

Simulation from Structural Survival Models under Complex Time-Varying Data Structures

Jessica G. Young*

Miguel A. Hernán*[†]

Sally Picciotto*

James M. Robins*[‡]

Abstract

Standard approaches to estimating the effect of a time-varying exposure on survival may be biased in the presence of time-varying confounders themselves affected by prior exposure. Methods involving estimation of structural models are becoming more widely used alternatives that do not suffer from this bias. In the case of survival outcomes, these include inverse probability weighted estimation of Cox Marginal Structural Models (Cox MSM) and g-estimation of Structural Nested Accelerated Failure Time Models (SNAFTM). While it is known how to simulate from a SNAFTM in the presence of time-dependent confounders affected by prior exposure, it was previously unknown how to simulate from a Cox MSM in this case. Here we describe how to appropriately simulate data from a Cox MSM under this data structure.

Key Words: causal inference, survival analysis, simulation, marginal structural models, structural nested models

1. Introduction

Standard approaches to estimating the causal effect of a time-varying treatment or exposure on survival (e.g., by fitting a time-dependent Cox model for the hazard at time t conditional on a function of treatment and covariate history through time t [Cox and Oakes, 1984]) may be biased in the presence of a time-varying covariate [Robins, 1986, Hernán et al., 2004] that is

1. a time-dependent confounder, i.e., it affects both future risk of failure and treatment and
2. affected by past treatment.

Robins and collaborators have developed methods to appropriately adjust for measured time-varying confounders that are affected by past treatment [Robins and Hernán, 2009]. In the failure time setting, these methods include inverse probability weighting (IPW) of Cox Marginal Structural Models (Cox MSM) [Robins, 1998, Hernán et al., 2000] and g-estimation of Structural Nested Accelerated Failure Time Models (SNAFTM) [Robins et al., 1992, 1993, Hernán et al., 2005]. The relative performance of these models, as well as the adequacy of the software used to estimate their parameters, can be evaluated via simulated data in which the causal effect of interest is known. Although methods for simulating from a SNAFTM have been published in the literature [Robins, 1992], no method for simulating from a Cox MSM has been published. This paper provides such a method.

In §2 we define the data structure of interest. In §3 we formally define the SNAFTM and Cox MSM. In §4 we review how complex longitudinal data with time-dependent confounders can be generated from a SNAFTM. In §5 we describe how complex longitudinal data with time-dependent confounders can be generated from a Cox MSM. In §6 we provide conclusions.

2. General Data Structure

2.1 Observed data structure

We consider data from a longitudinal study where measurements are taken at interval times $0, 1, 2, \dots, m, \dots, K, K+1$. We assume we observe n i.i.d. copies of

$$O = \{\bar{L}_{K+1}, \bar{A}_K\}$$

where

$$\bar{L}_{K+1} = \{\bar{Y}_{K+1}, \bar{V}_K\};$$

$\bar{V}_K = (V_0, V_1, \dots, V_m, \dots, V_K)$ with V_m a time-varying covariate measured at the start of the interval $[m, m+1)$ and assumed constant in this interval; $\bar{A}_K = (A_0, A_1, \dots, A_m, \dots, A_K)$ with A_m a dichotomous treatment ($1 = \textit{treated}$, $0 = \textit{untreated}$) measured immediately after V_m and assumed constant over the interval $[m, m+1)$; $Y_m = I(T < m)$

*Department of Epidemiology, Harvard School of Public Health

[†]Harvard-MIT Division of Health Sciences and Technology

[‡]Department of Biostatistics, Harvard School of Public Health

where T is a failure time variable; $\bar{Y}_{K+1} = (Y_u; 0 \leq u < K+1)$ with $Y_u = I(T < u)$ if T is exactly observed and $\bar{Y}_{K+1} = (Y_0, Y_1, \dots, Y_m, \dots, Y_K, Y_{K+1})$ if T is interval-censored. By convention, $Y_0 = 0$ and if $Y_m = 1$ then $V_m = 0$, $A_m = 0$, and $Y_{m+1} = 1$.

For simplicity, we do not allow for right-censoring in our discussion. However, the nature of the simulation problem described below is unaffected by the presence of such censoring. For more on how censoring is accounted for in estimation methods for structural survival models see Hernán et al. [2000, 2005].

