	1→0	Exp	.1593	.02818	.04286	.00374	.02617

Table 35: Bias and Standard Deviation(SD) of $\hat{P}_{jk}(s,t|Z=1)$; j=1, k=0,2 and MSE of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{P}_{jk}(s,t|Z=1)$; j=1, k=0,2 using Pseudo-Observation Method of Illness-Death Model with Recovery from Additional Simulations for 1→2 and 1→0 Transitions

with the slope than the intercept. At $n = 200$ despite the bias, MSE of the slope when using Aalen-Johansen estimator was quite smaller than MSE of the slope when using Titman estimator. At $n = 500$, we see that MSE of the slope when using Titman estimator improves dramatically relative to MSE of the slope when using Aalen-Johansen estimator. However, MSE of the slope when using Aalen-Johansen estimator is still lower. Table 34 shows the bias and standard deviation of the intercept and slope for illness to death and illness to healthy transitions from the illness-death model with recovery using pseudo-observation method for $n = 750$. Table 35 shows the bias and standard deviation of the predicted transition probability when $Z = 1$ when using pseudo-observation method for the illness-death model with recovery for $n = 750$. Table 35 also shows the MSE for the intercept, slope, and predicted transition probability when $Z = 1$. We see that at $n = 750$, MSE of the slope using Titman estimator is lower than MSE of the slope using Aalen-Johansen estimator for both transitions. We believe that MSE of the slope using Titman estimator will remain lower than MSE of the slope using Aalen-Johansen estimator for $n > 750$.

Figure 7 and Figure 8 show the histograms and normal QQ-plots for various

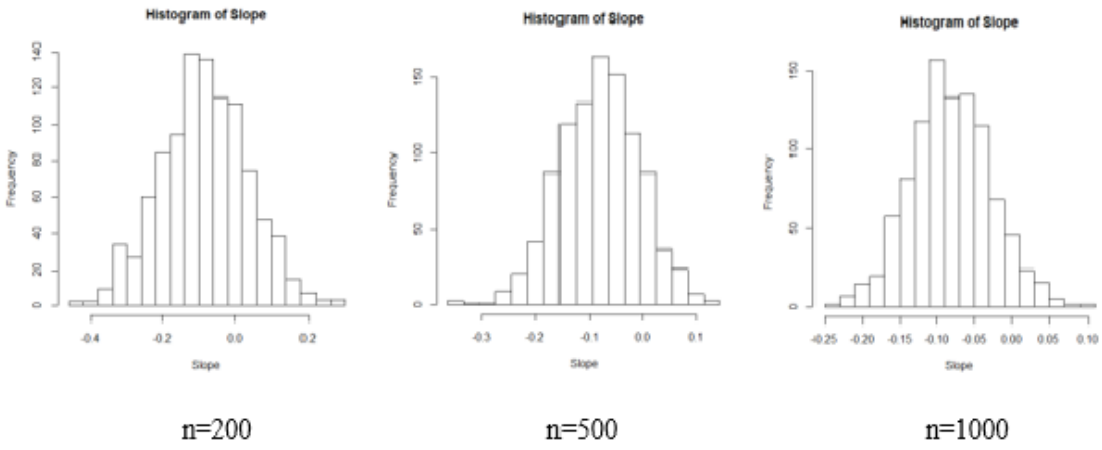

Figure 7: Histogram of the Slope Estimate for Illness-Death Model with Recovery with Frailty Effect using Pseudo-Observation Method with Titman Estimator

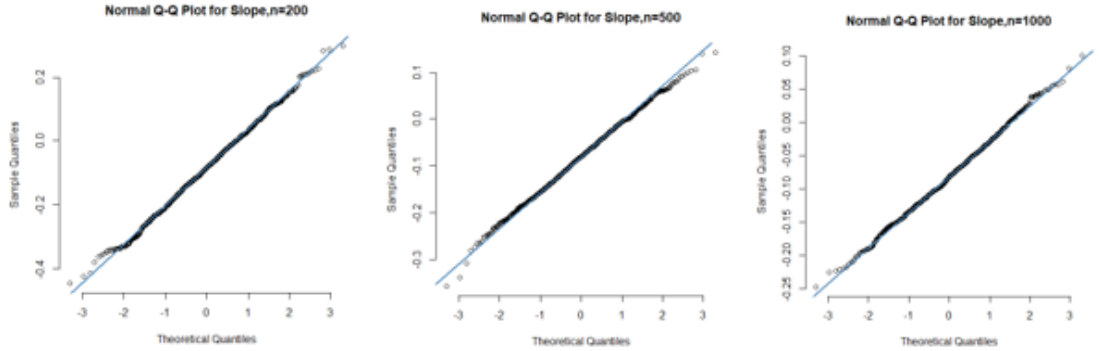

Figure 8: Normal QQ-Plot of the Slope Estimate for Illness-Death Model with Recovery with Frailty Effect using Pseudo-Observation Method with Titman Estimator

n	P-value
200	.5743
350	.5564
500	.5682
700	.9834
1000	.6062

Table 36: Shapiro-Wilk Test for Normality for Illness-Death Model with Recovery with Frailty Effect of the Slope using Pseudo-Observation Method with Titman Estimator

sample sizes, respectively. Table 36 shows the p-values from Shapiro-Wilk test of the slope. Slope estimates are from the illness to healthy transition from the illness-death model with recovery. The histograms show a distribution similar to the normal distribution. The QQ-plots show some mild departure at the ends, but departure from normality does not seem to be an issue. The p-values from Shapiro-Wilk test show insignificant results for all sample sizes listed. Hence, this supports the asymptotic normality result of Theorem 1.

6 Analyzing Liver Cirrhosis Data

6.1 Introduction and Preliminary Work

In the previous sections, we studied our method based on simulations. Now, we will use pseudo-observation method to analyze a dataset. We will be using the liver cirrhosis data available in the *mstate* package in R. The dataset was first introduced in [4]. It was made available in R by [11]. The dataset consists of 488 individuals who have their prothrombin levels monitored. Abnormal level of prothrombin can be lethal. It can lead to liver disease and liver bleeding ([20], [30]). Another motivation of the study was to investigate the usefulness of the prednisone medicine. In the study, 251 individuals took prednisone and the remaining 237 individuals were placebo. Individuals entered the study having normal prothrombin level or having low prothrombin level. In the multi-state model, there will be three states: normal prothrombin level (state 0), low prothrombin level (state 1), and death (state 2). It is

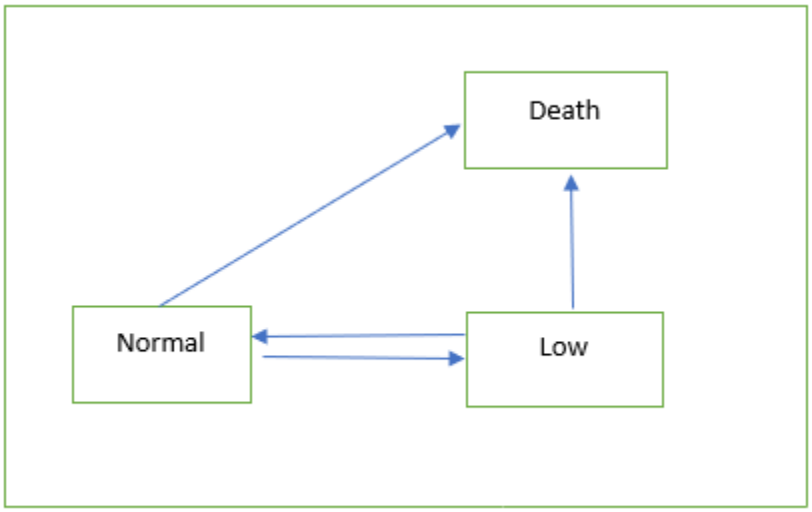

Figure 9: Multi-State Model of Liver Cirrhosis Study

possible for an individual with normal prothrombin level or low prothrombin level to transition to death. It is possible for an individual with low prothrombin level to tran-

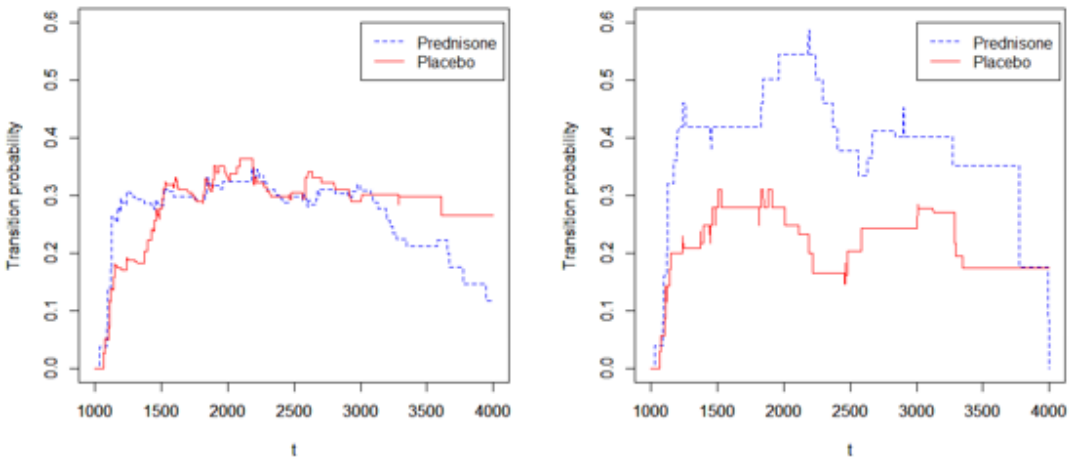

Figure 10: Transition Probability, $P_{10}(s = 1000, t)$ and Various Values of t for Aalen-Johansen estimator(left) and Titman estimator(right)

