Econometric Modeling of Self-Exciting Process

by

Yuhao Li

A dissertation submitted by in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Economics

Universidad Carlos III de Madrid

Advisor:

Miguel A. Delgado

May 2019

To my parents: 李晓青,吴玉华 *for everything you gave me*

Acknowledgments

My mother gave me a dictionary as a gift before I went to the elementary school, on the flyleaf, she wrote: ”Wish you grow up in learning and become a doctor.” Of course, I paid no attention to what she wrote then as I couldn’t read at that time (hence the dictionary), but I do have memories about this book because it has full colored figures, which helped me attract attention from my fellow classmates. It caught my eyes again in the Spring festive of this year when I was organizing my bookshelf, and these words revealed themselves in a most unexpected way. I showed it to my mom, but she said the word ’doctor’ is just a metaphor.

Nevertheless, what a prophecy!

Now, here I am, writing the acknowledgments of my thesis! I would not be able to do so without the help and guidance of Miguel. If my mom accidentally planted the seed, it is you that fertilize the fruit. I owe you so many thanks for so many things, I just do not know where to begin. Thank you for pushing me to produce quality research, for being patient and encouraging when I failed, for being supportive when I needed and for being a friend. Muchas Gracias!

I am grateful to Winfred Stute, without his lectures in Carlos III, this thesis may never exist.

Special thanks to wonderful Rui and Minghai: despite our differences, I am most enjoy your companions both at the office and home. I will secretly admit that I skipped some cleaning duties, but will officially deny any accuse.

To Julius, Miguel Angel and Pancho, I will remember our Amsterdam trip forever. It is an honor and glory to fight with you and eventually brought the Econometric Game Trophy back to Madrid.

I am lucky to have best colleagues during the Ph.D. journey: Junyi, Bea, Tomas, Ursula, Xiaojun, Yunrong, Meng, Mian, Weifeng, Chang, Boyao, and my fantastic officemates: Cleiton and Federico Masera.

My best buddy Qi Rui, thank you for all the meaningful and meaningless chats across six time zones.

Finally, big thanks to my family, my parents Li Xiaoqing and Wu Yuhua, my grandmoms, Li Yaqun and Sun Kaimin, for all the unconditional love and support.

Financial support from the Ministerio de Economía y Competitividad (Grant Number: BES 2015-072230) is gratefully acknowledged.

Dissertation Advisor: Miguel A. Delgado Yuhao Li

Econometric Modeling of Self-Exciting Process

Abstract

This thesis focuses on modeling state dependence, a phenomenon where past experiences do alter the path of future events. The general idea behind this concept is that differences among individuals are not merely explained by their characteristics, but also by their past experiences. Typical examples of state dependence in economics include the incidence of accidents, labor force participation and unemployment, consumer's purchase behaviors, learning etc.

I use the self-exciting process to incorporate the state dependent structure in economic models. The self-exciting process is a counting process whose filtration includes a σ-field that is generated by the process itself.

Chapter I introduces notation and provides an introduction to the self-exciting process. A minimum distance estimation (MDE) method is also introduced. Monte Carlo exercises are performed to investigate the MDE performance. I also provide a comparison among the self-exciting process approach and existing methods such as counting data regression and duration models.

Chapter II studies how past doctor visiting records could alter the preference individuals' future medical consumption choices, especially through the channel of a cost-sharing health insurance plan. This study contributes to the existing health insurance literature by providing additional evidence that supports the shadow price theory. While most previous studies that endorse the shadow price use specific company or social security designs, I investigate the shadow price using a genuine health insurance contract where medical cost has multiple sources.

Chapter III contributes to the study of work absence. Early works in this literature usually focus on absence duration and assume independence among the duration. I took another approach in this study where both absence duration and working duration (the length of a working spell until an absence occurred) are analyzed. Special attention is given to studying how the state dependent absence score could affect these duration. I study a particular firm who has installed a experience-rated work absence regulation. I also distinguish between short and long term absences and find that individuals have different state dependent reaction to different types of absences.

