Analyzing Dynamic Multiple Spell Durations Using Counting Processes*

Yuhao Li LIYUHAO.ECON@WHU.EDU.CN
Economics and Management School
Wuhan University

Draft Time: December, 2022. Click Here for Latest Version

Abstract

I propose a counting process approach to analyze multiple spell duration data. These data are doubly stochastic in a sense that both durations for an individual and the number of durations within a fixed period are random. I allow unobserved individual heterogeneities to enter into the model as fixed effects. In addition, conditional on the individual fixed effect, durations are state dependent. A first-difference transformation is developed to cancel fixed effects, and a minimum distance estimator is re-introduced with simplified proofs. Finite sample properties are investigated in simulations. The approach is applied to studying an individual's work absence decisions.

JEL Classification: C41, J22

Keywords: Multiple Spell Duration, Counting Process, State Dependence, Fixed Effect, Work Absence

1. Introduction

Duration data describes the amount of time that elapsed until a given event, or the length of time spent in a given state. Typical Economical and social phenomenons measured in terms of duration include duration of unemployment, duration of a strike, duration between two purchases of an individual, age of a woman at birth of first child, etc. Given an i.i.d sample $T_1, T_2, \ldots, T_n$ of durations from the distribution function F, one often characterizes a duration model through the hazard rate $h(t)$:

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

where f is the corresponding probability density function. $h(t)$ is also the basic quantity in the counting process approach to duration analysis, e.g., see Andersen et al. (2012). In this

*. I am grateful to Miguel A. Delgado for supports and guidance throughout this project, to Winfried Stute for his inspiration and valuable comments, and to Michael A. Nolan for sharing the work absence data. I also thank conference and seminar participants at the IAAE Nicosia,EWMES Naples, SAEe Madrid, UC3M, DUFE. All errors are mine.

approach, the duration data are represented as single-event counting processes $N_i(t), i = 1, \ldots, n$ where

$$N_i(t) = \mathbb{I}\{T_i \le t\}$$

counts 1 if the event $\{T_i \le t\}$ happens and otherwise $N_i(t) = 0$ throughout. The dynamics of $N_i(t)$ is described by its (random) intensity process which is a product:

$$\lambda_i(t|\mathcal{F}_i(t-)) = h(t)\mathbb{I}\{T_i > t\}$$

of the hazard rate and the random process $\mathbb{I}\{T_i > t\}$ indicating whether i is at risk just before time t. The interpretation of the intensity process is that $\lambda_i(t)dt = \mathbb{E}(dN_i(t)|\mathcal{F}_i(t-))$ the conditional expectation of the jump size $dN_i(t)$ at time t given the observed history $\mathcal{F}_i(t-)$ of individual i in $(0, t)$.

In this paper, I extend the single spell duration analysis to a dynamic multiple spell duration analysis using the counting process approach. I contribute to the literature in the following areas:

Allowing sample data to be doubly stochastic. In practice, researchers often encounter duration data that is collected within a fixed time interval, e,g., duration data of purchases within a calendar year. For a fixed observation window $\mathcal{T} = (0, \tau]$, a set of durations is denoted as $\{T_{ij}\}$ where for every individual $i = 1, \ldots, n$, I let $j = 1, \ldots, n_i$ to denote the cardinal of each duration. Crucially, the numbers of duration $\{n_i\}$ for each individual are i.i.d realizations of a random variable N, and within each individual, the length of a duration (or a spell) T_{ij} follows a distribution F_{ij}. For this reason, I call such data doubly stochastic. A close but essentially different data structure is the duration panel data, where within the observation window $\mathcal{T}$, the numbers of duration for different individuals are often identical. In the panel data case, the analysis is performed in a multiple spell duration model, where researchers construct economic models by specifying a multivariate joint distribution, e.g., Heckman and Walker (1990); Honoré (1993). However, in the context of the data $\{T_{ij}\}$, this framework implies a sample selection problem and loss of information: One has to fix an integer n^*, individuals with $n_i < n^*$ would be ignored from the estimation, while individuals with $n_i > n^*$ would not have their information fully utilized by researchers. Further notice that, although the data structure presented here is similar to that of an unbalanced panel data, the source of this unbalanced phenomenon is different. None of the conventional reasons, like rotating, randomly missing data, pooling cross-sectional and time-series data, nonresponsive, censoring or selection bias [1]is the source for the randomness of N_i. Rather, it is the underlying data generating process that makes N_i vary across individuals.

Allowing durations to be state dependent. If past experiences of an individual have genuine effects on his/her future behaviors, such structural relationship is called the

1. see Baltagi and Song (2006); Hsiao (2014) for references

state dependence. State dependence has long been documented in economic literature. For example, in consumer data, researchers in both marketing and economics have observed a form of persistence in choice data whereby consumers have a higher probability of choosing products that they have purchased before. In labor economics, there are experience-rating mechanisms that produce state dependent data. For example, a company might introduce an experience-oriented absence regulation that links a worker's benefit to his or her absence score calculated over a period. In automobile insurance, an individual's insurance premium is calculated based on the number of claims occurred in the previous year. Most literature assumes that, conditional on observed and unobserved individual heterogeneities, durations are independent.

Allowing a fixed effect unobserved heterogeneity. In a fixed effect framework, I allow an individual's unobserved heterogeneity to be correlated with other explanatory variables. This setting not only makes a model more flexible, but also is essentially necessary when durations are state dependent, since some covariates are constructed using past durations. Since Heckman (1978, 1981), distinguishing the impacts of unobserved heterogeneity from those of state dependence has been a central issue in empirical work. This distinction has implications for the interpretation and policy implications of many observed phenomena. For example, studies on unemployment durations seek to identify the duration dependence in job finding from unobserved heterogeneity, e.g., Kroft et al. (2013) and the literature cited therein. In the single spell duration literature, key results are that if the unobserved heterogeneity ν satisfies a tail condition $\mathbb{E}\nu < \infty$, then the model can be identified (Elbers and Ridder (1982); Heckman and Singer (1984)) and are that if no assumption on the tail distribution is made, then for a duration model with $\mathbb{E}\nu < \infty$, there are observationally equivalent models with $\mathbb{E}\nu = \infty$, see Ridder (1990). Multiple spell duration models require weaker identification assumptions, Honoré (1993) demonstrates one does not need to impose tail distribution restrictions to identify the model if multiple spell duration data are available. Different from existing literature, I eliminate the individual fixed effect through a first difference transformation.

Constructing a dynamic model for work absence. Work absences are not uncommon and are costly for both firms and employers. Yet, compare to the unemployment studies (both work absences and unemployments are interpretations of work flow), economists pay unproportionately less attention to this issue. In management and psychological literature, researchers often choose simple models to study the influence of some covariates. Using the proposed framework, I build a dynamic model to analysis how past absence records affect future absence decisions. These decisions include the decision to ask for a leave, and the decision of the length of an absence.

I adopt a counting process approach to analyze the duration data $\{T_{ij}\}$. For an individual i, let $S_{ij} = \sum_{k=1}^{j} T_{ik}$, a counting process for this individual:

$$N_i(s) = \sum_{j=1}^{\infty} \mathbb{I}\{S_{ij} \leq s\}$$

counts the number of S_{ij} that fall below s. To avoid confusion, I will use the notation $T(t)$ to denote the duration and $S(s)$ to denote the constructed event time throughout this paper. There are two reasons for choosing this approach. First, using a counting process in a time interval $\mathcal{T}$, one can completely describe the probability structure of a stochastic process $\{T_{ij}\}_{j\geq 1}$. Specifically, the randomness of durations $\{T_{ij}\}$ (or equivalently, $\{S_{ij}\}$) as well as the randomness of counts N_i can be written concisely. State dependence structures among durations can also be specified easily in an intensity function of a counting process. Second, one can use the martingale relationship between the counting process and its cumulative intensity function to construct moment restrictions. These moment restrictions provide an estimation channel. To give an impression, considering again the i.i.d durations $\{T_i\}_{i=1,\dots,n}$ represented as single-event counting processes $\{N_i(t) = \mathbb{I}\{T_i \leq t\}\}_{i=1,\dots,n}$. Define the cumulative intensity function as

$$\Lambda_i(t|\mathcal{F}_i(t-)) = \int_0^t \frac{f(z)}{1-F(z)} \mathbb{I}\{T_i > z\} dz$$

A continuum of moment restrictions is

$$\mathbb{E}\left(N_1(t) - \Lambda_1(t|\mathcal{F}_1(t-))\right) = \mathbb{E}M_1(t) = 0, \forall t \in \mathcal{T}$$

The paper is organized as the following. Section 2 introduces the model. For an individual i, begin with specifications for durations $\{T_{ij}\}_{j\geq 1}$, I show a cumulative intensity function $\Lambda_i(s|\mathcal{F}_i(s-))$ of a counting process $N_i(s)$ can be constructed from the specified hazard rates. Different count probabilities are also given in this section. In section 3, I introduce a first difference transformation to 'cancel' the individual fixed effect. In section 4, I advocate a minimum distance estimation method. This method is based on the martingale relationship between a counting process and its cumulative intensity. I provide simple proofs on the estimator's consistency and asymptotic normality results. Section 5 studies a work absence application. I show that, in general, workers would react to his/her absence score, with the exception when the absence reason is health related. Section 6 discusses some related topics and Section 7 concludes the whole paper.

2. Multiple Spell Duration Model in A Counting Process Framework

The presentation of the framework consists of three parts. First, I will specify the duration T_{ij}. Second, I will characterize $N_i(s)$ on a fixed time interval $\mathcal{T} = (0, \tau]$ using the specified hazard rates. Lastly, count probabilities are presented in an evolutionary way.

2.1 A Generalized Accelerated Failure Time Model for Duration

I specify a duration random variable T_{ij} as:

$$L(T_{ij};\alpha_0) = G(x_{ij};\beta_0)\nu_i u_{ij} \tag{1}$$

or

$$\begin{aligned}\log L(T_{ij};\alpha_0) &= \log G(x_{ij};\beta_0) + \log \nu_i + \log u_{ij} \\ &= \log G(x_{ij};\beta_0) + \eta_i + \varepsilon_{ij}\end{aligned}$$

where

- $L : [0,+\infty) \to \mathbb{R}_+$ is a known monotone function up to parameters with $L(0) = 0, L(+\infty) = +\infty$. This function measures current duration dependence.
- $x_{ij} \in X \subset \mathbb{R}^q$ is a q-vector of time-dependent explanatory variables that might contain state dependent elements.
- $G : \mathbb{R}^d \to \mathbb{R}_+$ takes on at least two distinct values on X.
- α_0 and β_0 represent vectors of true parameters. ν_i is an unobserved heterogeneity.
- The error terms $\{u_{ij}\}$ are assumed to be i.i.d over individuals and durations, and are independent of both x_{ij} and ν_i.

The above specification is the Generalized Accelerate Failure Time (GAFT) model, introduced by Ridder (1990). This model is non-parametrically identified up to a normalization.

In a GAFT model, one does not specify its hazard rate, but transform the duration. The transformed duration is a dependent variable of a model with multiplicative (or additive if the logarithm is performed) disturbance. As Ridder (1990) summarized, 'a GAFT model is characterized by the transformation of the dependent variable, the specification of the regression function, and the choice of the error distribution.' Durations that are modelled by (non)linear models have a strong attraction to econometricians. In addition, classical duration models such as the Proportional Hazards (PH) and Mixed Proportional Hazards (MPH) are contained in this specification.

From a GAFT specification, one can easily derive the hazard rate of T_{ij}:

$$\begin{aligned}P(T_{ij} \le t) &= P(L(T_{ij};\alpha_0) \le L(t;\alpha_0)) \\ &= P(u_{ij} \le L(t;\alpha_0)\,(G(x_{ij};\beta_0)\nu_i)^{-1}) \\ &= F_u\left(L(t;\alpha_0)\,(G(x_{ij};\beta_0)\nu_i)^{-1}\right)\end{aligned}$$

Recall the survival function $\bar{F}_{ij}(t) = P(T_{ij} > t) = 1 - F_{ij}(t)$ and the integrated hazard rate $H_{ij}(t)$ are linked via

$$\bar{F}_{ij}(t) = \exp(-H_{ij}(t))$$

Hence, the integrated hazard rate is

$$H_{ij}(t) = -\log\left(1 - F_u\left(L(t;\alpha_0)\left(G(x_{ij};\beta_0)\nu_i\right)^{-1}\right)\right)$$

Here, for simplicity, I assume $L(t;\alpha_0)$ is a non-decreasing function, and F_u is the probability distribution of u. If $L(t) = \Lambda_0(t)$ (the integrated base hazard), $G(x_{ij};\beta_0) = \exp(-x_{ij}^\top\beta_0)$, $\nu_i = \exp(-\eta_i)$, and $u_{ij} \sim EXP(1)$, one has an MPH model:

$$h_{ij}(t) = \lambda_0(t)\exp(x_{ij}^\top\beta + \eta_i)$$

This GAFT specification has three features. First, the composite error term $\nu_i u_{ij}$ enters the model multiplicatively. Taking the logarithm, this structure restriction is equivalent to the additive separability assumption made in most dynamic models, see Aguirregabiria and Mira (2010), and provides a first-difference channel to swipe out the unobserved heterogeneity ν_i. Second, ν_i enters into the model as an individual fixed effect, i.e., I do not restrict the conditional distribution of ν. The third feature of the model is that I put minimum restrictions on the state dependent structure. The time dependent explanatory variable $\{x_{ij}\}_{j\geq 1}$ is included in a $\sigma-$algebra generated by all the history information:

$$x_{ij} \in \sigma(\{t_{is}\} : s < j) \tag{2}$$

Special cases are the $q-$th order lagged duration: $G(x_{ij};\beta_0) = \sum_{k=1}^{q} t_{i(j-k)}\beta_k$, and the permanent memory effect: $G(x_{ij};\beta_0) = \beta_0\sum_{k=1}^{j-1} t_{ik}$.