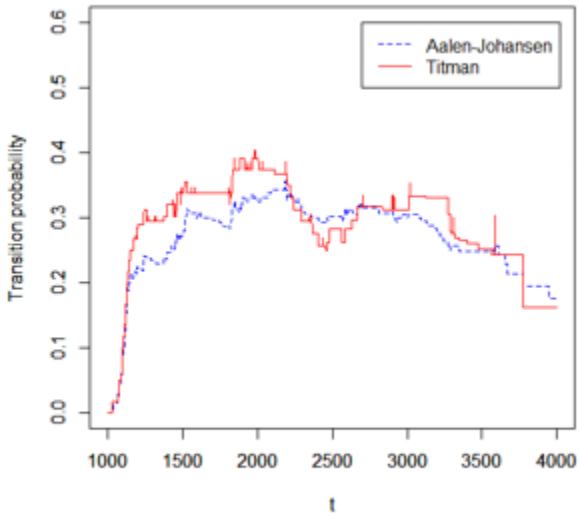

Figure 11: Transition Probability, $P_{10}(s = 1000, t)$ and Various Values of t for Aalen-Johansen estimator and Titman estimator(Groups not separated)

sition to normal prothrombin level. Thus, the multi-state model is an illness-death model with recovery. See Figure 9. [28] analyzed this data using Titman estimator as a non-parametric estimator. That is, analysis was not done using regression. The prednisone group and placebo group were analyzed separately. $s = 1000$ and various values of t were used. The low prothrombin level (state 1) to normal prothrombin level (state 0) transition was mainly studied. Figure 10 shows the plot of the transition probability, $P_{10}(s = 1000, t)$ and various values of t for Aalen-Johansen estimator and Titman estimator. The plots are reproduced from [28]. As [28] mentioned, Titman estimator detects a difference in the transition probability between the two groups for majority of t. Aalen-Johansen estimator does not detect this effect. [28] suggests that the group variable could be an unobserved effect that Titman estimator detects, but Aalen-Johansen estimator does not detect the effect. Therefore, a frailty effect could be present in the data. Figure 11 shows the transition probability, $P_{10}(s = 1000, t)$ and various values of t with Aalen-Johansen estimator and Titman estimator in the same plot. As we can see, there are many areas in the plot that the two estimators are different.

6.2 Markov Testing

In Section 6.1, we see signs of frailty effect in the dataset. In some cases, based on the scientific literature, we can expect non-Markov behavior in the multi-state model in a particular application. However, if prior knowledge of the behavior of the transition is not known, it is best to test whether the Markov property holds. In using pseudo-observations, it is essential to know if the Markov property holds. If the Markov property holds, using Aalen-Johansen estimator will be better. If the Markov property does not hold, Aalen-Johansen estimator may not be the best option based

on the simulation results. [24] created a Markov test using Kendall's τ. However, the method is only applicable to illness-death model without recovery. Titman and Putter[29] created a Markov test that works for general cases. In this section, we will describe [29] 's Markov test. It is important since we will be using it to test the Markov property for the liver cirrhosis data in Section 6.3.

[29]'s method is motivated by the log-rank test. In order to use the log-rank test, there needs to be two different groups of individuals. $\mathcal{S} = \{X_i(s) = j, Y_i(s) = 1\}$ and $\mathcal{S}^c = \{X_i(s) \neq j, Y_i(s) = 1\}$. If the process is Markov, at time $t > s$, the transition intensity should be the same. The transition probabilities are functions of the transtion rates, $\alpha_{lm}(t)$. Let $\alpha_{lm}(t)$ for $l \in \mathcal{R}_j$ where $\mathcal{R}_j$ is the set of states reachable from state j. To test the Markov property, $H_0 : \alpha_{lm}(t|X(s) = j) = \alpha_{lm}(t|X(s) \neq j)$. Let $\delta_i^{(j)}(s) = I(X(s) = j)$. The log-rank statistic for transition from state l to state m is

$$U_s^{(j)}(l, m) = \sum_{i=1}^{n} \int_s^{\tau} \left(\delta_i^{(j)}(s) - \frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t)}{\sum_k Y_{kl}(t)} \right) dN_i^{lm}(t) \tag{50}$$

where $Y_{kl}(t) = I(X_k(t-) = l)Y_k(t)$ is the risk indicator of individual k transitioning from state l to state m. τ is the maximum time possible for t. $N^{lm}(t)$ is the counting process of individuals transitioning from state l to state m. There must be at least one state besides state j that state l is reachable. If there is no other state that state l is reachable, the log-rank statistic will be uniformly 0. Thus, it is necessary that state $l \in \mathcal{R}_j \cap \mathcal{R}_{j^c}$ where $R_{j^c} = \bigcup_{j' \neq j} \mathcal{R}_{j'}$. For example, consider an illness-death model with recovery (healthy=0, illness=1, death=2). Suppose that $j = 0$ and $j^{'} = 1$. Let $l = 1$ and $m = 0$. Suppose we are interested in the illness to healthy transition. $\mathcal{R}_0 = \{0, 1, 2\}$ and $\mathcal{R}_1 = \{0, 1, 2\}$. Indeed, $l = 1 \in \mathcal{R}_0 \cap \mathcal{R}_1 = \{0, 1, 2\}$. Even though state 2 is in the set, l and j cannot be in state 2 since it is an absorbing state.

There are two approaches to test for Markov using the log-rank statistics. One

method is the local test, where log-rank statistics are computed individually. The hypothesis tests are tested individually. Under the null hypothesis, the log-rank statistics are asymptotically independent. The other method is the global test where s is put into an interval, $s \in [t_0, t_{max}] \subset [0, \tau]$. Thus, $t_0 \leq s \leq t_{max}$ and $s \leq t \leq \tau$. In the global test, the asymptotic covariance is of interest for $s, s^{'} \in [t_0, t_{max}]$ where $s \leq s^{'}$. For the transition from state l to state m for individual i, the asymptotic covariance is

$$
\begin{aligned}
&Cov(U_s^{(j)}, U_{s^{'}}^{(j)}) \\
&= \sum_i \int_{s^{'}}^{\tau} Y_{il}(t) \left(\delta_i^{(j)}(s) - \frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t)}{\sum_k Y_{kl}(t)} \right) \left(\delta_i^{(j)}(s^{'}) - \frac{\sum_k \delta_k^{(j)}(s') Y_{kl}(t)}{\sum_k Y_{kl}(t)} \right) d\hat{\Lambda}(t)
\end{aligned}
\tag{51}
$$

where $\hat{\Lambda}(t) = \int_0^t \frac{\sum_k dN_k^{(lm)}(u)}{\sum_k Y_{kl}(u)}$ which is the Nelson-Aalen estimator.

To find estimates of the covariance and variance, consider the kth transition time from state l to state m. The kth transition time is denoted as $t_{(k)}$. Let n_k be the number of individuals at risk at $t_{(k)}$ such that $n_k = \sum_i Y_i(t_{(k)})$. n_k can be decomposed as $n_{k11} + n_{k01} + n_{k10} + n_{k00}$ where $n_{k11} = \sum_i Y_i(t_{(k)}) \delta_k^{(j)}(s) \delta_i^{(j)}(s^{'})$, $n_{k01} = \sum_i Y_i(t_{(k)})(1 - \delta_k^{(j)}(s)) \delta_i^{(j)}(s^{'})$, $n_{k10} = \sum_i Y_i(t_{(k)}) \delta_k^{(j)}(s)(1 - \delta_i^{(j)}(s^{'}))$, and $n_{k00} = \sum_i Y_i(t_{(k)})(1 - \delta_k^{(j)}(s))(1 - \delta_i^{(j)}(s^{'}))$. The covariance can be estimated as

$$
\widehat{Cov}(U_s^{(j)}, U_{s^{'}}^{(j)}) = \sum_{k: t_{(k)} \geq s^{'}} \frac{n_{k11} n_{k00} - n_{k01} n_{k10}}{n_k^2} \tag{52}
$$

If $s = s^{'}$, the variance can be estimated as

$$\widehat{Var}(U_s^{(j)}) = \sum_{k:t_{(k)} \geq s'} \frac{n_{k11}(n_k - n_{k11})}{n_k^2} \tag{53}$$