2.2 Full data structure

Let $g = \bar{a}$ for $\bar{a} \equiv \bar{a}_K$ in the support of $\bar{A}_K$ denote a *treatment regime*. An example of a treatment regime is the regime $g = (1, 1, \dots, 1) = \bar{1}$ denoting *continuous treatment*. Let T_g and $\bar{V}_{K,g}$ represent the failure time and covariate history, respectively, a subject would have experienced had she, possibly contrary to fact, followed treatment regime $g = \bar{a}$. We say a subject follows treatment regime $\bar{a}$ if the subject takes treatment a_m at time m if alive at m and, by convention, treatment 0 if dead (i.e., failed). We can think of the observed data structure as a missing data structure where the full, unobserved data structure consists of the observed data, the counterfactual covariate history $\bar{V}_{K,g}$ and counterfactual failure time T_g for all possible $g = \bar{a}$. Given $g = \bar{a}$, we will use $T_g, T_{\bar{a}}$ and $T_{g=\bar{a}}$ interchangeably.

We assume the following three identifying assumptions [Robins and Hernán, 2009]:

1. Consistency:

Let $\bar{a}_m$ be the first m components of $\bar{a}$. Given $g = \bar{a}$, if $\bar{A}_m = \bar{a}_m$ and either $T < m+1$ or $T_g < m+1$ then $T_g = T$, $\bar{Y}_{m+1,g} = \bar{Y}_{m+1}$ and $\bar{V}_{m+1,g} = \bar{V}_{m+1}$ where

$$Y_{m,g} = I(T_g < m),$$

$$\bar{Y}_{K+1,g} = (Y_{0,g}, \dots, Y_{K+1,g})$$

in the interval-censored case and

$$\bar{Y}_{K+1,g} = (Y_{u,g}; 0 \leq u \leq K+1)$$

with

$$Y_{u,g} = I(T_g < u)$$

in the exactly observed case.

2. Conditional exchangeability: For any regime g and $m \in [0, K]$

$$(T_g, \bar{V}_{K+1,g}) \perp\!\!\!\perp A_m | \bar{V}_m, \bar{A}_{m-1}, Y_m = 0$$

3. Positivity: $f_{\bar{A}_{m-1}, \bar{V}_m, Y_m}(\bar{a}_{m-1}, \bar{v}_m, 0) \neq 0 \implies \Pr(A_m = a_m | \bar{V}_m, \bar{A}_{m-1}, Y_m = 0) > 0$ w.p.1 for all a_m in the support of A_m , $0 \leq m \leq K$.

The assumption of conditional exchangeability is also known as the assumption of sequential randomization or the assumption of no unmeasured confounding given the measured covariates $\bar{V}_m$. Note that (unconditional) exchangeability, or no confounding due to either measured or unmeasured covariates, is expressed as

$$T_g \perp\!\!\!\perp A_m | \bar{A}_{m-1}, Y_m = 0 \quad (1)$$

Figure 1 depicts, for a subset of time points, a causal directed acyclic graph (DAG) [Pearl, 1995, Spirtes et al., 1993] for the interval-censored case. The vertices (nodes) of the causal DAG represent variables, both measured and unmeasured. A causal DAG is one in which the directed edges (arrows) represent direct causal effects and all common causes of any pair of variables are included in the graph. The node U represents unobserved determinants of failure time and V_m . The causal DAG in Figure 1 implies conditional exchangeability because there are no arrows from U directly into treatment at any time. The V_m are measured time-dependent confounders that are also affected by previous treatment $\bar{A}_{m-1}$.

3. Structural failure time models

Let $T_{\bar{0}}$ be the counterfactual failure time under the treatment regime $g = \bar{0} \equiv \bar{0}_K = (0, 0, \dots, 0)$ (*never treated during follow-up*). We define models such that

$$T_{\bar{0}} \text{ and } \int_0^{T_{\bar{a}}} \exp\{\gamma_{aft}(t, \bar{a}_t, \bar{V}_{t,\bar{a}}, \psi_{aft})\} dt \text{ have the same distribution} \quad (2)$$