2.2 Characterizing a Counting Process

As described in the introduction section, one can construct a counting process $N_i(s)$ for each individual. In this subsection, I discuss how to use the duration specification to characterize $N_i(s)$.

First, note that a counting process $N_i(s)$ can be viewed as an aggregation of single-event counting processes:

$$N_i(s) = \sum_{j=1}^{\infty}\mathbb{I}\{S_{ij} \leq s\} = \sum_{j=1}^{\infty} N_{ij}(s)$$

where $N_{ij}(s) = \mathbb{I}\{S_{ij} \leq s\} = \mathbb{I}\{T_{ij} \leq s - s_{i(j-1)}\}$, where $\{s_{ij}\}_j$ are realizations of $\{S_{ij}\}_j$. Let $h_{ij}(t) = dH_{ij}(t)/dt$ be the corresponding hazard rate, one might write down the intensity function for $N_{ij}(s)$ as:

$$\lambda_{ij}(s|\mathcal{F}_{ij}(s-)) = h_{ij}(s - s_{i(j-1)})\mathbb{I}\{S_{ij} > s > s_{i(j-1)}\}$$

The intensity function is identical to the hazard rate if an 'event' is still at risk, but the value of the intensity would vanish to zero if an 'event' has occurred. By the construction of $N_i(s)$, it is reasonable to say that its intensity function $\lambda_i(s|\mathcal{F}_i(s-))$ should be:

$$\lambda_i(s|\mathcal{F}_i(s-)) = \sum_{j=1}^{\infty} h_{ij}(s - s_{i(j-1)})\mathbb{I}\{S_{ij} > s > s_{i(j-1)}\}$$

The following theorem states that this is indeed the correct specification.

Theorem 1 *Given a filtration $\mathcal{F}_i(s-)$ such that $\sigma(N_i(z) : z < s) \subseteq \mathcal{F}_i(s-)$, the cumulative intensity function $\Lambda_i(s|\mathcal{F}_i(s-)) = \int_0^s \lambda_i(z|\mathcal{F}_i(z-))dz$ of $N_i(s)$ is given by:*

$$
\begin{aligned}
\Lambda_i(s|\mathcal{F}_i(s-)) =& \Lambda_i(s_{i(j-1)}|\mathcal{F}_i(s_{i(j-1)}-)) \\
&+ \int_{s_{i(j-1)}}^{s} h_{ij}(z - s_{i(j-1)})\mathbb{I}\{S_{ij} > z > s_{i(j-1)}\}dz
\end{aligned} \tag{3}
$$

where j here is the largest number such that $s_{i(j-1)} < s$.

Proof See Appendix B. ■

Theorem 1 can be used to present the well-known Doob-Meyer decomposition result. Let $M_i(s) = N_i(s) - \Lambda_i(s|\mathcal{F}_i(s-))$, the Doob-Meyer decomposition states that $M_i(s)$ is a martingale:

$$
\begin{aligned}
\mathbb{E}\left(M_i(s)|\mathcal{F}_i(u)\right) &= \mathbb{E}\left(M_i(u) + M_i(s) - M_i(u)|\mathcal{F}_i(u)\right) \\
&= M_i(u) + \mathbb{E}\left(\int_u^s dN_i(z) - d\Lambda_i(z|\mathcal{F}_i(z-))|\mathcal{F}_i(u)\right) \\
&= M_i(u)
\end{aligned}
$$

where $u < s$, and

$$
\mathbb{E}M_1(s) = \mathbb{E}\left(N_1(s) - \Lambda_1(s|\mathcal{F}_1(s-))\right) = 0, \quad \forall s \in \mathcal{T} \tag{4}
$$

This continuum of moment restrictions gives us a channel to estimate the model. I will discuss the estimation method in the next section.

2.3 Count Probabilities

A key property of the data set $\{T_{ij}\}$ is that the number of durations for each individual $N_i(\tau)$ is not fixed. In this subsection, I describe the probability of $N_i(s) = m$.

Consider two modified intensity functions:

$$
\lambda_{N_i(s)}(s, \mathcal{F}_i(u)) = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr\{N_i(s+\Delta) - N_i(s) = 1 | N_i(s), \mathcal{F}_i(u)\}
$$

where $u < s$, and

$$
\tilde{\lambda}_{N_i(s)}(s) = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr\{N_i(s+\Delta) - N_i(s) = 1 | N_i(s)\}
$$

$\lambda_{N_i(s)}(s, \mathcal{F}_i(u))$ is an intensity that is conditional on history $\mathcal{F}_i(u)$ as well as the present state $N_i(s)$. $\tilde{\lambda}_{N_i(s)}(s)$ is an intensity function that is only conditional on the current state, while positions of past events do not matter.

The following lemmas from Rubin (1972) describe the probability of $N_i(s) = m$.

Lemma 2 *For a counting process, $\forall s > u \geq 0$, $m = 0, 1, 2, \ldots$, one has*

$$\frac{\partial}{\partial s}\Pr\{N_i(s) = m|\mathcal{F}_i(u)\} = \lambda_{m-1}(s, \mathcal{F}_i(u))\Pr\{N_i(s) = m-1|\mathcal{F}_i(u)\} - \lambda_m(s, \mathcal{F}_i(u))\Pr\{N_i(s) = m|\mathcal{F}_i(u)\}$$

Lemma 3 *Let $p_m(s) = \Pr\{N_i(s) = m\}$ is given by*

$$\frac{\partial}{\partial s}p_0(s) = -\tilde{\lambda}_0(s)p_0(s)$$

$$\frac{\partial}{\partial s}p_m(s) = \tilde{\lambda}_{m-1}(s)p_{m-1}(s) - \tilde{\lambda}_m p_m(s), \quad m \geq 1$$

Lemma 1 states that an evolution of a conditional probability satisfies a difference-differential equation, while Lemma 2 shows that a similar equation also governs the evolution of an unconditional probability. Brief proofs of these lemmas can be found in Appendix B.

3. A First Difference Transformation

I introduce a first difference transformation to cancel ν_i in this section. Some illustration examples are also presented here.

3.1 Method

The composite error term $\nu_i u_{ij}$ is additive after one takes the logarithm of the GAFT model. This structure provides a way to perform the first difference operation to cancel ν_i:

$$\Delta \log L(T_{ij}; \alpha_0) = \log\left(\frac{G(x_{ij}; \beta_0)}{G(x_{i(j-1)}; \beta_0)}\right) + \Delta\varepsilon_{ij}, \quad j \geq 2$$

where $\Delta y_{ij} = y_{ij} - y_{i(j-1)}$. In addtion, suppose that

- $\Delta \log L(T_{ij}; \alpha_0) = \alpha_0 \Delta\tilde{L}(T_{ij})$
- $\tilde{L}(T_{ij}) : \mathbb{R}_+ \to \mathbb{R}_+$ is a known function with no unknown parameters.

Then the first difference result could be rewritten as:

$$\Delta\tilde{L}(T_{ij}) = \log\left(\frac{G(x_{ij}; \beta_0)}{G(x_{i(j-1)}; \beta_0)}\right)^{1/\alpha_0} + \log\left(\frac{u_{ij}}{u_{i(j-1)}}\right)^{1/\alpha_0}$$

One can calculate a series of new 'durations' $\{\exp(\Delta\tilde{L}(T_{ij}))\}_{j\geq 2}$ for each individual i. The distribution function of $\exp(\Delta\tilde{L}(T_{ij}))$ can be easily derived:

$$\begin{aligned}
F_{\exp(\Delta\tilde{L}(T_{ij}))}(t) &= \Pr\{\exp(\Delta\tilde{L}(T_{ij})) < t\} \\
&= F_{\Delta\varepsilon_{ij}}\left(t^{\alpha_0} + \log\left(\frac{G(x_{i(j-1)}; \beta_0)}{G(x_{ij}; \beta_0)}\right)\right) \\
&= F_{\tilde{u}_{ij}}\left(t^{\alpha_0}\frac{G(x_{i(j-1)}; \beta_0)}{G(x_{ij}; \beta_0)}\right)
\end{aligned}$$

where $\tilde{u}_{ij} = u_{ij}/u_{i(j-1)}$, F_y is the distribution function of a random variable y. Further notice that

$$\begin{aligned}
F_{\varepsilon_2-\varepsilon_1}(a) &= \iint_{\varepsilon_2-\varepsilon_1\le a} f_\varepsilon(\varepsilon_1)f_\varepsilon(\varepsilon_2)d\varepsilon_1 d\varepsilon_2 \\
&= \int_{\varepsilon_1=-\infty}^{\infty}\int_{-\infty}^{a+\varepsilon_1} f_\varepsilon(\varepsilon_2)d\varepsilon_2 f_\varepsilon(\varepsilon_1)d\varepsilon_1 \\
&= \int_{-\infty}^{\infty} F_\varepsilon(a+\varepsilon_1)dF_\varepsilon(\varepsilon_1)
\end{aligned}$$

and

$$F_{\tilde{u}}(u) = P(\varepsilon_2-\varepsilon_1 \le \log(u)) = \int_{-\infty}^{\infty} F_\varepsilon(\log(u)+\varepsilon)dF_\varepsilon(\varepsilon)$$

Once one specifies a distribution for ε (or equivalently, u), the distribution of $\Delta\tilde{L}(T_{ij})$ follows immediately.

In this study, I focus on one popular choice of ε_{ij}: the Type I Extreme Value distribution, or equivalent, u_{ij} is the unit rate Exponential Distribution (Lee, 2008). Other useful specifications for ε (or u), see Ridder (1990) for reference. Thus,

$$\begin{aligned}
F_{\tilde{u}}(u) &= \int_{-\infty}^{\infty} [1-\exp(-\exp(\log u+\varepsilon))]\exp(\varepsilon-\exp(\varepsilon))d\varepsilon \\
&= \frac{u}{1+u}
\end{aligned}$$

and

$$F_{\exp(\Delta\tilde{L}(T_{ij}))}(t) = \frac{1}{1+t^{-\alpha_0}\frac{G(x_{ij};\beta_0)}{G(x_{i(j-1)};\beta_0)}}$$

The corresponding integrated hazard rate $\tilde{H}_{ij}$ for $\exp(\Delta\tilde{L}(T_{ij}))$ reads:

$$\tilde{H}_{ij}(t) = \log\left(1+t^{\alpha_0}\frac{G(x_{i(j-1)};\beta_0)}{G(x_{ij};\beta_0)}\right)$$

Like before, one could represent the data set $\{\exp(\Delta\tilde{L}(T_{ij}))\}_{i,j}$ using counting processes. For each individual i, let $\tilde{S}_{ij} = \sum_{k=2}^{j+1}\exp(\Delta\tilde{L}(T_{ik})), j \ge 1$, and one may construct a counting process:

$$\tilde{N}_i(s) = \sum_{j=1}^{\infty}\mathbb{I}\{\tilde{S}_{ij} \le s\} \tag{5}$$

Using Theorem 1, the corresponding cumulative intensity function is:

$$\begin{aligned}
\tilde{\Lambda}_i(s|\mathcal{F}_i(s)) =& \tilde{\Lambda}_i(\tilde{s}_{i(j-1)}|\mathcal{F}_i(\tilde{s}_{i(j-1)}-)) \\
&+ \int_{\tilde{s}_{i(j-1)}}^{s}\tilde{h}_{ij}(z-\tilde{s}_{i(j-1)})\mathbb{I}\{\tilde{S}_{ij} > z > \tilde{s}_{i(j-1)}\}dz
\end{aligned} \tag{6}$$

In the original counting process, the observation window is fixed: $\mathcal{T} = (0, \tau]$, but this is not the case for the transformed counting process $\tilde{N}_i$. This new observation window is $\mathcal{E}_i = (0, e_i]$, where

$$e_i = \sum_{j=2}^{n_i} \exp(\Delta \tilde{L}(t_{ij})) + \exp\left(\tilde{L}(\tau - s_{i(n_i)}) - \tilde{L}(t_{i(n_i)})\right)$$

Remark. A few words on the error term ε_{ij} are in order. So far, I have assumed they are i.i.d such that the new 'first-differenced error term' $\Delta\varepsilon_{ij}$ is well-defined, and its distribution can be derived easily. However, the error terms could also be correlated:

$$\varepsilon_{ij} = \rho + \varepsilon_{i(j-1)} + v_{ij}$$

or equivalently for u_{ij}:

$$u_{ij} = \exp(\rho + v_{ij}) u_{i(j-1)}$$

where $\{v_{ij}\}$ are i.i.d copies of a random variable v. When $\rho = 0$, ε_{ij} follows a random walk process.