To derive these results, note that $\sum_i Y_i(t_{(k)})\delta_k^{(j)}(s) = n_{k11}+n_{k10}$ and $\sum_i Y_i(t_{(k)})\delta_k^{(j)}(s^{'}) = n_{k11} + n_{k01}$.

$$\begin{aligned}
&Cov(U_s^{(j)}, U_{s'}^{(j)}) \\
&= \sum_i \int_{s'}^{\tau} Y_{il}(t) \left(\delta_i^{(j)}(s) - \frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t)}{\sum_k Y_{kl}(t)} \right) \left(\delta_i^{(j)}(s^{'}) - \frac{\sum_k \delta_k^{(j)}(s^{'}) Y_{kl}(t)}{\sum_k Y_{kl}(t)} \right) d\hat{\Lambda}(t) \\
&= \sum_i \int_{s'}^{\tau} Y_{il}(t) \left(\delta_i^{(j)}(s) - \frac{n_{k11} + n_{k10}}{n_k} \right) \left(\delta_i^{(j)}(s^{'}) - \frac{n_{k11} + n_{k01}}{n_k} \right) d\hat{\Lambda}(t) \\
&= \int_{s'}^{\tau} \sum_i Y_{il}(t) \delta_i^{(j)}(s) \delta_i^{(j)}(s^{'}) - \frac{n_{k11} + n_{k01}}{n_k} \sum_i Y_{il}(t) \delta_i^{(j)}(s) \\
&- \frac{n_{k11} + n_{k10}}{n_k} \sum_i Y_{il}(t) \delta_i^{(j)}(s^{'}) + \left(\frac{n_{k11} + n_{k10}}{n_k} \right) \left(\frac{n_{k11} + n_{k01}}{n_k} \right) \sum_i Y_{il}(t) d\hat{\Lambda}(t) \\
&= \int_{s'}^{\tau} n_{k11} - \frac{n_{k11} + n_{k01}(n_{k11} + n_{k10})}{n_k} - \frac{n_{k11} + n_{k10}(n_{k11} + n_{k01})}{n_k} \\
&+ \frac{n_k(n_{k11} + n_{k10})(n_{k11} + n_{k01})}{n_k^2} d\hat{\Lambda}(t) \\
&= \int_{s'}^{\tau} n_{k11} - \frac{2(n_{k11} + n_{k01})(n_{k11} + n_{k10})}{n_k} + \frac{(n_{k11} + n_{k01})(n_{k11} + n_{k10})}{n_k} d\hat{\Lambda}(t) \\
&= \int_{s'}^{\tau} n_{k11} - \frac{(n_{k11} + n_{k01})(n_{k11} + n_{k10})}{n_k} d\hat{\Lambda}(t) \\
&= \int_{s'}^{\tau} \frac{n_{k11} n_k - (n_{k11} + n_{k01})(n_{k11} + n_{k10})}{n_k} d\hat{\Lambda}(t) \\
&= \int_{s'}^{\tau} \frac{n_{k11}(n_{k11} + n_{k01} + n_{k10} + n_{k00}) - (n_{k11} + n_{k01})(n_{k11} + n_{k10})}{n_k} d\hat{\Lambda}(t)
\end{aligned}$$

By expanding the terms,

$$
\begin{aligned}
&= \int_{s'}^{\tau} \frac{n_{k11}n_{k00} - n_{k01}n_{k10}}{n_k} d\hat{\Lambda}(t) \\
&= \int_{s'}^{\tau} \frac{n_{k11}n_{k00} - n_{k01}n_{k10}}{n_k} \, d \int_0^t \frac{\sum_k dN_k^{(lm)}(u)}{\sum_k Y_k(u)} \\
&= \int_{s'}^{\tau} \frac{n_{k11}n_{k00} - n_{k01}n_{k10}}{n_k} \frac{dN^{(lm)}(t)}{n_k} \\
&= \int_{s'}^{\tau} \frac{n_{k11}n_{k00} - n_{k01}n_{k10}}{n_k^2} dN^{(lm)}(t) \\
&= \sum_{k:t_{(k)} \geq s'} \frac{n_{k11}n_{k00} - n_{k01}n_{k10}}{n_k^2}
\end{aligned}
$$

$$
\begin{aligned}
&\widehat{Var}(U_s^{(j)}) \\
&= \sum_i \int_s^{\tau} Y_{il}(t) \left(\delta_i^{(j)}(s) - \frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t)}{\sum_k Y_{kl}(t)} \right) \left(\delta_i^{(j)}(s) - \frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t)}{\sum_k Y_{kl}(t)} \right) d\hat{\Lambda}(t) \\
&= \int_s^{\tau} \sum_i Y_{il}(t) \left(\delta_i^{(j)}(s) - \frac{n_{k11}}{n_k} \right) \left(\delta_i^{(j)}(s) - \frac{n_{k11}}{n_k} \right) d\hat{\Lambda}(t) \\
&= \int_s^{\tau} \sum_i \left(Y_{il}(t) \delta_i^{(j)}(s) \delta_i^{(j)}(s) - 2 \frac{n_{k11}}{n_k} \sum_i Y_{il}(t) \delta_i^{(j)}(s) + \frac{n_{k11}^2}{n_k^2} \sum_i Y_i(t) \right) d\hat{\Lambda}(t) \\
&= \int_s^{\tau} \left(n_{k11} - 2 \frac{n_{k11}^2}{n_k} + \frac{n_k n_{k11}^2}{n_k^2} \right) d\hat{\Lambda}(t) \\
&= \int_s^{\tau} \left(n_{k11} - 2 \frac{n_{k11}^2}{n_k} + \frac{n_{k11}^2}{n_k} \right) d\hat{\Lambda}(t) \\
&= \int_s^{\tau} \left(n_{k11} - \frac{n_{k11}^2}{n_k} \right) d\hat{\Lambda}(t) \\
&= \int_s^{\tau} \frac{n_k n_{k11} - n_{k11}^2}{n_k} d\hat{\Lambda}(t) \\
&= \int_s^{\tau} \frac{n_{k11}(n_k - n_{k11})}{n_k} d\hat{\Lambda}(t)
\end{aligned}
$$

$$
\begin{aligned}
&= \int_s^\tau \frac{n_{k11}(n_k - n_{k11})}{n_k}\, d \int_0^t \frac{\sum_k dN_k^{(lm)}(u)}{\sum_k Y_k(u)} \\
&= \int_s^\tau \frac{n_{k11}(n_k - n_{k11})}{n_k} \frac{dN^{(lm)}(t)}{n_k} \\
&= \int_s^\tau \frac{n_{k11}(n_k - n_{k11})}{n_k^2} dN^{(lm)}(t) \\
&= \sum_{k:t_{(k)} \geq s} \frac{n_{k11}(n_k - n_{k11})}{n_k^2}
\end{aligned}
$$

□

For the local test, each of the $U_s^{(j)}(l,m)$ are asymptotically independent. When $U_s^{(j)}(l,m)$ is standardized,

$$\bar{U}_s^{(j)}(l,m) = \frac{U_s^{(j)}(l,m)}{\sqrt{\widehat{Var}(U_s^{(j)})}}. \tag{54}$$

Then when combining the standardized log-rank statistics, it becomes a chi-square statistic, $\sum \bar{U}_s^{(j)}(l,m)^2$. The sum is over the practitioner's desired values of s.

For the global test, under the null hypothesis, the process, $\{\bar{U}_s^{(j)}(l,m), s \in [t_0, t_{max}]\}$ converges to a zero-mean Gaussian process with covariance function,

$$\frac{Cov(U_s^{(j)}, U_{s'}^{(j)})}{\sqrt{Var(U_s^{(j)})Var(U_{s'}^{(j)})}}. \tag{55}$$

Furthermore, for the global test, Markov test can be tested based on a summary statistic of the process $\{\bar{U}_s^{(j)}(l,m), s \in [t_0, t_{max}]\}$. Commonly used summary statistics are mean, maximum, and weighted mean. Weights are important because there may be areas in the timeline that may have only few individuals. Without weighting, truncating s may be necessary. Areas with few individuals should be down-weighted

compared to other areas in the timeline. [29] proposes the weight to be

$$w(s) = \frac{\sqrt{d_j(s) n_j^{(1)}(s) n_j^{(0)}(s)}}{n_j^{(1)}(s) + n_j^{(0)}(s)} \tag{56}$$

where $n_j^{(1)}(s) = \sum_i \delta_i^{(j)}(s) Y_{il}(s)$, $n_j^{(0)}(s) = \sum_i (1 - \delta_i^{(j)}(s)) Y_{il}(s)$, and $d_j(s) = N^{(lm)}(\tau) - N^{(lm)}(s)$.