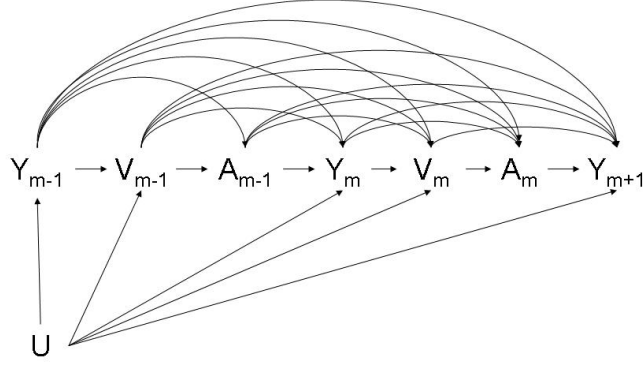


Figure 1: Directed Acyclic Graph (DAG) Representing the Full Data Structure.

and

$$\lambda_{T_{\bar{a}}}(t) = \lambda_0(t) \exp\{\gamma_{msm}(t, \bar{a}_t, \psi_{msm})\}, \quad (3)$$

as a SNAFTM and Cox MSM, respectively, where $\gamma_{aft}(t, \bar{a}_t, \bar{v}_t, \psi_{aft})$ and $\gamma_{msm}(t, \bar{a}_t, \psi_{msm})$ are known functions, continuous in t and differentiable wrt to t except at $t = 1, 2, \dots, K$; $\lambda_{T_{\bar{a}}}(t)$ and $\lambda_0(t)$ are the hazard functions for $T_{\bar{a}}$ and $T_{\bar{0}}$, respectively, evaluated at t ; and ψ_{aft} , ψ_{msm} denote parameter vectors encoding the causal effect of treatment on survival in their corresponding models. For a detailed description of these models and their estimation procedures see Hernán et al. [2000, 2005].

4. SNAFTM data generation

Robins [1992] shows that data adhering to the SNAFTM (2) with $\psi_{aft} = \psi$ can be generated under identifying assumptions 1-3 of §2.2 using the following algorithm.

Algorithm: For each of n simulated subjects:

step 1: Simulate the counterfactual $T_{\bar{0}}$ from a failure time distribution (e.g. Weibull, Exponential) with hazard $\lambda_0(t)$.

Then for each $m \in [0, K]$ implement steps 2-4:

step 2: Simulate V_m from some distribution conditional on a function of past variables. This function must include $T_{\bar{0}}$ and components of $\bar{A}_{m-1}$ if we wish to guarantee that V_m is a confounder and affected by prior treatment. For example, for binary V_m , we might assume the logistic regression model $\text{logit}[\Pr(V_m = 1 | \bar{V}_{m-1}, \bar{A}_{m-1}, T_{\bar{0}}, Y_m = 0; \beta)] = \beta_0 + \beta_1 T_{\bar{0}} + \beta_2 A_{m-1} + \beta_3 V_{m-1}$.

step 3: Simulate A_m from some distribution conditional on a function of past variables. This function must include some function of $\bar{V}_m$ in order to guarantee that V_m is a confounder and must exclude $T_{\bar{0}}$ to guarantee conditional exchangeability. For example, we might assume the logistic regression model $\text{logit}[\Pr(A_m = 1 | \bar{V}_m, \bar{A}_{m-1}, Y_m = 0; \alpha)] = \alpha_0 + \alpha_1 V_m + \alpha_2 V_{m-1} + \alpha_3 A_{m-1}$.

step 4: Given a function $\gamma_{aft}(t, \bar{A}_t, \bar{V}_t, \psi)$, a choice of ψ , the generated values of $T_{\bar{0}}$, $\bar{V}_m$, $\bar{A}_m$ from steps 1-3, and setting $V_t = V_m$ and $A_t = A_m$ for $t \in [m, m+1)$, let T be the solution to

$$T_{\bar{0}} = \int_0^T \exp\{\gamma_{aft}(t, \bar{A}_t, \bar{V}_t, \psi)\} dt \quad (4)$$

Finally, redefine $V_l = 0, A_l = 0$ for $l > T$.