3.2 Examples

Example 1. Specifying a duration T_{ij} as:

$$T_{ij}^{1+\alpha_0}/(1+\alpha_0) = \exp(x_{ij}\beta_0 + \eta_i) u_{ij}$$

or

$$(1+\alpha_0)\log(T_{ij}) - \log(1+\alpha_0) = x_{ij}\beta_0 + \eta_i + \varepsilon_{ij}$$

where x_{ij} is a state dependent variable. $x_{ij} = t_{i(j-1)}$ corresponds to the lagged duration model studied in Heckman and Borjas (1980); Heckman et al. (1985); Honoré (1993), while $x_{ij} = \sum_{k=1}^{j-1} t_{ik}$ corresponds to a permanent memory effect. If one specifies $u_{ij} \sim EXP(1)$, this GAFT model has a mixed proportional hazard rate:

$$h_{ij}(t) = t^{\alpha_0} \exp(-x_{ij}\beta_0 - \eta_i)$$

The first difference transformation swipe out η_i:

$$(1+\alpha_0)\Delta \log(T_{ij}) = \Delta x_{ij}\beta_0 + \Delta\varepsilon_{ij}$$

The integrated hazard rate for $\exp(\Delta \log(T_{ij}))$ is:

$$\tilde{H}_{ij}(t) = \log\left(1 + t^{(1+\alpha_0)} \exp((-\Delta x_{ij}\beta_0)\right), j \geq 2$$

Let $\tilde{S}_{ij} = \sum_{k=2}^{j+1} \exp(\Delta \log(T_{ik})), j \geq 1$, and construct a new counting process:

$$\tilde{N}_i(s) = \sum_{j=1}^{\infty} \mathbb{I}\{\tilde{S}_{ij} \leq s\}$$

its corresponding cumulative intensity is:

$$\begin{aligned}\tilde{\Lambda}_i(s|\mathcal{F}_i(s)) =& \tilde{\Lambda}_i(\tilde{s}_{i(j-1)}|\mathcal{F}_i(\tilde{s}_{i(j-1)}-)) \\ &+ \log\left(1 + (s - \tilde{s}_{i(j-1)})^{(1+\alpha_0)}\exp(-\Delta x_{ij}\beta_0)\right)\mathbb{I}\{\tilde{S}_{ij} > s > \tilde{s}_{i(j-1)}\}\end{aligned}$$

Example 2. Another interesting example is when $L(T_{ij}) = \exp(T_{ij})$, $G(x_{ij};\beta_0) = \exp(x_{ij}\beta_0)$, $x_{ij} = t_{i(j-1)}$, and $u_{ij} \sim EXP(1)$:

$$\exp(T_{ij}) = \exp(t_{i(j-1)}\beta_0)\nu_i u_{ij}$$

or

$$T_{ij} = t_{i(j-1)}\beta_0 + \log\nu_i + \varepsilon_{ij}$$

This is the classical AR(1) dynamic model. The integrated hazard rate for $\exp(\Delta T_{ij}), j \geq 2$ is:

$$\tilde{H}_{ij}(t) = \log\left(1 + t\exp(-\Delta t_{i(j-1)}\beta_0)\right)$$

with an initial value t_{i0} given, one can accordingly construct a new counting process using $\{\exp(\Delta t_{ij})\}_{j\geq 2}$ and find its cumulative intensity function.

Provided that the data is balanced and $|\beta_0| < 1$, conventional GMM based estimation methods (e.g., the Anderson-Hsiao Estimator, Arellano-Bond Estimator and system GMM) would use lags and levels as internal instrument variables. When $|\beta_0|$ is close to unit, these methods would encounter the weak IV problem (Blundell and Bond, 1998). When $|\beta_0| = 1$, the IV relevance condition is no longer valid. Moreover, in a relatively long period, the IV proliferation problem would be severe. The estimation method proposed in this study is not based on instrument variables and is free of the above-mentioned issues. The disadvantage of this counting process approach, however, is a full characterization of the model distribution, hence, is more sensitive to misspecification.

4. A Minimum Distance Estimation Method

In this study, I advocate a minimum distance method to estimate model parameters. Likelihood methods based on a joint distribution specification would create a sample selection bias. I discuss this issue in detail in Section 6.

4.1 Estimation Theories

The continuum of moment restrictions in Equation 4 provides a channel for estimation. Carrasco and Florens (2000) developed a GMM method for a continuum of moment conditions. However, to estimate the inverse of a non-invertible covariance operator, one has to perform a Tikhonov regularization. This method is complicated and requires additional knowledge on a tuning parameter. Kopperschmidt and Stute (2013) propose a minimum distance method to estimate self-exciting processes under multiple observations.

Their asymptotic properties results are based on U-statistic arguments. In this section, I re-introduce their estimator with simplified proofs.

Denote $\theta \in \Theta \subset \mathbb{R}^q$ as parameters of interest, $\mathcal{T}$ as the time space, $\mathcal{M} = \{\Lambda(s;\theta) : s \in \mathcal{T}, \theta \in \Theta\}$ as the associated model for the cumulative intensity, and let

$$M(s;\theta) = \mathbb{E}\left(N_1(s) - \Lambda_1(s;\theta|\mathcal{F}_1(s-))\right)$$

be the moment restriction. By Doob-Meyer decomposition result, one has

$$M(s;\theta_0) = 0 \quad \forall s \in \mathcal{T}$$

where θ_0 is a vector of true parameters. I impose the following assumptions:

- A1. For each $\varepsilon > 0$.

  $$\inf_{||\theta-\theta_0||\geq\varepsilon} ||\mathbb{E}\Lambda(\cdot,\theta_0) - \mathbb{E}\Lambda(\cdot,\theta)||_{\mathbb{E}\Lambda(\cdot,\theta_0)} > 0$$

  where

  $$||f||_\mu = \left[\int_{\mathcal{T}} f^2(z)\mu(dz)\right]^{1/2}$$

  is a semi-norm.

- A2. The process $(s;\theta) \to \Lambda(s;\theta)$ is continuous with probability one.

- A3. $\Lambda(s;\theta)$ is bounded in s and θ.

- A4. $\Theta \subset \mathbb{R}^k$ is compact.

Assumption A1 is a weak identification condition. A2 suggests that $\Lambda(s,\theta)$ has a (random) Lebesgue intensity $\lambda(s,\theta)$ with values in an appropriate Skorokhod space. This guarantees continuity (but not differentiability) of $\Lambda(s,\theta)$ in s and allows for unexpected jumps in the intensity function. A3 is used in ÖZTÜRK and Hettmansperger (1997), and A4 is standard in the literature.

By Assumption A1, one has

$$P\left(M(s;\theta) = 0\right) < 1, \quad \theta \neq \theta_0$$

thus, $M(s;\theta) \neq 0$ in a non-null space of $\mathcal{T}$, and one has

$$\int_{\mathcal{T}} M(z;\theta_0)^2 \mathbb{E}\Lambda(dz;\theta_0) = 0$$

but

$$\int_{\mathcal{T}} M(z;\theta)^2 \mathbb{E}\Lambda(dz;\theta_0) \neq 0 \quad \forall \theta \neq \theta_0$$

Hence,

$$\theta_0 = \arg\min_{\theta\in\Theta} \int_{\mathcal{T}} M(z;\theta)^2 \mathbb{E}\Lambda(dz;\theta_0)$$

By Lemma 3 of Kopperschmidt and Stute (2013), the above equation can be re-written as:

$$\theta_0 = \arg\min_{\theta\in\Theta} ||\mathbb{E}\Lambda(\cdot;\theta_0) - \mathbb{E}\Lambda(\cdot;\theta)||^2_{\mathbb{E}\Lambda(\cdot;\theta_0)}$$

By Lemma 5 of the same paper, one has

$$||\bar{N}_n - \bar{\Lambda}_n(\cdot;\theta)||^2_{\bar{N}_n} \xrightarrow{p} ||\mathbb{E}\Lambda(\cdot;\theta_0) - \mathbb{E}\Lambda(\cdot;\theta)||^2_{\mathbb{E}\Lambda(\cdot;\theta_0)}$$

where

$$\bar{N}_n = \frac{1}{n}\sum_{i=1}^n N_i, \quad \bar{\Lambda}_n(\cdot;\theta) = \frac{1}{n}\sum_{i=1}^n \Lambda_i(\cdot;\theta)$$

One can write the minimum distance estimator as

$$\begin{aligned}\hat{\theta}_n &= \arg\min_{\theta\in\Theta} ||\bar{N}_n - \bar{\Lambda}_n(\cdot;\theta)||^2_{\bar{N}_n} \\ &= \arg\min_{\theta\in\Theta} \int_{\mathcal{T}} \bar{M}_n(z;\theta)^2 \bar{N}_n(dz)\end{aligned}$$

where $\bar{M}_n(s;\theta) = \bar{N}_n(s) - \bar{\Lambda}_n(s;\theta)$, and $M_i(s;\theta) = N_i(s) - \Lambda_i(s;\theta)$. The quantity $||\bar{N}_n - \bar{\Lambda}_n(\cdot;\theta)||^2_{\bar{N}_n}$ represents an overall measure of fit of $\bar{\Lambda}_n(\cdot;\theta)$ to $\bar{N}_n$. This objective function is a weighted Cramér-von Mises statistic, which can be interpreted as a minimum distance estimator. The use of a counting measure is particularly convenient, as it turns a continuous integration into a discrete finite-many summation.

Theorem 4 *Under Assumptions A1-A4, one has*

$$\hat{\theta}_n \xrightarrow{a.s} \theta_0$$

Proof See Appendix B ■

In order to obtain the asymptotic normality, additional assumptions are required.

- A5. $\Lambda(s;\cdot)$ is once differentiable in a neighborhood of θ_0 and satisfies $\dot{\Lambda}(s;\theta)$ is square integrable w.r.t $\mathbb{E}\Lambda(\cdot;\theta_0)$ where $\mathcal{N}_0$ is a neighborhood of θ_0 and $\dot{\Lambda}(s;\theta) = \partial\Lambda(s;\theta)/\partial\theta$.
- A6. $\theta_0 \in int(\Theta)$.

Assumption A5 is a standard smoothness condition. A5 is unchanged if one replaces $\Lambda(s;\cdot)$ by $M(s;\cdot)$, and $\dot{\Lambda}(s;\theta)$ by $\dot{M}(s;\theta) = \mathbb{E}\dot{M}_1(s;\theta) = \partial M(s;\theta)/\partial\theta$. Assumption A6 is standard.

Theorem 5 *Under Assumptions A1-A6, one has*

$$\sqrt{n}\left(\int_{\mathcal{T}} \dot{M}(z;\theta_0)\dot{M}(z;\theta_0)^\top \mathbb{E}\Lambda(dz;\theta_0)\right)(\hat{\theta}_n - \theta_0) \xrightarrow{d} \int_{\mathcal{T}} \dot{M}(z;\theta_0) B_\Gamma \mathbb{E}\Lambda(dz;\theta_0)$$

where B_Γ denotes a centered Gaussian process with covariance structure given by $\Gamma(s_1,s_2) = \mathbb{E}(M_1(s_1;\theta_0)M_1(s_2;\theta_0))$.

Proof See Appendix B. ■

This theorem naturally leads to the following corollary.