To avoid the computation to be too expensive, [29] suggests computing $\bar{U}_s^{(j)}(l, m)$ for selected grid time points, $s_1, ..., s_L$. Then the null distribution is the Gaussian process only using the selected grid time points. If the global test is based on a summary statistic using the mean or maximum, the mean of the log-rank statistics and maximum of the log-rank statistics only using the selected grid time points will be computed, respectively.

Instead of using Gaussian process, [18] created an approximation method using wild bootstrap.Wild bootstrap is used more commonly in regression. The residuals are multiplied to independent, identically distributed(i.i.d.) random variables G_{ih} $i = 1, ..., n$; $h = 1, ..., N_i^{(lm)}(\tau)$. $N_i^{(lm)}(\tau)$ is the total number of transitions from state l to state m for individual i. The residual portion in (50) is $\delta_i^{(j)}(s) - \frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t)}{\sum_k Y_{kl}(t)}$. $\delta_i^{(j)}(s)$ can be treated as the observed binary variable. $\frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t)}{\sum_k Y_{kl}(t)}$ can be treated as the predicted proportion of individuals who are at state j at time s. The mean and variance of G_{ih} are 0 and 1, respectively. This was a requirement that [32] and [19] made when establishing wild bootstrap. It would make sense to use standard normal distribution. However, [7] showed that using centered Poisson random variables is better than using standard normal random variables. The wild bootstrap form of (50) is

$$\sum_{i=1}^{n} \sum_{h=N_i(s)+1}^{N_i(\tau)} \left(\delta_i^{(j)}(s) - \frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t_{ih})}{\sum_k Y_{kl}(t_{ih})} \right) G_{ih} \tag{57}$$

where t_{ih} is the time of the hth transition from state l to state m for individual i. (57) will be resampled.

So far, a specific state j has been tested. [29] suggest testing the Markov property in general by utilizing a vector of log-rank statistics. Let $\mathcal{R}_l^*$ be the set of states reachable to state l. Let $\mathbf{U}(s)$ be a vector of log-rank statistics, $U_s^{(j)}(l,m)$. $r^* = |\mathcal{R}_l^*|$ is the cardinality of $\mathcal{R}_l^*$. Let Ω be a $r^* \times r^*$ singular covariance matrix where the $(j,\ j^{'})$th element of Ω is $Cov\left(U_s^{(j)}(l,m), U_s^{(j^{'})}(l,m)\right)$. There is one constraint. In, $U_s^{(j)}(l,m)$, $\delta_i^{(j)}(s)$ is unknown. Hence, it is being estimated by $\frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t)}{\sum_k Y_{kl}(t)}$. Under the null hypothesis, $K_s(l,m) = \mathbf{U}_{(r)}(s)^T \Omega_{(r,r)}^{-1} \mathbf{U}_{(r)}(s)$ follows a chi-square distribution with $r^* - 1$ degrees of freedom. The one degrees of freedom lost is from the constraint. $\mathbf{U}_{(r)}(s)$ is $\mathbf{U}(s)$ without the rth element and $\Omega_{(r,r)}$ is Ω without the rth row and rth column. This is so that Ω becomes a non-singular matrix. r can be chosen arbitrarily. When using summary statistics for global tests, the mean,maximum, and weighted mean of $K_s(l,m)$ from $[t_0, t_{max}]$ can be computed. This summarizes [29]'s method to test the Markov property.

Non-Markov effects can also come from frailty. Commenges and Andersen [8] established a score test for homogeneity for survival data. This can be extended to multi-state models by considering the stratified version of this test by making each transition intensity a strata ([29]). This would be referred to as the stratified Commenges-Andersen test. Let individual i have the transition intensity modeled as $\alpha_{lmi}(t; z_i) = \alpha_{lm0}(t, z_i) \exp(\sigma\epsilon_i + \beta_{lm} z_i)$ where ϵ_i are i.i.d. random variables of an unspecified distribution with mean 0 and variance 1. Let z_i be a covariate for individual i. The frailty portion of the model is $\exp(\sigma\epsilon_i)$. Then the score test to test for frailty is $H_0 : \sigma = 0$ where σ^2 is the variance of the frailty distribution. That is, under H_0, frailty effect does not exist.

6.3 Data Analysis

In order to use pseudo-observation method for the liver cirrhosis data, we will need to test whether the Markov property holds. In the analysis, we will mainly focus on the illness to healthy transition.We will use the log-rank test. We will also use stratified Commenges-Andersen test to test for frailty, in particular.

Regarding the log-rank test, the $MarkovTest$ function from the $mstate$ package in R was used. The $MarkovTest$ function was created by [29].The global test was used. For the summary statistic, mean, maximum and weighted mean were used. When testing specific j, $\{\bar{U}_s^{(j)}(l,m) \;\; s \in [0, 2000]\}$. For the weight, the recommended formula for $w(s)$ was used. For the overall chi-square test, $K_s(l,m)$ was computed for mean, maximum, and weighted mean. There were 2,000 grid times: 1,2,. . . ,2000. The wild bootstrap was replicated 1,000 times. G_{ih} are i.i.d. centered Poisson random variables.

$\bar{U}_s^{(j)}(l,m)$	Test Stat	P-value
Mean (j=1)	.7771	.206
Mean (j=2)	.7771	.206
Max (j=1)	3.212	.058
Max (j=2)	3.212	.058
Weighted Mean (j=1)	.9833	.118
Weighted Mean (j=2)	.9833	.118
$K_s(l,m)$	Test Stat	P-value
Mean	.9495	.026
Max	10.317	.026
Weighted Mean	2.2812	.026

Table 37: Statistic and P-value for the State-Specific Test and Overall Test for the Transition of Low Prothrombin Level to Normal Prothrombin Level

Table 37 shows the statistic and p-value for the state-specific test and the overall test for the transition of low prothrombin level to normal prothrombin level. For the state-specific test, $j = 1$ and $j = 2$ were used. Looking at the result, log-rank

test gives mix results. From the state-specific tests, there is mild evidence at best of violation of the Markov property. However, in the overall chi-square test, there seems to be evidence of violation of the Markov property. When testing for frailty, the p-value from the stratifed Commenges-Andersen test is .0019. Thus, there is significant evidence that frailty effect exists. To summarize, we can say that there is a violation of the Markov property due to the frailty effect. We would need to be cautious on concluding whether a pathological non-Markov behavior exists.

Now, we will create a semi-parametric regression model of the data using pseudo-observation method. The treatment variable (prednisone or placebo) will be a binary covariate in the regression model. Time $s = 1000$ and various t will be used. Only one pseudo-observation is needed. For the mean function, the identity link function will be used. From the quasi-likelihood function, the variance function relative to the mean, will follow a normal distribution. Table 38 shows the results from GEE when using $t = 1500, 2000, 2500,$ and 3000. Pseudo-observations using Titman estimator and Aalen-Johansen estimator were considered. Since Titman estimator is a landmark estimator, only individuals who are at the low prothrombin level at time s and who are at risk at time s will be part of the subsample. It is of interest to know how large the subsamples are. Table 39 shows the subsample size of each state and number of individuals who are censored for various s.

Overall, the parameter estimates for each t roughly matches what we see in Figure 10. When using Aalen-Johansen estimator, the slope is close to 0. In many cases when using Titman estimator, the slope is relatively larger than 0. Looking at the graph and slope estimates, it seems that Titman estimator detects the treatment covariate and Aalen-Johansen estimator does not detect the covariate. The insignificant slope when using Aalen-Johansen estimator is shown by the high p-value.

Another important point is that the standard error when using Titman estimator

s=1000, t=1500					
Estimator	Parameter	Estimate	Standard Error	Test Stat	P-value
Titman	β_0	$\hat{\beta}_0$ =.2663	.0952	7.8256	.0052
Titman	β_1	$\hat{\beta}_1$ =.1389	.12783	1.18	.2774
AJ	β_0	$\hat{\beta}_0$ =.2716	.06529	17.3073	<.0001
AJ	β_1	$\hat{\beta}_1$ =.0024	.08704	.0007	.9784
s=1000, t=2000					
Estimator	Parameter	Estimate	Standard Error	Test Stat	P-value
Titman	β_0	$\hat{\beta}_0$ =.2555	.09706	6.9294	.0085
Titman	β_1	$\hat{\beta}_1$ = .2636	.13131	4.0283	.0447
AJ	β_0	$\hat{\beta}_0$ =.3337	.05738	33.8329	<.0001
AJ	β_1	$\hat{\beta}_1$ =-.0132	.07897	.028	.867
s=1000, t=2500					
Estimator	Parameter	Estimate	Standard Error	Test Stat	P-value
Titman	β_0	$\hat{\beta}_0$ =.1928	.08702	4.9091	.0267
Titman	β_1	$\hat{\beta}_1$ =.1733	.12559	1.9039	.1677
AJ	β_0	$\hat{\beta}_0$ =.3072	.05108	36.1775	<.0001
AJ	β_1	$\hat{\beta}_1$ =-.0106	.07172	.02192	.8823
s=1000, t=3000					
Estimator	Parameter	Estimate	Standard Error	Test Stat	P-value
Titman	β_0	$\hat{\beta}_0$ =.2311	.09568	5.8332	.0157
Titman	β_1	$\hat{\beta}_1$ =.1583	.13196	1.4384	.2304
AJ	β_0	$\hat{\beta}_0$ =.2907	.0509	32.6201	<.0001
AJ	β_1	$\hat{\beta}_1$ =.0175	.07143	.0597	.807