Often $\gamma_{aft}(t, \bar{A}_t, \bar{V}_t, \psi)$ can be written as $\gamma_{aft}^*(\bar{A}_{int(t)}, \bar{V}_{int(t)}, \psi)$ for $int(t)$ the largest integer less than t . In that case, the solution T to the previous equation can be obtained exactly with the following algorithm:

- if $T_0 > \int_0^{m+1} \exp\{\gamma_{aft}^*(\bar{A}_{int(t)}, \bar{V}_{int(t)}, \psi)\} dt$ then $Y_{m+1} = 0$;
- else if $T_0 \leq \int_0^{m+1} \exp\{\gamma_{aft}^*(\bar{A}_{int(t)}, \bar{V}_{int(t)}, \psi)\} dt$ then $Y_{m+1} = 1$ and $T \in (m, m+1]$ with $T = m + (T_0 - \int_0^m \exp\{\gamma_{aft}^*(\bar{A}_{int(t)}, \bar{V}_{int(t)}, \psi)\} dt) \exp\{-\gamma_{aft}(\bar{A}_m, \bar{V}_m, \psi)\}$.

5. Cox MSM data generation

We now prove that data adhering to the Cox MSM (3) with $\psi_{msm} = \psi$ and baseline hazard $\lambda_0(t)$ can be generated under assumptions 1-3 of §2.2 by the algorithm described in the previous section except that step 4 is modified such that T now solves the equation

$$T_0 = h(T, \bar{a}_T, \psi)$$

where $h(t, \bar{a}_t, \psi)$ is the solution to

$$\int_0^{h(t, \bar{a}_t, \psi)} \lambda_0(u) du = \int_0^t \lambda_0(u) \exp\{\gamma_{msm}(u, \bar{a}_u, \psi)\} du.$$

We further prove that this last result can be restated as follows: Data adhering to the Cox MSM (3) with $\psi_{msm} = \psi$ and baseline hazard $\lambda_0(t)$ can be generated under assumptions 1-3 of §2.2 by simulating data from a SNAFTM with parameter ψ such that

$$\gamma_{aft}(t, \bar{a}_t, \bar{v}_t, \psi) = \ln \left\{ \frac{\lambda_0(t)}{\lambda_0\{h(t, \bar{a}_t, \psi)\}} \right\} + \gamma_{msm}(t, \bar{a}_t, \psi) \quad (5)$$

In particular if T_0 has an exponential distribution so $\lambda_0(t)$ does not depend on t , $\gamma_{aft}(t, \bar{a}_t, \bar{v}_t, \psi) = \gamma_{msm}(t, \bar{a}_t, \psi)$.

Proof: Under a Cox MSM and our identifying assumptions, it is known that $S_{\bar{a}}(t) \equiv \Pr[T_{\bar{a}} > t] = \exp\left\{-\int_0^t \lambda_0(u) \exp\{\gamma_{msm}(u, \bar{a}_u, \psi)\} du\right\}$. Thus $S_{\bar{a}}^{-1}(S_{\bar{a}}(T_{\bar{a}})) = h(T_{\bar{a}}, \bar{a}_{T_{\bar{a}}}, \psi)$ and T_0 have the same marginal distribution, where $S_0(t) \equiv \Pr[T_0 > t]$. The Cox MSM assumption plus our identifying assumptions place no further restrictions on the joint distribution of $T_{\bar{a}}$ and T_0 . Thus we may assume they have a degenerate joint distribution satisfying $T_0 = h(T_{\bar{a}}, \bar{a}_{T_{\bar{a}}}, \psi)$. But our algorithm generates data from a joint distribution satisfying the three identifiability conditions plus $T_0 = h(T_{\bar{a}}, \bar{a}_{T_{\bar{a}}}, \psi)$.

To find $\gamma_{aft}(t, \bar{a}_t, \bar{v}_t, \psi)$ that generates this same data we must find $\gamma_{aft}(t, \bar{a}_t, \bar{V}_{t, \bar{a}}, \psi_{aft})$ that satisfies

$$h(T_{\bar{a}}, \bar{a}_{T_{\bar{a}}}, \psi) = \int_0^{T_{\bar{a}}} \exp\{\gamma_{aft}(t, \bar{a}_t, \bar{V}_{t, \bar{a}}, \psi_{aft})\} dt.$$

Differentiating both sides wrt $T_{\bar{a}}$ and evaluating at $T_{\bar{a}} = t$, we obtain

$$dh(t, \bar{a}_t, \psi) / dt = \exp\{\gamma_{aft}(t, \bar{a}_t, \bar{V}_{t, \bar{a}}, \psi_{aft})\}.$$