Corollary 6 *Under Assumptions A1-A6, one has*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Omega)$$

where

$$\begin{aligned}\Omega = &\left(\int_{\mathcal{T}} \dot{M}(z;\theta_0)\dot{M}(z;\theta_0)^\top \mathbb{E}\Lambda(dz;\theta_0)\right)^{-1} \times \\ &\int_{\mathcal{T}}\int_{\mathcal{T}} \dot{M}(s_1;\theta_0)\dot{M}(s_2;\theta_0)^\top \Gamma(s_1,s_2)\mathbb{E}\Lambda(ds_1;\theta_0)\mathbb{E}\Lambda(ds_2;\theta_0)\times \\ &\left(\int_{\mathcal{T}} \dot{M}(z;\theta_0)\dot{M}(z;\theta_0)^\top \mathbb{E}\Lambda(dz;\theta_0)\right)^{-1}\end{aligned}$$

Proof This result follows immediately from Theorem 3 and the fact that the integrated weighted Gaussian process follows a normal distribution. ■

Remark. A transformation of the expression for Ω might simplify its estimation. Notice that

$$\int_{\mathcal{T}} \dot{M}(z;\theta_0)B_\Gamma \mathbb{E}\Lambda(dz;\theta_0) = \sqrt{n}\int_{\mathcal{T}} \bar{M}_n(z;\theta_0)\dot{M}(z;\theta_0)\mathbb{E}\Lambda(dz;\theta_0) + o_p(1)$$

and

$$\begin{aligned}\sqrt{n}\int_{\mathcal{T}} \bar{M}_n(z;\theta_0)\dot{M}(z;\theta_0)\mathbb{E}\Lambda(dz;\theta_0) &= \sqrt{n}\int_{\mathcal{T}} \bar{M}_n(z;\theta_0)\mathbb{E}\frac{\partial}{\partial\theta}\Lambda(z;\theta)\mathbb{E}\Lambda(dz;\theta_0)\mid_{\theta=\theta_0} \\ &= \sqrt{n}\int_{\mathcal{T}}\int_{s^*}^{T} \mathbb{E}\frac{\partial}{\partial\theta}\Lambda(z;\theta)\mathbb{E}\Lambda(dz;\theta_0)\bar{M}_n(ds^*;\theta_0)\mid_{\theta=\theta_0} \\ &= \sqrt{n}\int_{\mathcal{T}} \psi(s^*)\bar{M}_n(ds^*;\theta_0)\end{aligned}$$

where

$$\psi(s^*) = \int_{s^*}^{T} \mathbb{E}\frac{\partial}{\partial\theta}\Lambda(z;\theta)\mathbb{E}\Lambda(dz;\theta_0)$$

Here, the second equation is the outcome of applying Fubini's Theorem. Thus,

$$\sqrt{n}\left(\int_{\mathcal{T}} \dot{M}(z;\theta_0)\dot{M}(z;\theta_0)^\top \mathbb{E}\Lambda(dz;\theta_0)\right)(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, C)$$

where C is a $k \times k$ matrix with entries

$$C_{ij} = \int_{\mathcal{T}} \psi_i(z)\psi_j(z)\mathbb{E}\Lambda(dz;\theta_0)$$

Notice that $\psi(s^*)$ can be estimated by

$$\hat{\psi}(s^*) = \int_{s^*}^{T} \frac{\partial}{\partial \theta} \bar{\Lambda}_n(z;\theta) \bar{N}_n(dz) \mid_{\theta=\hat{\theta}} = \frac{1}{N_n} \sum_{l:s_l>s^*} \frac{\partial}{\partial \theta} \bar{\Lambda}_n(s_l;\theta) \mid_{\theta=\hat{\theta}}$$

where N_n and s_l are the number of events and event times of the average process $\bar{N}_n((0,\tau])$, respectively. Similarly, C_{ij} is estimated by

$$\hat{C}_{ij} = \int_{\mathcal{T}} \hat{\psi}_i(z)\hat{\psi}_j(z)\bar{N}_n(dz) = \frac{1}{N_n} \sum_{l=1}^{N_n} \hat{\psi}_i(s_l)\hat{\psi}_j(s_l)$$

The term $\left(\int_{\mathcal{T}} \dot{M}(z;\theta_0)\dot{M}(z;\theta_0)^\top \mathbb{E}\Lambda(dz;\theta_0)\right)$ can be estimated in the same way, and hence, its estimation expression is omitted here.

4.2 Simulation Studies

I use three data generating processes (DGPs) to study estimators' finite sample performance. The specifications are:

1. $t_{ij} = \exp\left(-\beta \sum_{k=1}^{j-1} t_{ik}\right) \nu_i u_{ij}$, with $\beta = -0.15$, and $\nu_i \sim U(1.25, 3.5)$
2. $t_{ij}^{\alpha} = \exp\left(-\beta \sum_{k=1}^{j-1} t_{ik}\right) \nu_i u_{ij}$, with $\alpha = 0.5, \beta = -0.05$, and $\nu_i \sim U(1.25, 3.0)$
3. $t_{ij}^{\alpha} = \exp\left(-\beta t_{i(j-1)}\right) \nu_i u_{ij}$, with $\alpha = 1.25, \beta = 1.65$, and $\nu_i \sim U(4.25, 9.0)$

I use the standard exponential error term $u_{ij} \sim EXP(1)$ in all DGPs. For each DGP, I run 200 replications. The sample size is set at $n = 100(200)$. Because of the intrinsic dynamic structure of each DGP, the terminating times are specified as

1. DGP1: $\tau = 3500(5000)$
2. DGP2: $\tau = 3500(5000)$
3. DGP3: $\tau = 10(15)$

The simulation algorithm is documented in Appendix C, the first difference transformation is used to cancel ν_i, and new counting processes are constructed for estimation. Tables 1-3 report the results. I also plot the estimated density function (kde estimation) of these estimators. Figures can be found in Appendix D.

Insert Tables 1 to 3

5. An Empirical illustration

I apply the proposed method to studying the dynamic mechanism of an individual's work absence decisions. This application concerns how the absence score would affect a worker's absence decisions. These decisions include the decision to apply for an absence, which is measured by the working duration and the decision to return to work after an absence, measured by the absence duration. The inclusion of absence score creates a state-dependent structure in the econometric models. The absence score is calculated as the accumulation of absence days and is linked to job benefits. One could reasonably assume that workers should have arranged their absence decisions dynamically. Another state dependent effect would be the duration effect: The longer one stays in a state (working or absence), the more (or less) likely one is to continue to stay in such a state.

Work absences are not uncommon in both developed and developing countries. U.S. Bureau of Labor Statistics (2005) data reveals that, on any given day, approximately 3.3% of the U.S. workforce does not report to work. Duflo et al. (2012) reports the absence rate in an Indian NGO teacher program could be as high as 35%. Moreover, work absences are costly for both workers and firms. For workers, although the social security covers illness-related absences in some countries in the form of sick pay, the replacement rates are, in general, less than 100%. For firms, arguably, labor costs are the single most considerable budgetary expense. Fister-Gale (2003) cites research showing that absenteeism costs in one survey population accounted for as high as 14.3% of the total payroll. Early works by Allen (1981) and Barmby et al. (1991) demonstrate the importance of financial aspects in explaining absence behavior. A group of Norwegian economists contribute significantly to this field. Markussen et al. (2011) shows that employee heterogeneity drives most cross-section variation in absenteeism. Fevang et al. (2014) show that Norway's social security system of short-term pay liability creates a sick pay trap: firms are discouraged from letting long-term sick workers back into work. Applied psychologists and management specialists contribute most to the work absenteeism literature. In general, psychological literature argues, according to Steers and Rhodes (1978), that job dissatisfaction represents the primary cause of absenteeism. In management literature, however, this view has been challenged. Increased understanding of the importance of so-called trigger absence behavior has emerged from the management literature (Steel et al., 2007). These literatures argue that absence score is a significant work absence decision-making factor. However, no satisfied empirical work has been done to support this claim.

5.1 The Data

The data in this application is obtained from a UK-based manufacturing firm who produced a homogeneous product. Other publications that use the same data (or a subset of the data) are Barmby et al. (1991) and Barmby et al. (1995). The data consists of detailed absence records: the beginning and ending dates of absences, types of absences (sick-leave, maternity release, jury service, work accident etc.) as well as individual characteristics. I use the data from the calendar year 1987 to 1988. In total, there are 749 workers with 5718 absence records. Figure 1 shows the histogram distribution of the length of absences. Among all the absences, one-day off leaves account for more than half. Absences longer than 5 days are rare.

Insert Figure 1

The firm has introduced an experience-oriented absence policy. In general, workers with lower cumulative absence scores will receive a better sick-leave benefit, but all employees are entitled to receive the government statutory sick-pay (SSP) benefit if they meet the criteria. Workers are categorized into three groups: Grade A workers are paid with their full wage (including bonuses) less the SSP, Grade B workers are paid with their basic wages less SSP, and Grade C workers receive no benefits from the firm. Each day of absence attracts a certain number of 'points', mostly 1 point, depending on the cause of this absence. To simplify the analysis, I assume that one day-off represents a 1 point of absence score. The firm's regulation states that Grade A workers have less than 21 points, Grade B workers have 21 to 41 points, and Grade C workers are those above 41 points. Workers are categorized into these three grades based on their absence records over the previous two years.

There is no abnormal behavior around the cut-off points 21 and 41. To show this, I non-parametrically estimate the absence score density function at the end of the years 1987 and 1988. Figure 2 plots the result. The P.D.Fs are smooth around these cut-off absence scores. Some possible explanations to this smoothness could be 1) It is difficult to foresee the occurrence of a future absence, 2) the absence regulation renews every two years, and last year's absence records (1988) that determine 1989's sick pay benefit are also the records to determine the benefit for 1990; hence the absence score is updated in a 'smooth' way, and 3) the absence score will affect only the sick-pay benefit (which is stochastic: only receive the benefit when ill), but not the salary (which is deterministic), hence the incentive to 'control' the absence scores around the cut-off points are not strong.

Insert Figure 2

5.2 An Econometric Model for Work Absences

I focus on modeling the working process (i.e., 'ask for a leave' decisions), the absence process (i.e., 'return to work' decisions) model shares the identical structure up to parameters and hence will be omitted in description. Figure 3 illustrates one possible work absence record.

Insert Figure 3

Here, solid lines are working durations $\{t_{ij}\}$ and dashed lines are absence durations $\{\tilde{t}_{ij}\}$.

Some words on the absence process are in order. I distinguish absences by their duration length: Recall, a worker is eligible to receive the government SSP benefit if the sick-leave duration is longer than three days. Any absence is a long-term absence if it lasts more than three days, and an absence is a short-term absence if it lasts within three days. Almost all the long-term absences in the data-set are associated with sick-leaves, while most short-term absences are nonmedical. Modeling the short-term absence duration is not a primary interest for two reasons. First, from the perspective of management, it is more important to understand the frequency of short-term absences rather than to understand their duration lengths. Second, from the perspective of modeling, since short-term absences can only last one, two or three days, a multinomial framework is better suited. Importantly, there is no distinction between short and long-term absences in the 'ask for a leave' model. This is because when a worker is making such decisions, s/he is unable to determine the precise length of the absence duration due to incomplete information. For example, the cause of a headache may be the lack of rest or a serious illness, but a worker can hardly know the true cause prior alone.

The working duration model shares a similar specification with the one mentioned in *Example 1*:

$$T_{ij}^{\alpha+1} = (\alpha+1)\exp(\beta x_{ij} - \eta_i)u_{ij}$$

where η_i is an individual's unobserved heterogeneity, $u_{ij} \overset{i.i.d}{\sim} EXP(1)$, and α measures the duration dependence. x_{ij} is the absence score before $j-$th absence, $x_{i0} = 0$, $x_{ij} = \sum_{k=1}^{j-1} \tilde{t}_{ik}$. Since

$$s = \sum_{k=1}^{j} t_{ik} + \sum_{k=1}^{j-1} \tilde{t}_{ik}$$

is a calendar time, x_{ij} can also be constructed using working durations: $x_{ij} = s - \sum_{k=1}^{j} t_{ik}$, i.e., $x_{ij} \in \mathcal{F}_i(s-)$, and $\mathcal{F}_i(s) = \sigma\left(\{t_{ik}\}_{k=1,\ldots,j}, \{\tilde{t}_{ik}\}_{k=1,\ldots,j-1}\right)$.

Counting processes $\{N_i\}_{i=1,\ldots,n}$ consisting of $\{t_{ij}\}_{i=1,\ldots,n;j=1,\ldots,n_i}$ as well as their cumulative intensity functions $\{\Lambda_i\}_{i=1,\ldots,n}$ can be constructed. After a first difference transformation, a vector of minimum distance estimators could be obtained.

5.3 Main Results

Table 4 presents the main results. Column (1) reports the state dependent effects of pooled attendance duration, where I do not distinguish short and long-term absences. Both duration dependent and absence score effects are significant. A positive duration dependence is observed: The previous absence is more likely to trigger the next one if they are far away from each other. This result is consistent with the standard labor-leisure theory.

Insert Table 4

The cause of an absence obviously plays a significant role in the decision-making process. I investigate how individuals react to the absence score if they have high confidence that the cause of an absence is related to a health issue. Notice that almost all the long-term absences in our data are related to health problems. Absences that follow immediately these long-term recoveries might be related to health issues. A typical example of such an absence could be the medical re-check. Column (2) of table 4 studies these 'ask for a leave' decisions, and the results indicate that both duration dependence and the absence score coefficient are no longer significant: The dynamic effects will be irrelevant when workers know (or at least they thought they knew) the causes of absences are health related.

Column (3) reports the results of long-term absence duration. Again, one observes a positive duration dependence: The longer one stays in absence, the more likely she/he is going to return to work. However, individuals do not respond to the absence score. This result is consistent with the conclusion before: Long-term absences are sick-leaves and when it comes to health issues, the absence score is no longer relevant. This also implies that one might assume that long-term recovery durations are i.i.d, and could reconsider a conventional duration model for long-term recoveries.