Table 38: GEE Output using Pseudo-Observation for Low to Normal Prothrombin Level Transition for t=1500,2000,2500, and 3000

to create pseudo-observation is drastically higher than the standard error when using Aalen-Johansen estimator to create pseudo-observations. In general, the standard error when using Aalen-Johansen estimator being lower than the standard error when using Titman estimator is expected. However, the difference is quite substantial. When using Titman estimator, the high standard error for the slope caused it to be insignificant in many cases. Looking at Table 39, there are only 61 individuals in the subsample of low prothrombin level state at $s = 1000$. We believe that is substantially

State	$s = 1000$	$s = 500$	$s = 365$
Normal	179	213	234
Low	61	93	98
Dead	172	124	109
Censored	76	58	47

Table 39: Subsample Size for Each State for Various s

low and contributes to the high standard error. At $s = 1000$, we see that over half of the sample is either dead or censored. If we set s to be smaller, fewer individuals are dead or censored. More individuals are in the normal or low prothrombin level state. However, far more individuals are in the normal prothrombin level state than the lower prothrombin level state. Setting s to be smaller will likely lower the standard error, but it is hard to say that the p-value will get lower. This data analysis shows the importance of having a high landmark subsample size when using Titman estimator to create pseudo-observations.

7 Discussion

7.1 Another Approach

In our study of multi-state models, counting processes played a critical role in finding the non-parametric estimator for non-Markov transition probabilities. Counting processes are used pervasively in multi-state models, including survival models and competing risks models. However, counting processes are not the only way to study transition probabilities in multi-state models. Applications of graphical models are relevant to areas such as spatial statistics, statistical learning, and agent-based simulation. One application is the Susceptible-Infectious-Recovered (SIR) model. We will study the literature and methods. SIR model is expressed as a differential equation.

$$\begin{cases} \frac{dS}{dt} = & bN - \lambda S - dS \\ \frac{dI}{dt} = & \lambda S - gI - dI \\ \frac{dR}{dt} = & gI - dR \end{cases} \tag{58}$$

where S is the number of susceptible individuals, I is the number of infectious individuals, R is the number of recovered individuals, N is the population size, b is the birth rate, d is the natural death rate, g is the rate of recovery from infection, and λ is the rate at which susceptible individuals become infected. SIR models express the rate of change of susceptible individuals, infectious individuals, and recovered individuals. The SIR model is appropriate for diseases with lifetime immunity. Therefore, an illness-death model with recovery when the transition of illness to healthy occurring at most once seems to be reasonably analogous to the SIR model. [16] reviewed the literature and methods regarding SIR models and other network models. SIR models are well-used to model spread of infectious diseases. However, SIR models assume

all individuals are able to interact with one another. This is not realistic when the population is large. In many large population settings, individuals can interact with few individuals. Alternatively, a deterministic mean-field models are used to approximate average quantity of the stochastic model. The behavior of a spread of disease is considered to be stochastic. Two common ways to find proportion of susceptible, infectious, and recovered individuals are message passing (MP) method and edge-based compartmental model (EBCM). [26] describes these methods.

MP method analyzes each individual separately and all possible ways of an individual getting infected by its neighbor in a graphical structure. MP method was established by [15]. Let the test individual, u be an individual that can be infected, but cannot infect other individuals. Let v be a neighbor of u. Message is defined as the probability of v not able to transmit the disease to u by time t. The message is denoted as $H^{u \leftarrow v}(t)$. To find the probability of $H^{u \leftarrow v}(t)$, there are two cases to consider. The first case is that v is intially infected at time 0 and does not transmit the disease to u by time t. This probability is $\int_0^t f(a)da$ where a is the length of time until transmisson occurs. a is also known as the age of infection since the infection started at time 0. $f(a)$ is the density of an individual who was infected at time 0 transmitting the disease to a neighbor by time t. For the other case, suppose that v was not intially infected at time 0, and was infected by one of its neighbors later. After v was infected, suppose that it takes a length of time of $a < t$ to attempt to transmit the disease to u. If v was infected at $t' > t - a$, u will be infected after time t. In addition, v must have been intially susceptible to the disease and avoided transmisson of the disease from its neighbors until after time $t - a$. Suppose that the probability of being initially susceptible is e. The probability of this case can be modeled as $e \int_0^t f(a) \prod_{w \in \mathcal{N}(v) \setminus u} H^{v \leftarrow w}(t - a)da$. $\mathcal{N}(v)$ is the set of neighbors of v. Adding the two cases,

$$H^{u\leftarrow v}(t) = 1 - \int_0^t f(a) \left(1 - e \prod_{w\in\mathcal{N}(v)\setminus u} H^{v\leftarrow w}(t-a) \right) da. \tag{59}$$

Rather than analyzing a message between two individuals, the average message is denoted as, $H_1(t)$ is of interest. This requires configuration models(CM). An individual has k neighbors which is defined as a degree. $p_k \quad k = 0, 1, ...,$ is a probability distribution that a random individual in the graphical structure has degree k. p_k is defined as the degree distribution. The excess degree is the number of edges the neighbor has, excluding the edge with the test individual, u. Thus, excess degree is $k-1$. The moment generating functions for degree distribution and excess degree distribution are respectively, $G_0(x) = \sum_k p_k x^k$ and $G_1(x) = \frac{1}{G_0'(1)} \sum_k k p_k x^{k-1}$. $G_0'(1) = \sum_k k p_k$. To find the average message, consider $\prod_{w\in\mathcal{N}(v)\setminus u} H^{v\leftarrow w}(t-a)$ in (59). The excess degree is considered since the edge to u is excluded. In addition, the product in (59) is replaced with the moment generating function of the excess degree distribution, $G_1(x)$. The average message, $H_1(t)$ is expressed as

$$H_1(t) = 1 - \int_0^t f(a) \left(1 - e G_1(H_1(t-a)\right) da \tag{60}$$

with an initial condition of $H_1(0) = 1$. Then, the expression of the MP method regarding susceptible, recovered, and infectious proportion can be expressed as

$$\begin{aligned} \langle S\rangle(t) &= e G_0(H_1(t)) \\ \langle R\rangle(t) &= \int_0^t q(a) \left(1 - \langle S\rangle(t-a)\right) da \\ \langle I\rangle(t) &= 1 - \langle S\rangle(t) - \langle R\rangle(t) \end{aligned} \tag{61}$$

For $\langle S\rangle(t)$, e is the proportion of individuals who are intially susceptible to the disease.

In order for the individual to remain susceptible until time t, the individual must avoid the transmisson of the disease until time t. All neighbors are eligible to transmit the disease. Therefore, all k neighbors need to be considered. Thus, the moment generating function of the degree distribution, $G_0(x)$ is used. For $\langle R \rangle(t)$, $q(a)$ is denoted as the density of the duration of the infection. At time a, the individual has recovered, but needs to avoid getting susceptible again for the remaining $t - a$ time in order to stay recovered at time t.

Key difference between EBCM and MP method is that EBCM uses instantaneous rates of transmission and recovery by using hazard functions, and MP method uses raw densities. EBCM was created by [22]. However, it only works for systems with Markov behavior. [26] extended EBCM that accommodates non-Markov behavior. We will describe their extended EBCM. The hazard functions for transmission and recovery are denoted respectively, $\zeta(a)$ and $\rho(a)$. EBCM also utilizes the CM method. The motivation of the method is to find the probability of a random test individual is susceptible, infectious, or recovered at time t. The proportions of individuals being susceptible, infectious, and recovered at time t are denoted as $S(t)$, $I(t)$ and $R(t)$, respectively. The model also considers the behavior of the neighbors of the random test individual. The model needs to consider a random neighbor being susceptible, infectious neighbors who did not transmit the disease to the test individual, and recovered neighbors who did not transmit the disease to the test individual. The probability densities are respectively, $\Psi_S(t)$, $\Psi_I(t)$, and $\Psi_R(t)$. Since the edge to the test individual is excluded, the excess degree distribution would be used. EBCM is expressed as (62).