To evaluate $dh(t, \bar{a}_t, \psi) / dt$, we differentiate the identity

$$\int_0^{h(t, \bar{a}_t, \psi)} \lambda_0(u) du = \int_0^t \lambda_0(u) \exp\{\gamma_{msm}(u, \bar{a}_u, \psi)\} du$$

wrt to t to obtain

$$\{\lambda_0(h(t, \bar{a}_t, \psi))\} dh(t, \bar{a}_t, \psi) / dt = \lambda_0(t) \exp\{\gamma_{msm}(t, \bar{a}_t, \psi)\}.$$

Solving for $dh(t, \bar{a}_t, \psi) / dt$, we conclude that

$$\lambda_0(t) \exp\{\gamma_{msm}(t, \bar{a}_t, \psi)\} / \lambda_0(h(t, \bar{a}_t, \psi)) = \exp\{\gamma_{aft}(t, \bar{a}_t, \bar{V}_{t, \bar{a}}, \psi_{aft})\}. \quad (6)$$

The result (5) follows from taking the natural log of both sides of (6).

6. Conclusions

Despite the increasing popularity of the Cox MSM, no description of how to correctly simulate from such models has appeared in the literature. This has led some users to mistakenly conclude that these methods simply “don’t work” because they have used an incorrect simulation method in conducting simulation studies (see the appendix for an example of an incorrect simulation method). This paper has corrected this problem by describing how to appropriately simulate from a Cox MSM.

References

- D.R. Cox and D. Oakes. *Analysis of Survival Data*. London: Chapman and Hall, 1984.
- M.A. Hernán, B. Brumback, and J.M. Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5):561–570, 2000.
- M.A. Hernán, S. Hernández-Díaz, and J.M. Robins. A structural approach to selection bias. *Epidemiology*, 15: 615–625, 2004.
- M.A. Hernán, S.R. Cole, J. Margolick, M. Cohen, and J.M. Robins. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and Drug Safety*, 14(7):477–491, 2005.
- J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82:669–710, 1995.
- J.M. Robins. A new approach to causal inference in mortality studies with a sustained exposure period: application to the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512, 1986.
- J.M. Robins. Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, 79:321–34, 1992.
- J.M. Robins. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology*, pages 95–133. New York: Springer, 2000.
- J.M. Robins. Marginal structural models. In *1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, pages 1–10. American Statistical Association, 1998.
- J.M. Robins and M.A. Hernán. Estimation of the causal effects of time-varying exposures. In G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, editors, *Advances in Longitudinal Data Analysis*, pages 553–599. Boca Raton, FL: Chapman and Hall/CRC Press, 2009.
- J.M. Robins and L. Wasserman. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In D. Geiger and P. Shenoy, editors, *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 409–420. San Francisco: Morgan Kaufmann, 1997.
- J.M. Robins, D. Blevins, G. Ritter, and M. Wulfsohn. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*, 3:319–336, 1992.
- J.M. Robins, D. Blevins, G. Ritter, and M. Wulfsohn. Errata to g-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*, 4:189, 1993.
- P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search*. New York: Springer-Verlag, 1993.

Appendix

Here we present an example of a commonly used, yet incorrect, approach to Cox MSM data generation in the presence of time varying confounders affected by prior treatment. Consider the following algorithm:

For each $m \in [0 : K]$

step1: Simulate V_m by assuming a parametric model for its distribution conditional on past variables (e.g. the logistic regression model $\text{logit}[\Pr(V_m = 1 | \bar{V}_{m-1}, \bar{A}_{m-1}, Y_m = 0; \beta)] = \beta_0 + \beta_1 A_{m-1} + \beta_2 V_{m-1}$ with all components of β nonzero).

step 2: Simulate A_m by assuming a parametric model for its distribution conditional on past variables (e.g. the logistic regression model $\text{logit}[\Pr(A_m = 1 | \bar{V}_m, \bar{A}_{m-1}, Y_m = 0; \alpha)] = \alpha_0 + \alpha_1 V_m + \alpha_2 V_{m-1} + \alpha_3 A_{m-1}$ with all components of α nonzero).

step 3: Simulate Y_{m+1} by assuming a parametric model for its distribution conditional on past variables (e.g. the logistic regression model $\text{logit}[\Pr(Y_{m+1} = 1 | \bar{V}_m, \bar{A}_m, Y_m = 0; \theta)] = \theta_0 + \theta_1 V_m + \theta_2 V_{m-1} + \theta_3 A_m$ with all components of θ nonzero).