6. Related Topics and Discussion

In this section, I discuss some related topics, including estimating time-invariant covariates, estimating count statistics through a parametric bootstrap, a structural interpretation for a multi-state duration model, and problems when using likelihood-based estimation methods.

6.1 Estimating Time-invariant Covariates

In the model specification, observed heterogeneities are absorbed in the unobserved individual effect ν_i. The constructed counting process after the first difference transformation does not contain any time-invariant effect. One can, nevertheless, study these heterogeneities by focusing on one particular subset of data, where the state dependent elements are not present. One then can use Heckman and Singer's Non-Parametric Maximum Likelihood (NPML) method to estimate a conventional single spell duration model. In the work absence application, I use the initial duration of newly hired workers to estimate individual heterogeneity effects. The results are presented in Appendix D. One caveat of this strategy is that one must assume that these time-invariant effects are the same in both the initial duration and the subsequent duration.

6.2 Estimating Count Statistics

Since there is no closed form expression for (conditional) probabilities of $N_i(\tau) = m$, estimating count statistics by bootstrap might be a reasonable approach. In this subsection, I propose a parametric bootstrap procedure.

Suppose a GAFT model is specified and estimated using the methods mentioned before, and one obtain estimators $\hat{\theta} = \theta_0 + o_p(1)$ for parameters of time-dependent covariates. The goal is to estimate count probabilities and statistics conditional on a given history up to $s_{i1} = t_{i1}$, i.e., the occurrence time for the first event.

Preliminary: Random Time Change Theorem. One can use the cumulative intensity to transform the original event times $\{s_{ij}\}_{j\geq 1}$ to another sequence of event times $\{\zeta_{ij}\}_{j\geq 1}$ to form a homogeneous Poisson process with unit intensity. Equivalently, the original non i.i.d durations $\{t_{ij}\}_{j\geq 1}$ can be transformed into new durations $\{w_{ij}\}_{j\geq 1}, w_{ij} = \zeta_{ij} - \zeta_{i(j-1)}$, which are i.i.d $EXP(1)$, see Daley and Vere-Jones (2007). The time transformation is given by:

$$\zeta_{ij} = \Lambda_i(s_{ij}; \theta_0), \quad \zeta_{i0} = 0$$

the associated transformed durations w_{ij} are given by

$$w_{ij} = \zeta_{ij} - \zeta_{i(j-1)} = \Lambda_i(s_{ij}; \theta_0) - \Lambda_i(s_{i(j-1);\theta_0})$$

Denote $\tilde{N}_i, \tilde{\Lambda}_i$ a counting process and its cumulative intensity function after one performs the first difference transformation to the original data. Given information $s_{i1} = t_{i1}$, the bootstrap algorithm is:

1. Generate a w^*_{ij} from $EXP(1)$ distribution, the bootstrap transformed event time is updated by $\zeta^*_{ij} = \zeta^*_{i(j-1)} + w^*_{ij}$, and consequently, the event times $\tilde{s}_{ij}$ in the counting process $\tilde{N}_i$ is constructed by $\tilde{s}_{ij} = \tilde{\Lambda}_i^{-1}(\zeta^*_{ij}; \hat{\theta})$

2. Calculate the duration $\Delta\tilde{L}(t_{ij}) = \tilde{s}_{ij} - \tilde{s}_{i(j-1)}$, and generate the original duration t_{ij} by
$$t_{ij} = \tilde{L}^{-1}(\tilde{s}_{ij} - \tilde{s}_{i(j-1)} + \tilde{L}(t_{i(j-1)}))$$

3. Update the original event time by
$$s_{ij} = s_{i(j-1)} + t_{ij}$$

Repeat Steps 1-3 if $s_{ij} < \tau$, and stop the procedure otherwise. Repeat the bootstrap B times, and for k-th iteration, denote $n_i^k = N_i(\tau)$, the number of events occurred before the terminal time τ. One can estimate the following count probabilities and statistics:

- $\widehat{\Pr}\{N_i(\tau) = m | \mathcal{F}_i(t_{i1})\} = \sum_{k=1}^{B} \mathbb{I}\{n_i^k = m\}/B$
- $\widehat{\mathbb{E}}\left(N_i(\tau) | \mathcal{F}_i(t_{i1})\right) = \sum_{k=1}^{B} n_i^k / B$

6.3 A Structural Interpretation for a Multi-State Duration Model

In a multi-state duration model, an individual would transit from one state to another state in a continuous time setup. For example, in the work absence application, a worker exists a working period and enters into an absence period by asking for leave. Here, I show that the reduced form of such a model can be specified as a GAFT model. The strategy presented here simplifies the interdependent duration model of Honor and De Paula (2010); Honoré and de Paula (2018) and is in line with the classical repeated search model in which an individual would compare utilities in two states, see Mortensen (1986) for a review. The transition time from state A to state B depends solely on the difference in the discounted future utilities between these two states. The levels of utilities do not matter.

Let $0 \leq K_{i1} < M_{i1} < K_{i2} < M_{i2} \cdots, K_{ij}, M_{ij} \in (0, \tau]$, where $\{K_{ij}\}_{j\geq 1}$ and $\{K_{ij}\}_{j\geq 1}$ are the random variables representing beginning times of corresponding states. For example, K_{ij} could be the starting date of jth employment, while M_{ij} is the starting date of jth unemployment. Fix any $j = 1, 2, \ldots, n_i$, conditional on information up to time $k_{i(j-1)}$, the utility of an individual i, who chooses to switch from one state to another state at a time $M_{i(j-1)}$ is

$$\int_{k_{i(j-1)}}^{M_{i(j-1)}} \nu_i u_{ij} e^{-\rho z} dz + \int_{M_{i(j-1)}}^{\mathbb{E}(K_{ij})-M_{i(j-1)}} L(z) G(x_{ij}) e^{-\rho z} dz$$

or

$$\int_{0}^{T_{i(j-1)}} \nu_i u_{ij} e^{-\rho z'} dz' + \int_{T_{i(j-1)}}^{\mathbb{E}(K_{ij})-k_{i(j-1)}} L(z') G(x_{ij}) e^{-\rho z'} dz'$$

where

- $T_{i(j-1)} = M_{i(j-1)} - k_{i(j-1)}$ is the duration in state A, $\mathbb{E}(K_{ij}) - k_{i(j-1)} - T_{i(j-1)}$ is the elapsed time in state B, $z' = z - k_{i(j-1)}$.
- ν_i is the unobserved heterogeneity, u_{ij} is an i.i.d random variable that represents an individual's preference taste in the current state (state A).
- $G : \mathbb{R} \to \mathbb{R}_+$ is a function of the state dependent variable x_{ij} and $L : \mathbb{R}_+ \to \mathbb{R}_+$ represents the duration dependence.

I normalize the utility flow in state A to be an individual-specific constant rate (u_{ij}),and assume the utility flow would discount exponentially at ρ. Finally, the multiplicative structure for $L(z)G(x_{ij})$ is imposed because I want the resulting model to have a GAFT model.

The first order condition with respect to $T_{i(j-1)}$ is:

$$\nu_i u_{ij} e^{-\rho T^*_{i(j-1)}} - L(T^*_{i(j-1)}) G(x_{ij}) e^{-\rho T^*_{i(j-1)}} = 0$$

Thus,

$$T^*_{i(j-1)} = (G(x_{ij}))^{-1} \nu_i u_{ij}$$

which is the GAFT specification.

6.4 Likelihood Based Methods are Incompatible

The likelihood-based methods are commonly used in both the duration analysis and the counting process analysis. However, these methods are incompatible with the data set studied here. To demonstrate the point, let's first consider the joint density of event times $(S_{i1}, S_{i2}, \ldots, S_{i(N_i(\tau))})$ of a counting process (or equivalently, the durations):

$$f_{S_{i1},\ldots,S_{i(N_i(\tau))}}\left(s_{i1},\ldots,s_{i(N_i(\tau))}\right) = \exp\left\{-\int_0^\tau \lambda_i\left(t \mid \mathcal{F}_i(t-)\right)dt\right\} \times \prod_{j=1}^{N_i(\tau)} \lambda_i\left(s_{ij} \mid \mathcal{F}_i(s_{ij}-)\right)$$

see Kass et al. (2014) for the proof. The events times $S_{i1}, S_{i2}, \ldots, S_{i(N_i(\tau))}$ have two sources of randomness: One is due to the variability described by the point process p.d.f, and the second is due to the way $N_i(\tau)$ varies. Notice that random variables $N_i(\tau)$ and $\mathbb{I}\{S_{i(N_i(\tau)+1)} > \tau\}$ contain the same amount of information, while $\mathbb{I}\{S_{i(N_i(\tau)+1)} > \tau\} \in \mathcal{F}_i(\tau)$. Thus, the second source of randomness ultimately comes from the variation of $\mathcal{F}_i(s)$. Stochastic processes like the $\{S_{i1}, S_{i2}, \ldots, S_{i(N_i(\tau))}\}$ are called *Doubly Stochastic Processes*. A doubly stochastic model can be represented in two stages. In the first stage, the distribution of the outcome is parametrically specified by a certain density function. At the second stage, some elements in the density function are treated as being random. One classical doubly stochastic process is the Cox process, also known as the doubly stochastic Poisson process.

Considering a likelihood contributor conditional on event times $\{s_{ij}\}_{j=1,\ldots,n_i}$ from an individual i:

$$l_i = \prod_{k=1}^{n_i} \lambda_i(s_{ik} \mid \mathcal{F}_i(s_{ik}-)) \times \exp\left(-\sum_{k=1}^{n_i}\int_{s_{i(k-1)}}^{s_{ik}} \lambda_i(z \mid \mathcal{F}_i(z-))dz\right) \times \exp\left(-\int_{s_{i(n_i)}}^{\tau} \lambda_i(z \mid \mathcal{F}_i(z-))dz\right)$$

which is also the specification of the joint density of the occurrence times $s_{i1}, \ldots, s_{i(n_i)}$ and the number of occurrences $N_i(\tau) = n_i$, see Rubin (1972) for reference.

Using the likelihood function constructed from $\{l_i\}_{i=1,\ldots,n}$ would not lead to consistent estimators, since the likelihood contributors are not identical. One way to obtain identical likelihood contributors is to fix the number of occurrences across individuals:

$$L = \prod_{i:n_i=n^*} l_i^{n^*}$$

where $l_i^{n^*}$ is the likelihood contributor whose number of occurrences is n^*. However, this practice would lead to both a sample selection problem and a loss of information problem: Individuals with $n_i < n^*$ would be ignored from the estimation, while individuals with $n_i > n^*$ would not have their information fully utilized by researchers.

7. Conclusion

In this paper, I proposed a multiple spell duration model to study a set of duration data $\{T_{ij}\}$ where $i = 1, \ldots, n$ and for each individual i, $j = 1, \ldots, N_i$. These data are doubly stochastic: Both the durations $\{T_{ij}\}_{j\geq 1}$ and the number of duration N_i are random for a fixed observation window. In addition, within each individual, past durations have genuine effects on the future ones. These state dependent effects hold true after conditional on the unobserved heterogeneity. I use a counting process approach to construct the multiple spell duration model for two reasons. First, a counting process fully characterizes the doubly stochastic structure of the data. Second, a martingale relationship between the process and its cumulative intensity provides an estimation channel.

More specifically, fix an individual i, I specify a generalized accelerated failure time (GAFT) structure for each duration. This structure is broad and contains some important duration models, e.g., the mixed proportional hazard model. It also facilitates a researcher to perform a first difference transformation to cancel fixed effects. One can construct a counting process for each individual using his/her duration data. The corresponding (cumulative) intensity function can be written using the GAFT specification. Count probabilities can be specified using modified intensity functions, although such specifications are expressed through difference-differential equations.

I use the first difference transformation to cancel the fixed effects. Such a transformation is achievable with some mild restrictions on model specifications. A transformed counting process is constructed using first differenced durations. The associated (cumulative) intensity can be specified by imposing distribution assumptions on the error term of a GAFT model.

One can estimate the model by minimizing the distance between the transformed counting process and its cumulative intensity. I re-introduced the minimum distance estimator proposed by Kopperschmidt and Stute (2013), and provide simplified proofs. I also conduct simulation studies that employ the first difference transformation.

I use the proposed model to study a work absence application. A worker's absence decisions include a decision to 'ask for a leave' and a decision to 'return to work'. These absence decisions are affected by a worker's absence score, which depends on his/her past work absence records. I found that workers will respond to the absence score in general. Higher absence scores discourage future absence initials. However, this pattern does not apply to sick-leave absences.

Lastly, some related topics are discussed. These topics include how to estimate time-invariant covariates, how to estimate count statistics through a parametric bootstrap, and a structural interpretation of a double-state duration model. I also discuss reasons for which conventional likelihood-based methods are incompatible with the doubly stochastic data.