$\Theta(t)$ represents the probability that the test individual does not get the disease transmitted from its neighbors until time t. Hence, $\Theta(t)$ plays the role of the message. In EBCM, the differential equation of $\Theta(t)$ is also part of the model. It is monotoni-

cally decreasing, and depends on $\zeta(a)$ and $\psi_I(t,a)$. Since the hazard functions depend on the age of infection, a, joint densities need to be considered. $i(t,a)$ denotes the joint density of infectious individuals and has an age of infection, a. $\psi_I(t,a)$ denotes the joint density of an infectious neighbor that has not transmitted the disease to the test individual and has an age of infection, a. Furthermore, $\Psi_I(t) = \int_0^t \psi_I(t,a)da$ and $I(t) = \int_0^t i(t,a)da$. The rate of infectious neighbors and rate of infectious individuals are denoted as $\psi_I(t,0)$ and $i(t,0)$, respectively. That is, rate of infectious monitors the spread of the disease from intially-infected individuals. To compute $\psi_I(t,0)$, the boundary condition is the rate of susceptible neighbors getting infected. Then to compute $\psi_I(t,0)$, the interior of $-\Psi_S(t)$ is computed. Similarly, to compute $i(t,0)$, the boundary condition is the rate of susceptible individuals getting infected. Then

$$
\begin{cases}
\frac{d\Theta(t)}{dt} = -\int_0^t \zeta(a)\psi_I(t,a)da \\
\Psi_S(t) = eG_1(\Theta(t)) \\
\psi_I(t,0) = -\dot{\psi}_S(t) = (1-b)\delta(t) + bG_2(\Theta(t))\int_0^t \zeta(a)\psi_I(t,a)da \\
(\frac{\partial}{\partial t} + \frac{\partial}{\partial a})\psi_I(t,a) = -(\zeta(a)+\rho(a))\psi_I(t,a) \quad 0 < a \leq t \\
\Psi_I(t) = \int_0^t \psi_I(t,a)da \\
\Psi_R(t) = \Theta - \Psi_S - \Psi_i \\
S(t) = eG_0(\Theta(t)) \\
i(t,0) = -\dot{S}(t) = (1-b)\delta(t) + eG_0'(1)G_1(\Theta(t))\int_0^t \zeta(a)\psi_I(t,a)da \\
(\frac{\partial}{\partial t} + \frac{\partial}{\partial a})i(t,a) = -\rho(a)i(t,a) \quad 0 < a \leq t \\
I(t) = \int_0^t i(t,a)da \\
R(t) = 1 - S(t) - I(t)
\end{cases}
\tag{62}
$$

to compute $i(t,0)$, the interior of $-S(t)$ is computed. In the first term of $\psi_I(t,0)$ and $i(t,0)$, there is a $\delta(t)$. This is a Dirac delta function where the function is equal to zero when $t \neq 0$, but it spikes up to a large positive value when $t = 0$. The first term represents the beginning of the infection at $t = 0$. Finally, there are partial differential equations with respect to t and a for $\psi_I(t,a)$ and $i(t,a)$. This is because individuals and its neighbors eventually recover. This behavior will depend with respect to both t and a.

[26] showed that (61) and (62) are equivalent. They are both suitable for systems with non-Markov behavior. There is one fundamental difference between SIR models and multi-state models using counting processes. In SIR models, the states(nodes) of the graph or lattice are individuals, and the disease is traveling from state to state. In multi-state models uisng counting processes, there are less states in the model. It is the individuals who are traveling from state to state. We believe that both approaches are useful and have solutions to the non-Markov problem. However, the two approaches are not comparable.

7.2 Conclusions and Remarks

In the past, there have been only few regression models that accommodate for non-Markov models. Creating a regression model using pseudo-observations, itself is not a new method. However, our contribution is that we applied it to the regression model of transition probability to covariates in a multi-state model where non-Markov behavior may exist. This includes the pathological non-Markov effect and the frailty effect. This works for time-irreversible and time-reversible multi-state models. The only requirement to use pseudo-observation method is that we need an unbiased estimator. In the simulations of Section 3, Section 4, and Section 5, we saw that when

the non-parametric estimator is unbiased or has mild bias, building a regression model using a pseudo-observation, and estimating the parameter using GEE gives accurate results. This shows the validity of our method.

When using Titman estimator to create the pseudo-observation, the estimates were accurate in all cases for Markov and non-Markov models. While we cannot claim that using Titman estimator is always better, we found it as a reliable alternative when Aalen-Johansen estimator is biased. We recommend using [29]'s method to test for the Markov property. We also recommend testing for frailty using stratifed Commenges-Andersen test. If the Markov property holds, Aalen-Johansen estimator should be used due to its smaller standard deviation. If Markov property does not hold, we recommend practitioners to compute the non-parametric estimator, Aalen-Johansen estimator and Titman estimator. We know that Titman estimator is unbiased in all non-Markov cases. If the two estimators match or are close to eachother, Aalen-Johansen estimator should be used when creating the pseudo-observation. If the two estimators are quite apart from eachother, we believe using Titman estimator to create the pseudo-observation is a safer option. We noticed from the simulations that at times using Aalen-Johansen estimator to create pseudo-observations worked well, and there were times that it performed poorly. From the simulation, we saw that using Aalen-Johansen estimator to create the pseudo-observation lead to poor results for pathological non-Markov models for the illness-death model without recovery and the frailty case for the illness-death model with recovery. On the contrary, it lead to accurate results in the frailty case for the illness-death model without recovery and pathological non-Markov model for the illness-death model with recovery. As of right now, we do not have a theoretical reasoning of why it worked and did not work in certain cases. In addition, the bias also differs for various combinations of s and t. Bias may also differ based on the transition intensities of the multi-state model.

Hence, our choice of s, t , and the transition intensities could have an impact on the result of the simulation. Ultimately as future work, we would like to throughly investigate when and how Aalen-Johansen estimator is biased and unbiased. Once we know that, we can make better recommendations to when to use Titman estimator and when to use Aalen-Johansen estimator in creating pseudo-observations. With this all discussed, we are still convinced that Aalen-Johansen estimator is not the solution for every non-Markov case, and an alternative is absolutely necessary.

While using Titman estimator to create pseudo-observations lead to accurate parameter estimates, the standard deviation was higher than when using Aalen-Johansen estimator. This lead to higher MSE when using Titman estimator. This was due to the subsample from landmarking. [23] created the non-parametric estimator, Landmark Aalen-Johansen(LMAJ) estimator that showed that the standard deviation was less than Titman estimator. We believe that even if we create pseudo-observations using LMAJ estimator, the standard deviation of the parameter estimates from GEE could still be an issue. [29] claimed that landmark estimators had a trade-off between being unbiased and having a larger standard deviation.

As far as we know, solving the issue on larger standard deviation is still an ongoing problem, particularly for time-reversible models.[27] suggests pre-smoothing the estimator to reduce the standard deviation. They focused on the non-parametric estimator created by [10] in an illness-death model without recovery. The expression of the non-parametric estimator from [10] utilizes Kaplan-Meier estimator. [27] pre-smoothed Kaplan-Meier estimator using binary regression and non-parametric regression. While Titman estimator reduces to Kaplan-Meier estimator in some cases, in many cases it does not. If we want to extend pre-smoothing to Titman estimator, we would need to find a way to pre-smooth in cases where it does not reduce to Kaplan-Meier estimator. [27] showed that when they specified the pre-smoothing

function correctly, the standard deviation of the non-parametric estimator decreased with bias slightly going up. However when the pre-smoothing function is misspecified, the bias of the non-parametric estimator heavily went up. Hence, pre-smoothing functions may decrease the standard deviation, but with risks. If a pre-smoothing function is misspecified, it could do more damage than good.

It is possible for pseudo-observations to be computed simultaneously to create multiple pseudo-observations. The original motivation of GEE was based on longitudinal studies. We believe that our method can be extended to longitudinal studies. With s being fixed and individual i having m timepoints, t_1,...,t_m, m pseudo-observations can be created as (21). Our method creates a regression model based on the snapshot of time t. The benefit of the longitudinal approach is that we would be able to see the change of the parameter estimates from the regression model over time.

When the landmark sample was small, it contributed to even higher standard deviation for the non-parametric Titman estimator. When using Titman estimator to create pseudo-observations, the standard deviation of the parameter estimates from GEE also was substantially large. We believe that a higher sample size will benefit both estimator in creating pseudo-observations. We see the standard deviation decrease as the sample size increases. We believe that it is even more crucial when using Titman estimator, since we see major improvements as the sample size increases. Higher sample size is good, but we also need to make sure that each landmark subsample is abundantly large. In the liver cirrhosis dataset, the overall sample size was 488 which is fairly large. However, the landmark subsample when individuals had low prothrombin level at $s = 1000$ was only 61. Ultimately, we would recommend the overall sample size to be at least 500. We also recommend the landmark subsample size to be at least 100 as a bare-minimum, and sample size of over 150 would be ideal. We believe that bias is a bigger issue than higher standard deviation since standard

deviation will decrease as the sample size increases. Bias will remain an issue for any sample size. However regarding higher sample sizes, practitioners would need to consider the financial and time aspect of the study. This is because higher sample sizes will make the study financially expensive and more time-consuming.

References

[1] Aalen, O., and Johansen, S. (1978). An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations.*Scandinavian Journal of Statistics*. 5(3), 141-150.