Despite the ubiquity of this type of approach, data generated in this way will generally fail to follow a given Cox MSM with a specified value for its parameter. To see this more explicitly, consider the Cox MSM $\gamma_{msm}(t, \bar{a}_t, \psi_{msm}) = \psi_{msm} \times a_t$ with $\psi_{msm} = 1$. Note that, for sufficiently small time steps, the Cox MSM $\gamma_{msm}(t, \bar{a}_t, \psi_{msm}) = \psi_{msm} \times a_t$ is well-approximated by the discrete hazard model $\Pr[Y_{m+1, \bar{a}} = 1 | Y_{m, \bar{a}} = 0] / \Pr[Y_{m+1, \bar{0}} = 1 | Y_{m, \bar{0}} = 0] =$

$\exp(\psi_{msm} \times a_m)$ for $t \in [m, m+1)$. Now for data generated under the above algorithm, $\Pr[Y_{m+1,\bar{a}} = 1 | Y_{m,\bar{a}} = 0]$ can further be written in terms of the observed data distribution using the parametric version of Robin's g-formula [2000]:

$$\Pr[Y_{m+1,\bar{a}} = 1 | Y_{m,\bar{a}} = 0; \theta, \beta] = \sum_{\bar{V}_m} \Pr[Y_{m+1} = 1 | \bar{V}_m, \bar{A}_m = \bar{a}_m, Y_m = 0; \theta] w_g(m, \bar{V}_m; \theta, \beta) / \sum_{\bar{V}_m} w_g(m, \bar{V}_m; \theta, \beta), \quad (7)$$

$$w_g(m, \bar{V}_m; \theta, \beta) = f(V_0; \beta) \prod_{j=1}^m \Pr[Y_j = 0 | \bar{V}_j, \bar{A}_{j-1} = \bar{a}_{j-1}, Y_{j-1} = 0; \theta] f[V_j | \bar{A}_{j-1} = \bar{a}_{j-1}, \bar{V}_{j-1}, Y_j = 0; \beta] \quad (8)$$

with the factors in (7) and (8) defined by the models cited in steps 1 and 3 of the algorithm. Our algorithm can only succeed in simulating data from the Cox MSM $\gamma_{msm}(t, \bar{a}_t, \psi_{msm}) = \psi_{msm} \times a_t$ with $\psi_{msm} = 1$ if there exists a parameter vector (θ, β) with all components non-zero satisfying $\Pr[Y_{m+1,\bar{a}} = 1 | Y_{m,\bar{a}} = 0; \theta, \beta] / \Pr[Y_{m+1,\bar{0}} = 1 | Y_{m,\bar{0}} = 0; \theta, \beta] = \exp(a_m)$ for all regimes $\bar{a}$ and times m . It can be shown that no such vector (θ, β) exists for $m > 5$ and thus the algorithm fails. The proof is similar to the proof of the so called g-null paradox in [Robins and Wasserman, 1997].

Note that under the relaxation of the assumption of non-zero parameters, this method of data generation will succeed. For example if we set $\theta_1 = \theta_2 = 0$ (i.e., $\bar{V}_m$ does not predict the risk of failure in the next interval given $\bar{A}_m$), then for sufficiently small time steps, the data will approximately follow a Cox MSM with ψ_{msm} nearly equal to θ_3 . However when $\theta_1 = \theta_2 = 0$, $\bar{V}_m$ is no longer a confounder for the effect of treatment on survival so fitting the time-dependent discrete hazard model $\text{logit}[\Pr(Y_{m+1} = 1 | \bar{A}_m, Y_m = 0; \theta)] = \theta_0 + \theta_3 A_m$ is essentially equivalent to fitting the Cox MSM model $\gamma_{msm}(t, \bar{a}_t, \psi_{msm}) = \psi_{msm} \times a_t$ with $\psi_{msm} = \theta_3$. As a final example, if $\beta_1 = 0$ (so V_m does not depend on prior treatment) and $\theta_3 = 0$, then $\Pr[Y_{m+1,\bar{a}} = 1 | Y_{m,\bar{a}} = 0; \theta, \beta]$ will not depend on the treatment regime $\bar{a}$ and thus the generated data will follow a Cox MSM with $\psi_{msm} = 0$.