References

Aguirregabiria, V. and P. Mira (2010): "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156, 38–67.

Allen, S. G. (1981): "An empirical model of work attendance," *The Review of Economics and Statistics*, 77–87.

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (2012): *Statistical models based on counting processes*, Springer Science & Business Media.

Baltagi, B. H. and S. H. Song (2006): "Unbalanced panel data: A survey," *Statistical Papers*, 47, 493–523.

Barmby, T., C. Orme, and J. Treble (1995): "Worker absence histories: a panel data study," *Labour Economics*, 2, 53–65.

Barmby, T. A., C. D. Orme, and J. G. Treble (1991): "Worker absenteeism: An analysis using microdata," *The Economic Journal*, 101, 214–229.

Blundell, R. and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics*, 87, 115–143.

Carrasco, M. and J.-P. Florens (2000): "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, 16, 797–834.

Daley, D. J. and D. Vere-Jones (2007): *An introduction to the theory of point processes: volume II: general theory and structure*, vol. 1,2, Springer Science & Business Media.

Duflo, E., R. Hanna, and S. P. Rya (2012): "Incentives work: Getting teachers to come to school," *The American Economic Review*, 102, 1241–1278.

Elbers, C. and G. Ridder (1982): "True and spurious duration dependence: The identifiability of the proportional hazard model," *The Review of Economic Studies*, 49, 403–409.

Fevang, E., S. Markussen, and K. Røed (2014): "The sick pay trap," *Journal of Labor Economics*, 32, 305–336.

Fister-Gale, S. (2003): "Sickened by the cost of absenteeism, companies look for solutions," *Workforce*, 82, 72–75.

Heckman, J. and B. Singer (1984): "The identifiability of the proportional hazard model," *The Review of Economic Studies*, 51, 231–241.

Heckman, J. J. (1978): "Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence," in *Annales de l'INSEE*, JSTOR, 227–269.

——— (1981): "Heterogeneity and state dependence," in *Studies in labor markets*, University of Chicago Press, 91–140.

Heckman, J. J. and G. J. Borjas (1980): "Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence," *Economica*, 47, 247–283.

Heckman, J. J., V. J. Holtz, and J. R. Walker (1985): "New evidence on the timing and spacing of births," *The American Economic Review*, 75, 179–184.

Heckman, J. J. and J. R. Walker (1990): "The relationship between wages and income and the timing and spacing of births: Evidence from Swedish longitudinal data," *Econometrica: journal of the Econometric Society*, 1411–1441.

Honor, B. E. and Á. De Paula (2010): "Interdependent durations," *The Review of Economic Studies*, 77, 1138–1163.

Honoré, B. E. (1993): "Identification results for duration models with multiple spells," *The Review of Economic Studies*, 60, 241–246.

Honoré, B. E. and Á. de Paula (2018): "A new model for interdependent durations," *Quantitative Economics*, 9, 1299–1333.

Hsiao, C. (2014): *Analysis of panel data*, 54, Cambridge university press.

Kass, R. E., U. T. Eden, E. N. Brown, et al. (2014): *Analysis of neural data*, vol. 491, Springer.

Kopperschmidt, K. and W. Stute (2013): "The statistical analysis of self-exciting point processes," *Stast. Sinica*, 23, 1273–1298.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013): "Duration dependence and labor market conditions: Evidence from a field experiment," *The Quarterly Journal of Economics*, 128, 1123–1167.

Lee, S. (2008): "Estimating panel data duration models with censored data," *Econometric Theory*, 24, 1254–1276.

Markussen, S., K. Røed, O. J. Røgeberg, and S. Gaure (2011): "The anatomy of absenteeism," *Journal of health economics*, 30, 277–292.

Mortensen, D. T. (1986): "Job search and labor market analysis," *Handbook of labor economics*, 2, 849–919.

ÖZTÜRK, Ö. and T. P. Hettmansperger (1997): "Generalised weighted Cramér-von Mises distance estimators," *Biometrika*, 84, 283–294.

Rao, R. R. (1962): "Relations between weak and uniform convergence of measures with applications," *The Annals of Mathematical Statistics*, 659–680.

Ridder, G. (1990): "The non-parametric identification of generalized accelerated failure-time models," *The Review of Economic Studies*, 57, 167–181.

Rubin, I. (1972): "Regular point processes and their detection," *IEEE Transactions on Information Theory*, 18, 547–557.

Steel, R. P., J. R. Rentsch, and J. R. Van Scotter (2007): "Timeframes and absence frameworks: A test of Steers and Rhodes'(1978) model of attendance," *Journal of Management*, 33, 180–195.

Steers, R. M. and S. R. Rhodes (1978): "Major influences on employee attendance: A process model." *Journal of applied Psychology*, 63, 391.

A. Tables and Figures

Table 1: Results of DGP1

$N = 100, \tau = 3500$	True	Estimator	SD	MAD	CI95	CI90
β	-0.15	-0.166	0.029	0.019	100%	97%
$N = 100, \tau = 5000$						
β	-0.15	-0.172	0.029	0.018	98.5%	98%
$N = 200, \tau = 3500$						
β	-0.15	-0.165	0.020	0.013	99%	95.5%
$N = 200, \tau = 5000$						
β	-0.15	-0.167	0.020	0.011	97.5%	93.5%

* SD stands for the standard deviation of each simulation. MAD is the median absolute deviation. CI95(CI90) is the percentage of the 95%(90%) confidence interval generated by se that covers the true parameter.

Table 2: Results of DGP2

$N = 100, \tau = 3500$	True	Estimator	SD	MAD	CI95	CI90
α	0.5	0.508	0.069	0.058	97%	95.5%
β	-0.05	-0.053	0.041	0.024	99.5%	97%
$N = 100, \tau = 5000$						
α	0.5	0.531	0.070	0.061	92%	89%
β	-0.05	-0.080	0.055	0.027	98.5%	97%
$N = 200, \tau = 3500$						
α	0.5	0.512	0.051	0.043	97%	91.5%
β	-0.05	-0.054	0.030	0.019	98%	95%
$N = 200, \tau = 5000$						
α	0.5	0.519	0.050	0.049	95.5%	91.5%
β	-0.05	-0.058	0.030	0.017	100%	99%

* SD stands for the standard deviation of each simulation. MAD is the median absolute deviation. CI95(CI90) is the percentage of the 95%(90%) confidence interval generated by se that covers the true parameter.

B. Proofs

B.1 Theorem 1

I use Figure 4 to help illustrating the proof. Fix a time t, the value of $\Lambda_i(s)$ is:

Table 3: Results of DGP3

$N = 100, \tau = 10$	True	Estimator	SD	MAD	CI95	CI90
α	1.25	1.283	0.137	0.084	100%	99.5%
β	1.65	1.703	0.183	0.102	97.5%	99%
$N = 100, \tau = 15$						
α	1.25	1.312	0.109	0.072	99%	96%
β	1.65	1.746	0.152	0.087	99%	94%
$N = 200, \tau = 10$						
α	1.25	1.272	0.096	0.061	98.5%	96%
β	1.65	1.697	0.129	0.081	98%	97.5%
$N = 200, \tau = 15$						
α	1.25	1.323	0.077	0.047	94.5%	88.5%
β	1.65	1.761	0.107	0.071	92.5%	85.5%

* SD stands for the standard deviation of each simulation. MAD is the median absolute deviation. CI95(CI90) is the percentage of the 95%(90%) confidence interval generated by se that covers the true parameter.

Table 4: State-Dependent Effect

	pooled attendance	*sick-related attendance*	*long-term recovery*
	(1)	(2)	(3)
α	0.309***	0.053	1.122***
	(0.071)	(0.111)	(0.189)
β	0.916***	-0.079	-0.063
	(0.321)	(0.057)	(0.097)

Note: Column (1) reports attendance duration results. Column (2) reports the result for sick short-term attendance duration. Column (3) and (4) report the state-dependent results for short and long-term recovery duration, respectively. α is the duration dependent coefficient, while β is the absence score coefficient. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

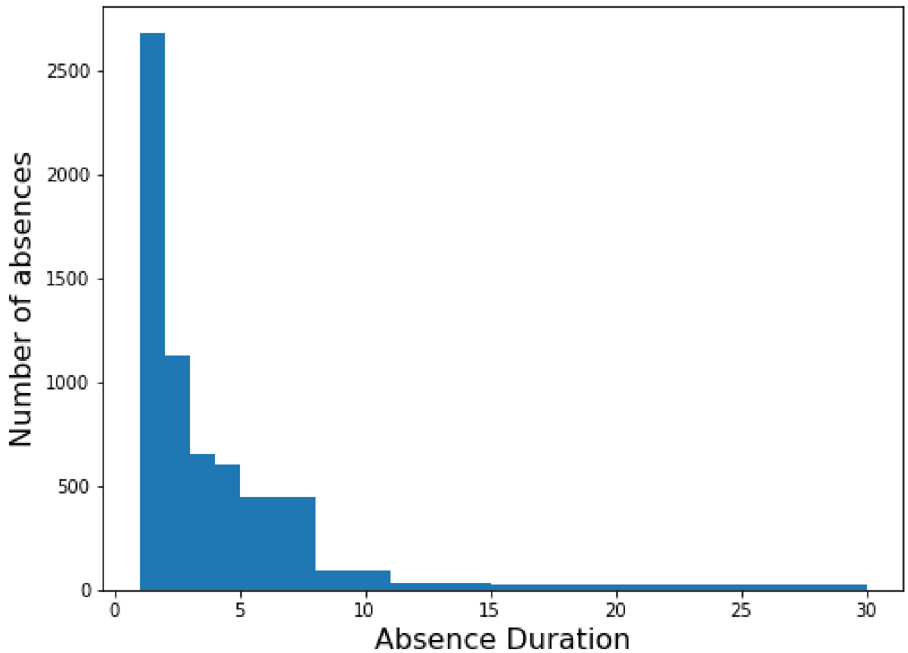
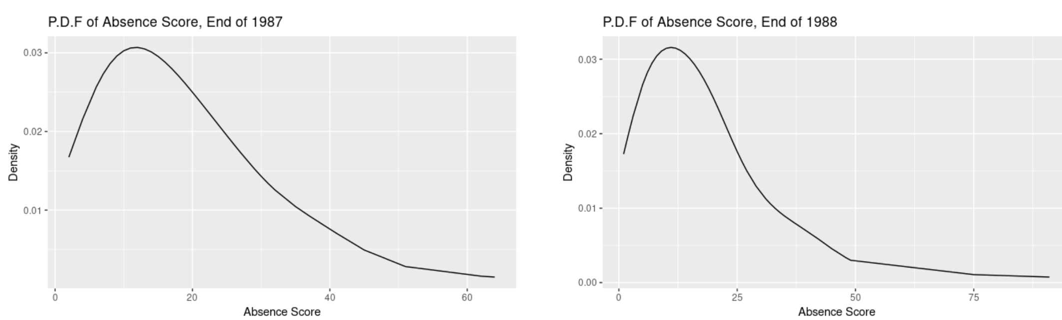