[2] Allignol, A., Beyersmann, J., Gerds, T., and Latouche, A.(2014). A competing risks approach for non-parametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis*. 20, 495-513.

[3] Andersen, P., Borgan, Ø., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Process*. Springer, New York

[4] Andersen, P., Klein, J., and Rosthoj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models.*Biometrika*. 90(1), 15-27.

[5] Azarang, L., Scheike, T. ,and de Uña-Alvarez J. (2017). Direct modeling of regression effects for transition probabilities in the progressive illness-death model. *Statistics in Medicine*. 36, 1964-1976.

[6] Balan, T., and Putter, H. (2020). A tutorial on frailty models. *Statistical Methods in Medical Research*. 29(11), 3424-3454.

[7] Beyersmann, J., Di Termini, S., and Pauly, M. (2013). Weak Convergence of the Wild Bootstrap for the Aalen-Johansen Estimator of the Cumulative Incidence Function of a Competing Risk. *Scandinavian Journal of Statistics*. 40(3), 387-402.

[8] Commenges, D., and Andersen, P. (1995). Score test of homogeneity for survival data. *Lifetime Data Analysis*. 1, 145-156.

[9] Datta, S., and Satten, G. (2001). Validity of the Aalen–Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-markov models. *Statistics & Probability Letters*. 55(4), 403-411.

[10] de Uña-Alvarez J., and Meira-Machado, L. (2015). Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. *Biometrics*. 71(2), 364-375.

[11] de Wreede, L., Fiocco, M., and Putter, H. (2011). mstate: An R Package for the Analysis of Competing Risks and Multi-State Models. *Journal of Statistical Software*. 38(7), 1-30.

[12] Halekoh, U., Højsgaard, S., and Yan, J. (2006). The R Package geepack for Generalized Estimation Equations. *Journal of Statistical Software*. 15(2), 1-11.

[13] Hoff, R., Putter, H., Mehlum, I., and Gran, J. (2019). Landmark estimation of transiton probabilities in non-Markov multi-state models with covariates. *Lifetime Data Analysis*. 25, 660-680.

[14] Jones, E., Epstein, D., and Garcia-Mochon, L. (2017). A Procedure for Deriving Formulas to Convert Transition Rates to Probabilities for Multistate Markov models. *Medical Decision Making*. 37(7), 779-789.

[15] Karrer, B., and Newman, M. (2010). A message passing approach for general epidemic models. *Physical Review E*. 82, 1-10.

[16] Keeling, M., and Eames, K. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*. 2(4), 295-307.

[17] Liang, K., and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*. 73(1), 13-22.

[18] Lin, D., Wei, L., and Ying, Z. (1993). Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika*. 80(3), 557-572.

[19] Liu, R. (1988). Bootstrap Procedures under some Non-I.I.D. Models. *The Annals of Statistics*. 16(4), 1696-1708.

[20] Mannucci, P. (1970). Estimation of prothrombin in liver disease. *Journal of Clinical Pathology*. 23(4), 291-295.

[21] Meira-Machado, L., de Uña-Alvarez J., and Cadarso-Suarez, C. (2006). Non-parametric estimation of transition probabilities in a non-Markov illness–death model. *Lifetime Data Analysis*. 12, 325-344.

[22] Miller, J., Slim, A., and Volz, E. (2012). Edge-based compartmental modelling for infectious disease spread. *Journal of the Royal Society Interface*. 9(70), 890-906.

[23] Putter, H., and Spitoni, C. (2018). Non-parametric estimation of transition probabilities in non-Markov multi-state models: The Landmark Aalen-Johansen estimator. *Statistical Methods in Medical Research*. 27(7), 2081-2092.

[24] Rodriguez-Girondo, M. and de Uña-Alvarez J. (2012). A non-parametric test for Markovianity in the illness-death model. *Statistics in Medicine*. 31(30), 4416-4427.

[25] Scheike, T., Zhang, M., and Gerds, T. (2008). Predicting cumulative incidence probability by direct binomial regression.*Biometrika*. 95(1), 205-220.

[26] Sherborne, N., Miller, J., Blyuss, K., and Kiss, I. (2018). Mean-field models for non-Markovian epidemics on networks. *Journal of Mathematical Biology*. 76, 755-778.

[27] Soutino, G., Meira-Machado, L., and Oliveira, P. (2022). A comparison of pres-moothing methods in the estimation of transition probabilities. *Communication in Statistics-Simulation and Computation*. 51(9), 5202-5221.

[28] Titman, A. (2015). Transition probability estimates for non-markov multi-state models. *Biometrics*. 71(4),1034–1041.

[29] Titman, A., and Putter, H. (2022). General tests of the Markov property in multi-state models. *Biostatistics*. 23(2), 380-396.

[30] Tripodi, A., Caldwell, S., Hoffman, M., Trotter J.,and Sanyal, A. (2007). Review article: the prothrombin time test as a measure of bleeding risk and prognosis in liver disease. *Alimentary Pharmacology & Therapeutics*. 26, 141-148.

[31] Wedderburn, R. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*. 61(3), 439-447.

[32] Wu, C. (1986).Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics*. 14(4), 1261-1295.

Appendix

Proof of Theorem 1

Theorem 1 states:

$\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically normal with mean zero and covariance matrix estimated $\hat{\Sigma} = \mathcal{I}(\hat{\beta})^{-1}\widehat{var}(U(\beta))\mathcal{I}(\hat{\beta})^{-1}$ where $\mathcal{I}(\beta) = \sum_i \frac{\partial\theta_i}{\partial\beta}^T V_i^{-1} \frac{\partial\theta_i}{\partial\beta}$ and $\widehat{var}(U(\beta)) = \sum_i U_i(\hat{\beta})U_i(\hat{\beta})^T$. $\mathcal{I}(\beta)$ is a $p \times p$ information matrix.

Note:

I.We will utilize the sketch of the proof from [17] and complete the proof.

II. In our application, we only need one pseudo-observation to create a regression model. However, in this proof, we prove the theorem in a general case where there are multiple pseudo-observations/time points.

III. Recall the working covariance matrix structure is $V_i = c(\phi)E_i^{1/2}R_i(\gamma)E_i^{1/2}$. Also, $\theta_i = \theta_i(Z_i) = E(h(X_i|Z_i)$. θ_i is the mean function.

IV. Important Assumptions include:

(1) $E(U(\beta) = 0$.

(2)$U_i(\beta)$ $i = 1, ..., n$ are independent.

(3)

i) $\widehat{\gamma}(\beta, \phi)$ is a $\sqrt{n}$ consistent estimator when β and ϕ are known. Hence, $\sqrt{n}(\widehat{\gamma}(\beta, \phi(\beta)) - \gamma) = O_p(1)$

ii) $\widehat{\phi}(\beta)$ is a $\sqrt{n}$ consistent estimator when β is known. Hence, $\sqrt{n}(\widehat{\phi}(\beta) - \phi) = O_p(1)$.

iii) $|\frac{\partial\widehat{\gamma}(\beta,\phi)}{\partial\phi}| = O_p(1)$.

(4) We assume mild regularity conditions.

Proof:

Consider the first-order Taylor series expansion about β,

$$\sum_{i=1}^{n} U_i\left(\widehat{\beta},\widehat{\gamma}(\widehat{\beta},\widehat{\phi}(\widehat{\beta}))\right) \approx \sum_{i=1}^{n} U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right) + \sum_{i=1}^{n} \frac{\partial^* U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)}{\partial\beta}(\widehat{\beta}-\beta)$$

$$\sum_{i=1}^{n} U_i\left(\widehat{\beta},\widehat{\gamma}(\widehat{\beta},\widehat{\phi}(\widehat{\beta}))\right) - \sum_{i=1}^{n} U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right) \approx \sum_{i=1}^{n} \frac{\partial^* U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)}{\partial\beta}(\widehat{\beta}-\beta)$$

$$-\sum_{i=1}^{n} U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right) \approx \sum_{i=1}^{n} \frac{\partial^* U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)}{\partial\beta}(\widehat{\beta}-\beta)$$

$$(\widehat{\beta}-\beta) \approx \left(\sum_{i=1}^{n} \frac{\partial^* U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)}{\partial\beta}\right)^{-1}\left(\sum_{i=1}^{n} U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)\right)$$

$$\sqrt{n}(\widehat{\beta}-\beta) \approx \left(\sum_{i=1}^{n} -\frac{\frac{\partial^* U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)}{n}}{\partial\beta}\right)^{-1}\left(\sum_{i=1}^{n} \frac{U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)}{\sqrt{n}}\right)$$

Note:

$$\frac{\partial^* U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)}{\partial\beta} = \frac{\partial\, U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)}{\partial\beta} + \frac{\partial\, U_i\left(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))\right)}{\partial\,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))}\left(\frac{\partial\,\widehat{\gamma}(\beta,\widehat{\phi}(\beta))}{\partial\beta}\right)$$

$=A_i + B_i C.$

Let β be fixed and do a Taylor series expansion about γ.