Figure 1: Most frequent absence duration



(a) At the end of 1987

(b) At the end of 1988

Figure 2: Non-parametric P.D.F of absence scores

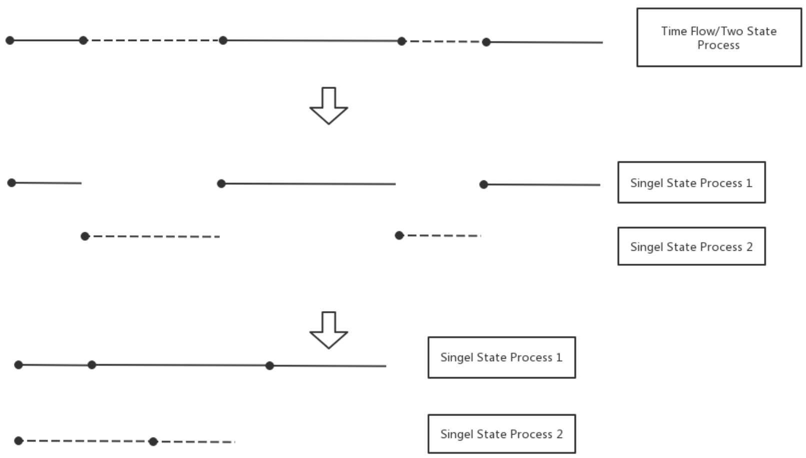


Figure 3: Decompose two state process into single state processes

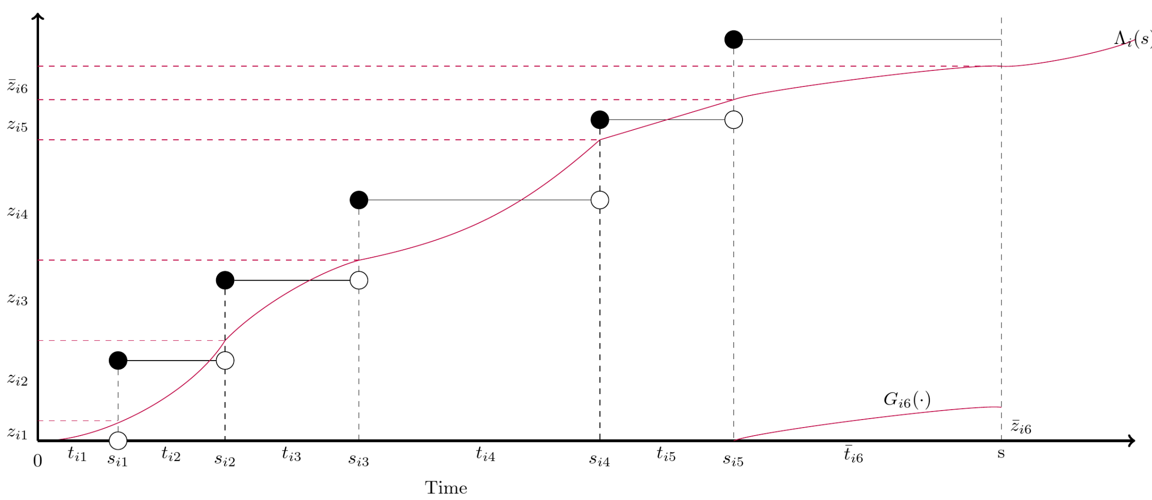


Figure 4: A Possible Realization of a Counting Process and its Cumulative Intensity

$$\Lambda_i(s) = \sum_{j=1}^{5} z_{ij} + \bar{z}_{i6}$$

where

$$z_{ij} = \Lambda_i(s_{ij}) - \Lambda_i(s_{i(j-1)}) = G_{ij}(t_{ij}), \quad j = 1, \ldots, 5, s_{i0} = 0$$

and

$$\bar{z}_{i6} = \Lambda_i(s) - \Lambda_i(s_{i5})$$

Here $G_{ij}(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ are non-decreasing functions. In the figure, t_{ij} are complete duration, $\bar{t}_{i6}$ is an incomplete duration. Our proof consists of two steps, first, I show that $z_{ij} = H_{ij}(t_{ij})$, where $H_{ij}(\cdot)$ are the corresponding integrated hazard rates of T_{ij}. Second, I show that $\bar{z}_{i6} = H_{i6}(\bar{t}_{i6})$.

From the randome time theorem, one knowns that $z_{ij}, j = 1, \ldots, 5$ are the realizations of an EXP(1) random variable, thus

$$P(Z_{ij} > x) = P(G_{ij}(T_{ij}) > x) = P(T_{ij} > G_{ij}^{-1}(x))$$

and,

$$1 - F_{ij}(G_{ij}^{-1}(x)) = \exp\left(-H_{ij}(G_{ij}^{-1}(x))\right) = \exp(-x)$$

In the above equation, the first equality comes from the relationship between the survival function and the integrated hazard rate. Thus,

$$G_{ij}(\cdot) = H_{ij}(\cdot)$$

$$z_{ij} = G_{ij}(t_{ij}) = H_{ij}(t_{ij})$$

Next, notice that

$$P(Z_{i6} > \bar{z}_{i6}) = P(G_{i6}(T_{i6}) > \bar{z}_{i6}) = P(T_{i6} > G_{i6}^{-1}(\bar{z}_{i6}))$$

thus,

$$1 - F_{i6}(G_{i6}^{-1}(\bar{z}_{i6})) = \exp(-\bar{z}_{i6})$$

Further notice that (as illustrated by the figure)

$$G_{i6}^{-1}(\bar{z}_{i6}) = \bar{t}_{i6}$$

Hence,

$$1 - F_{i6}(\bar{t}_{i6}) = \exp(-\bar{z}_{i6})$$

and

$$\bar{z}_{i6} = -\log\left(1 - F_{ij}(\bar{t}_{i6})\right) = \int_0^{\bar{t}_{i6}} h_{i6}(x)dx = H_{i6}(\bar{t}_{i6})$$

B.2 Theorem 2

Since $\int_{\mathcal{T}} M(z;\theta)^2 \mathbb{E}\Lambda(dz;\theta_0)$ has an unique minimizer at $\theta = \theta_0$, then, using the theory of M-estimator, one just need to show uniformly in θ

$$\int_{\mathcal{T}} \bar{M}_n(z;\theta)^2 \bar{N}_n(dt) \overset{a.s}{\to} \int_{\mathcal{T}} M(z;\theta)^2 \mathbb{E}\Lambda(dz;\theta_0)$$

Notice that A2 implies $\bar{M}_n(s;\theta)$ is continuous in θ. This result holds by applying the continuous mapping theorem and the fact that $\bar{N}_n(s), \bar{\Lambda}_n(s;\theta)$ are sample mean of i.i.d nondecreasing process, a Glivenko-Cantelli argument yields, with probability one, uniform convergence of $\bar{N}_n(s), \bar{\Lambda}_n(s;\theta)$ to $\mathbb{E}N_1(s) = \mathbb{E}\Lambda(s;\theta_0), \mathbb{E}\Lambda_1(s;\theta)$, respectively, uniform in s and compact subsets of Θ.

B.3 Theorem 3

The following Lemma is needed to prove Theorem 3.

Lemma A1. Let θ^* be a consistent estimator of θ_0, then

$$\frac{1}{n}\sum_{i=1}^{n} \dot{M}_i(s;\theta^*) \overset{a.s}{\to} \mathbb{E}\dot{M}_1(s;\theta_0)$$

Proof See Rao (1962) ■

The first order condition of the minimization problem is

$$\sum_{l=1}^{N_n} \left(\sum_{i=1}^{n} \dot{M}_i(s_l;\hat{\theta}_n)\right)\left(\sum_{i=1}^{n} M_i(s_l;\hat{\theta}_n)\right) = 0$$

where in a similar notation, $\dot{M}_i(s;\theta) = \partial M_i(s;\theta)/\partial\theta$. By the mean value theorem, one can find an estimator $\theta_n^* = \gamma\hat{\theta}_n + (1-\gamma)\theta_0, \gamma \in (0,1)$ such that

$$\sum_{l=1}^{N_n} \left(\sum_{i=1}^{n} \dot{M}_i(s_l;\hat{\theta}_n)\right)\left(\sum_{i=1}^{n} M_i(s_l;\theta_0)\right) + G_n(\hat{\theta}_n - \theta_0) = 0$$

where

$$G_n = \sum_{l=1}^{N_n} \left(\sum_{i=1}^{n} \dot{M}_i(s_l;\hat{\theta}_n)\right)\left(\sum_{i=1}^{n} \dot{M}_i^{\top}(s_l;\theta_n^*)\right)$$

Therefore

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta_0) &= n^3 G_n^{-1}\left(\frac{1}{n}\sum_{l=1}^{N_n}\left[\frac{1}{n}\sum_{i=1}^{n}\dot{M}(s_l;\hat{\theta}_n)\right]\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_i(s_l;\theta_0)\right]\right) \\
&= n^3 G_n^{-1}\int_{\mathcal{T}}\left[\frac{1}{n}\sum_{i=1}^{n}\dot{M}(z;\hat{\theta}_n)\right]\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_i(z;\theta_0)\right]\bar{N}_n(dz)
\end{aligned}$$

Notice that

$$
\begin{aligned}
n^3 G_n^{-1} &= \left(\frac{1}{n} \sum_{l=1}^{N_n} \left(\frac{1}{n} \sum_{i=1}^{n} \dot{M}_i(s_l; \hat{\theta}_n) \right) \left(\frac{1}{n} \sum_{i=1}^{n} \dot{M}_i^\top(s_l; \theta_n^*) \right) \right)^{-1} \\
&= \left(\int_{\mathcal{T}} \left(\frac{1}{n} \sum_{i=1}^{n} \dot{M}_i(z; \hat{\theta}_n) \right) \left(\frac{1}{n} \sum_{i=1}^{n} \dot{M}_i^\top(z; \theta_n^*) \right) \bar{N}_n(dz) \right)^{-1}
\end{aligned}
$$

thus, by LLN and Lemma A1,

$$
n^3 G_n^{-1} \overset{a.s}{\to} \left(\int_{\mathcal{T}} \dot{M}(z; \theta_0) \dot{M}(z; \theta_0)^\top \mathbb{E}\Lambda(dz; \theta_0) \right)^{-1}
$$

Similarly, by Lemma A1, $\frac{1}{n} \sum_{i=1}^{n} \dot{M}(s; \hat{\theta}_n) \overset{a.s}{\to} \dot{M}(s; \theta_0), \forall t \in \mathcal{T}$. Finally, by martingale CLT, one have

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_i(\cdot; \theta_0) \Rightarrow B_\Gamma(\cdot)
$$

where $\Rightarrow$ denotes weakly convergence, and B_Γ is a centered Gaussian process with covariance structure $\Gamma(s_1, s_2) = \mathbb{E}(M_1(s_1; \theta_0) M(s_2; \theta_0))$.

Putting everything together, one has the result stated in Theorem 3.

B.4 Lemma 1 and 2

$$
\begin{aligned}
&\frac{\partial}{\partial \theta} \Pr\{N_i(s) = m | \mathcal{F}_i(u)\} \\
&\quad = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr\{N_i(s + \Delta) = m | \mathcal{F}_i(u)\} - \Pr\{N_i(s) = m | \mathcal{F}_i(u)\} \\
&\quad = \lim_{\Delta \to 0} \frac{1}{\Delta} \left[\Pr\{N_i(s + \Delta) = m | \mathcal{F}_i(u), N_i(s) = m\} - 1 \right] \Pr\{N_i(s) = m | \mathcal{F}_i(u)\} \\
&\quad + \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr\{N_i(s + \Delta) = m | \mathcal{F}_i(u), N_i(s) = m - 1\} \Pr\{N_i(s) = m - 1 | \mathcal{F}_i(u)\}
\end{aligned}
$$

Recall the definition of $\lambda_{N_i(s)}(s, \mathcal{F}_i(u))$, one has the results asserted in Lemma 1. In the same way, Lemma 2 can be proved.

C. Simulation Algorithm

The *Time Rescaling Theorem* suggests generating a sequence of $EXP(1)$ random variables and then back-transforming to get the desired counting process. The following is the algorithm for generating a process on the interval $(0, \tau]$ with conditional intensity $\lambda_i(t \mid \mathcal{F}_i(t-))$:

1. Initialize $s_{i0} = 0$ and $j = 1$.

2. Sample z_{ij} from an $EXP(1)$ distribution.

3. Find s_{ij} as the solution to

$$z_{ij} = \int_{s_{i(j-1)}}^{s_{ij}} \lambda_i(t \mid \mathcal{F}_i(t-))dt$$

4. If $s_{ij} > \tau$ stop.

5. Set $j = j + 1$, update the filtration $\mathcal{F}_i()$ and go to step 2.

D. Miscellaneous Results

D.1 Simulation Results: Density Functions

The following figures report kernel density estimation results of the simulated studies.