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{U_i\left(\beta, \widehat{\gamma}(\beta, \widehat{\phi}(\beta))\right)}{\sqrt{n}} \\
&= \sum_{i=1}^{n} \frac{U_i(\beta,\gamma)}{\sqrt{n}} + \sum_{i=1}^{n} \frac{\frac{\partial U_i(\beta,\gamma)}{\sqrt{n}}}{\partial\gamma}\left(\widehat{\gamma}(\beta, \widehat{\phi}(\beta)) - \gamma\right) + o_p(1) \\
&= \sum_{i=1}^{n} \frac{U_i(\beta,\gamma)}{\sqrt{n}} + \sum_{i=1}^{n} \frac{\frac{\partial U_i(\beta,\gamma)}{\partial\gamma}}{n}\sqrt{n}\left(\widehat{\gamma}(\beta, \widehat{\phi}(\beta)) - \gamma\right) + o_p(1) \\
&= A^* + B^*C^* + o_p(1)
\end{aligned}$$

Consider,

$$B^* = \sum_{i=1}^{n} \frac{\frac{\partial U_i(\beta,\gamma)}{\partial\gamma}}{n}$$

For the ith individual, $\frac{\partial U_i(\beta,\gamma)}{\partial\gamma} = \sum_i \left(\frac{\partial\theta_i}{\partial\beta}\right)^T V_i^{-1}\frac{\partial V_i^{-1}}{\partial\gamma}V_i^{-1}(\widehat{\theta}_i - \theta_i)$ where $\frac{\partial V_i^{-1}}{\partial\gamma}$ is $k_i \times k_i$ matrix where a derivative is taken for each element with respect to γ. We see that $\frac{\partial U_i(\beta,\gamma)}{\partial\gamma}$ is a linear function of $\widehat{\theta}_i - \theta_i$. Then, $E(\widehat{\theta}_i - \theta_i) = 0$. We assume that the processes, $X_1(t)$,...,$X_n(t)$ are i.i.d. Therefore, the individuals are i.i.d. By Strong Law of large numbers, $B^* = \sum_{i=1}^{n} \frac{\frac{\partial U_i(\beta,\gamma)}{\partial\gamma}}{n}$ converges to the mean which is 0. Thus, $B^* = o_p(1)$.

Consider,

$$\begin{aligned}
C^* &= \sqrt{n}\left(\widehat{\gamma}(\beta, \widehat{\phi}(\beta)) - \gamma\right) \\
&= \sqrt{n}\left(\widehat{\gamma}(\beta, \widehat{\phi}(\beta)) - \widehat{\gamma}(\beta, \phi(\beta)) + \widehat{\gamma}(\beta, \phi(\beta)) - \gamma\right)
\end{aligned}$$

By identifying a first-order Taylor series expansion from the first two terms,

$$= \sqrt{n}\left(\frac{\partial\,\widehat{\gamma}(\beta,\phi(\beta))}{\partial\phi}(\widehat{\phi}(\beta)-\phi)+\widehat{\gamma}(\beta,\phi(\beta))-\gamma\right)$$

Using assumptions (3) i)-iii)

$$\sqrt{n}\left(\frac{\partial\,\widehat{\gamma}(\beta,\phi(\beta))}{\partial\phi}(\widehat{\phi}(\beta)-\phi)+\widehat{\gamma}(\beta,\phi(\beta))-\gamma\right)=O_p(1).$$

Consider $A^*=\sum_{i=1}^{n}\frac{U_i(\beta,\gamma)}{\sqrt{n}}=\sqrt{n}\sum_{i=1}^{n}\frac{U_i(\beta,\gamma)}{n}$. By assumption (1), $E(U_i(\beta,\gamma))=0$.

$$Cov(U_i(\beta,\gamma))=Cov\left(\left(\frac{\partial\theta_i}{\partial\beta}\right)^T V_i^{-1}(\widehat{\theta}_i-\theta_i)\right)$$
$$=\left(\frac{\partial\theta_i}{\partial\beta}\right)^T V_i^{-1}Cov(\widehat{\theta}_i)V_i^{-1}\left(\frac{\partial\theta_i}{\partial\beta}\right)$$

Note that $Cov(\widehat{\theta}_i)$ is a $k_i\times k_i$ matrix where $Cov(\widehat{\theta}_{il},\widehat{\theta}_{il'})$ for $l,l'\in 1,...,k_i$. Since $U_i(\beta,\gamma)$ are independent from assumption (2),

$$Cov\left(\sum_{i=1}^{n}U_i(\beta,\gamma)\right)=\sum_{i=1}^{n}\left(\frac{\partial\theta_i}{\partial\beta}\right)^T V_i^{-1}Cov(\widehat{\theta}_i)V_i^{-1}\left(\frac{\partial\theta_i}{\partial\beta}\right)$$

By Central Limit Theorem, $\sqrt{n}\sum_{i=1}^{n}U_i(\beta,\gamma)$ has an asymptotic distribution of normal with mean 0 and covariance matrix, $\sum_{i=1}^{n}\left(\frac{\partial\theta_i}{\partial\beta}\right)^T V_i^{-1}Cov(\widehat{\theta}_i)V_i^{-1}\left(\frac{\partial\theta_i}{\partial\beta}\right)$. Consider

$$A_i=\frac{\partial U_i(\beta,\widehat{\gamma}(\beta,\widehat{\phi}(\beta)))}{\partial\beta}.$$

The derivative of the score function, $\left(\frac{\partial\theta_i}{\partial\beta}\right)^T V_i^{-1}(\widehat{\theta}_i-\theta_i)$ with respect to β is $-\left(\frac{\partial\theta_i}{\partial\beta}\right)^T V_i^{-1}\left(\frac{\partial\theta_i}{\partial\beta}\right)$.

Consider

$$B_i = \frac{\partial U_i(\beta, \widehat{\gamma}(\beta, \widehat{\phi}(\beta)))}{\partial\, \widehat{\gamma}(\beta, \widehat{\phi}(\beta))}.$$

B_i will be a constant since β is known. Then, $\sum_i B_i$ would be a constant. Hence, $\sum_i B_i = o_p(n)$. Consider

$$C = \frac{\partial\, \widehat{\gamma}(\beta, \widehat{\phi}(\beta))}{\partial \beta}.$$

C is a constant. Therefore, $C = O_p(1)$.

Looking back at the original Taylor series expansion at the beginning, we are left with

$$\left(\frac{\sum_{i=1}^n \left(\frac{\partial \theta_i}{\partial \beta}\right)^T V_i^{-1} \left(\frac{\partial \theta_i}{\partial \beta}\right)}{n} \right)^{-1} \frac{\sqrt{n} \sum_{i=1}^n U_i(\beta, \gamma)}{n}$$
$$= \left(\sum_{i=1}^n \left(\frac{\partial \theta_i}{\partial \beta}\right)^T V_i^{-1} \left(\frac{\partial \theta_i}{\partial \beta}\right) \right)^{-1} \sqrt{n} \sum_{i=1}^n U_i(\beta, \gamma)$$

Note that $\sum_{i=1}^n \left(\frac{\partial \theta_i}{\partial \beta}\right)^T V_i^{-1} \left(\frac{\partial \theta_i}{\partial \beta}\right)$ is the information matrix, $\mathcal{I}(\beta)$.

The asymptotic covariance would be

$$\begin{aligned}
\Sigma &= \left(\sum_{i=1}^n \left(\frac{\partial \theta_i}{\partial \beta}\right)^T V_i^{-1} \left(\frac{\partial \theta_i}{\partial \beta}\right) \right)^{-1} \sum_{i=1}^n \left(\frac{\partial \theta_i}{\partial \beta}\right)^T V_i^{-1} Cov(\widehat{\theta}_i) V_i^{-1} \left(\frac{\partial \theta_i}{\partial \beta}\right) \\
&\times \left(\sum_{i=1}^n \left(\frac{\partial \theta_i}{\partial \beta}\right)^T V_i^{-1} \left(\frac{\partial \theta_i}{\partial \beta}\right) \right)^{-1} \\
&= \mathcal{I}(\beta)^{-1} Var(U(\beta)) \mathcal{I}(\beta)^{-1}
\end{aligned}$$

By Central Limit Theorem, $\sqrt{n}(\widehat{\beta} - \beta)$ has an asymptotic normal distribution with

mean 0 and covariance matrix, Σ. To estimate Σ, we plug in $\widehat{\beta}$ for β. We estimate $Cov(\widehat{\theta}_i)$ with $(\widehat{\theta}_i - \theta_i)(\widehat{\theta}_i - \theta_i)^T$. Then $\widehat{Var}(U(\beta)) = \sum_i U_i(\widehat{\beta})(U_i(\widehat{\beta})^T$. So, $\widehat{\Sigma} = \mathcal{I}(\widehat{\beta})^{-1}\widehat{Var}(U(\beta))\mathcal{I}(\widehat{\beta})^{-1}$. $\square$