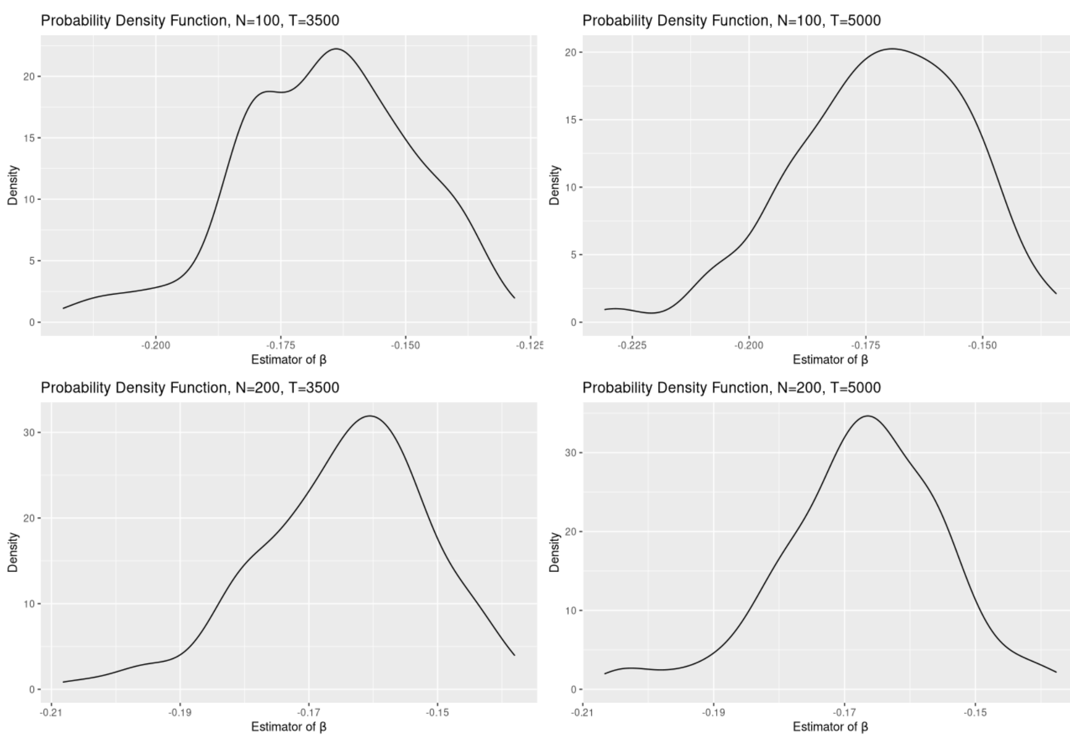


Figure 5: Density Function, DGP1

D.2 NPMLE Results

I specify the hazard rate for such initial duration as

$$h_{working}(d) = \exp(x_i^\top \beta + \eta_i)$$

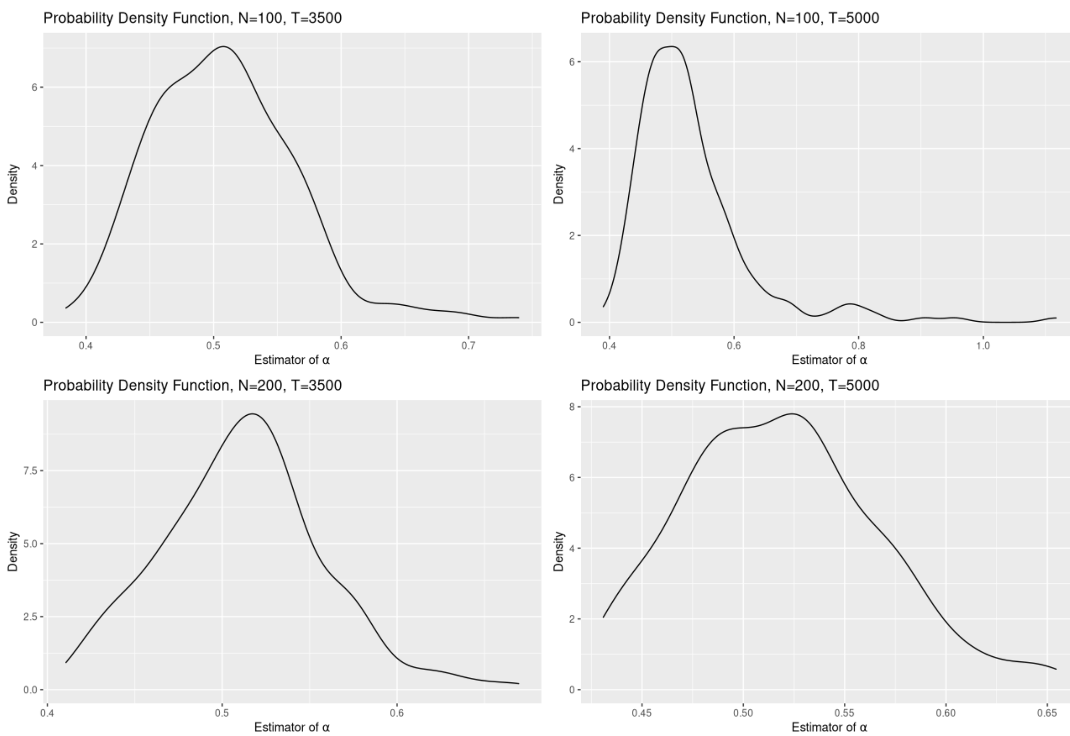


Figure 6: Density Function of $\hat{\alpha}$, DGP2

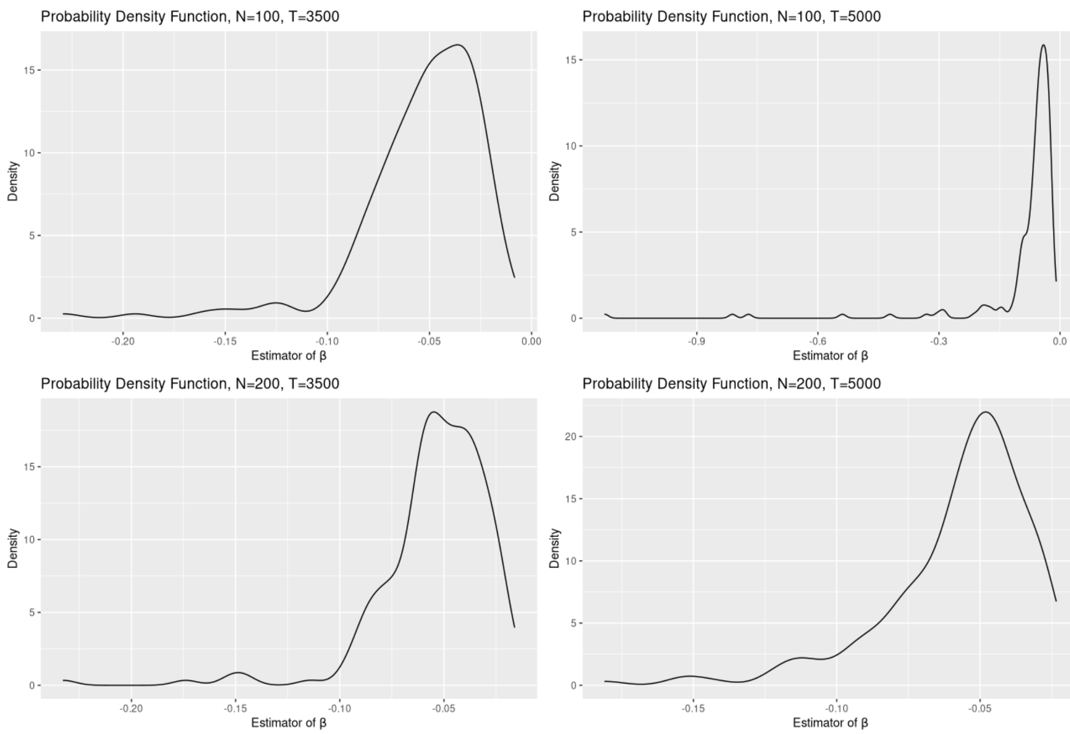


Figure 7: Density Function of $\hat{\beta}$, DGP2

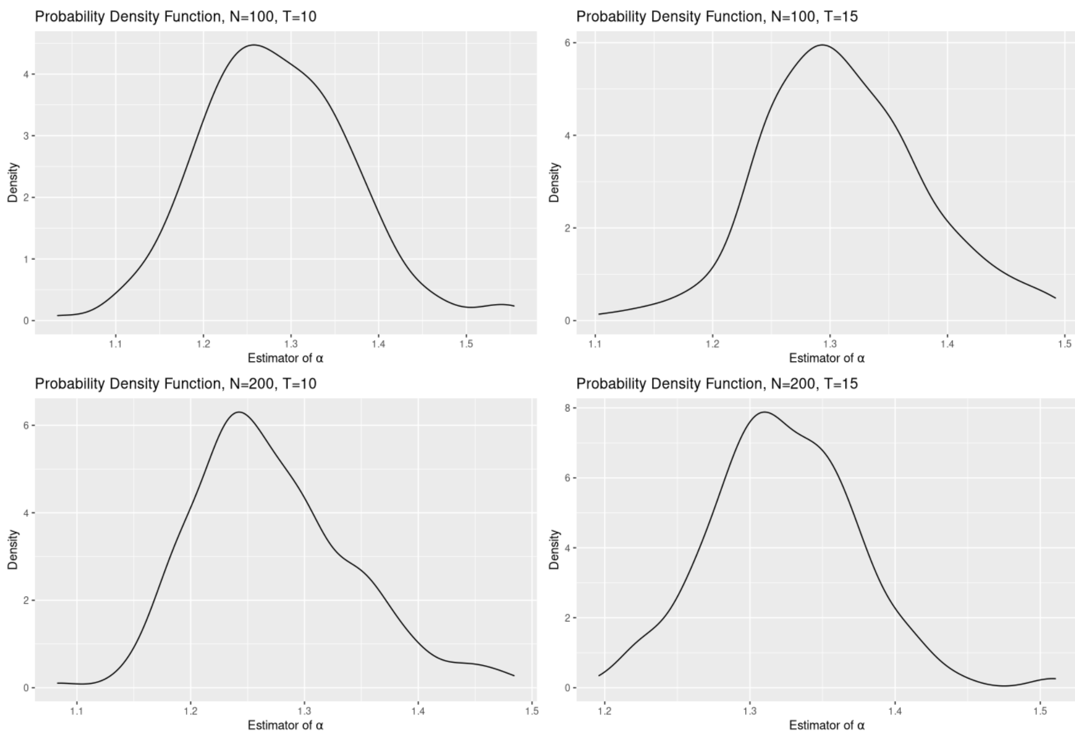


Figure 8: Density Function of $\hat{\alpha}$, DGP3

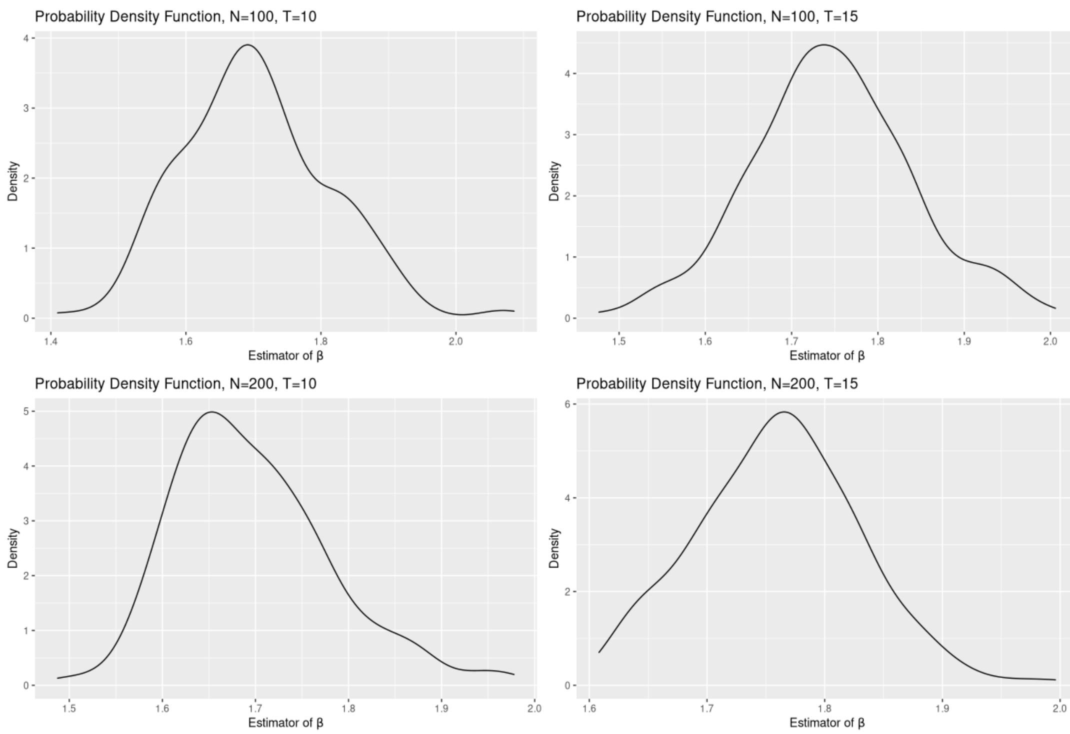


Figure 9: Density Function of $\hat{\beta}$, DGP3

Table 5: NPMLE Results

	Attendance	*Long-Term Absence*
	(1)	(2)
age	-0.056***	0.086***
	(0.015)	(0.012)
age2	0.107***	-0.167***
	(0.032)	(0.023)
male	0.002	0.217***
	(0.138)	(0.083)
full time	0.042	0.109
	(0.151)	(0.087)
marriage	0.052	0.013
	(0.208)	(0.087)
α	–	0.803***
	–	(0.039)

Note: $age2 = age^2/100$, male is a gender indicator with value one if a person is male and zero otherwise, full time is an indicator with value one if a person holds a full time labor contract, and zero otherwise. *p<0.1; **p<0.05; ***p<0.01

where ν_i is the individual unobserved heterogeneity, and x_i is a vector of time-invariant individual attributes including age, sex, full time/ part time job indicator and marriage status.

Results in Table 4 also suggest that long-term absence durations could be regarded as i.i.d variables. I specify the following hazard rate:

$$h_{absence}(t) = t^{\alpha} \exp(x_i^{\top} \beta + \eta_i)$$

where t is the long-term absence duration.

Heckman & Singer's Non-Parametric Maximum Likelihood Estimator (NPMLE) is employed to estimate the parameters. Table 5 presents results when using the initial attendance data from the newly-hired workers and the long term absence duration records from the whole sample.