Chapter IV investigates the classical unemployment duration problem. I restrict the attention to Spanish Youth, who are well known for their high job turnover rate. A crucial element in this literature is the separation of the state dependent and the unobserved heterogeneity. I did so by assuming a multiplicative duration structure and perform a first ratio transformation on the individual's unemployment duration to swipe out the unobserved heterogeneity. The new estimator could be regarded as an extension to the existing dynamic panel data model but it avoids using instrumental variables and could allow unit root autoregressive coefficient and non-stationarity process.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In many economic studies, history is a key component of the theory and the decision making process. Examples can be found in consumer behavior, unemployment and discrete choice models. The phenomenon that the past experience could alter the taste of future choices are known as the state dependence in the literature.

Conventionally, researchers tend to use dynamic panel data models or structural models to describe the state dependent structure. However, dynamic panel data requires the underlying processes to be stationary autoregressive, while structural models impose some restrictions that might be not realistic (e.g., individual are risk neutral) for the sake of easing computation burden. Moreover, sometimes it is too complicated to build a structural model, for example, when state variables are updated by both internal and external processes.

In this thesis, I use the self-exciting process to study the state dependent structure. Compare to existing methods, the self-exciting process is flexible in modelling, stationary assumption is not required as in the dynamic panel data models. The self-exciting process is a statistical model that does not dependent on economic theories, thus we are in the spectral of reduced form model.

Three papers are included in this thesis. The first paper studies the shadow price of health insurance, namely how the cumulative medical cost would alter an individual's utilization of medical service. The second paper investigate the dynamic behavior of a worker's work absence under a experience-rated absence regulation. The last paper focuses on the state dependent structure in unemployment.

All three applications are state-dependent in nature. The first application contributes to the health economic literature by providing additional evidence that supports the shadow price theory. Previous literature either use specific company or social security program policies to study the shadow price effect (Aron-Dine et al. 2015; Einav et al. 2015), while in this paper, we investigate this effect under a genuine health insurance where medical cost come from multiple sources (e.g., doctor-visitings, drug purchases). The second paper contributes the study of work absence. To author's best knowledge, it is one of very few works that takes dynamic into consideration. Early work absence literature focus on the absence duration, very often the duration are assumed to be i.i.d (Barmby et al. 1991; Markussen et al. 2011; Delgado & Kniesner 1997). The last paper studies the classical unemployment problem on how to separate the state-dependent effect and the unobserved heterogeneity (Heckman 1981).

The rest of this introduction chapter is dedicated to provide necessary knowledge on the self-exciting process. The estimating method for a parametric self-exciting process as well as the simulation exercises are also included in this chapter. I end this chapter by comparing the self-exciting process with classical count data regression and duration models.

1.1 Some Notations and Basic Concepts about Self-Exciting Process

Briefly speaking, a self-exciting process is a counting process (point process) whose filtration includes the counting process itself. We first introduce the idea of counting process. A counting process is expressed as:

$$N(t) = \sum_{i=1}^{\infty} \mathbb{I}\{t_i \leq t\} \tag{1.1}$$

where $t_i, i \in \mathcal{N}^+$ are occurrence times of realized events, $\mathbb{I}\{\cdot\}$ is the indicator function. An example of such a counting process is illustrated in figure 1.1

We may include marks in the counting process by extending the definition:

$$N(t, R^+) = \sum_{i=1}^{\infty} \mathbb{I}\{t_i \leq t, x_i \in R^+\} \tag{1.2}$$

where x_i is the mark associated with i^{th} event. In our health insurance application, an event is a medical utilization (e.g., a visit to a doctor) and a mark is the cumulative expenditure $x(t_i)$ upon that visiting.

In order to study the definition of the counting process in a rigorous way, introduction of some concepts is in order.

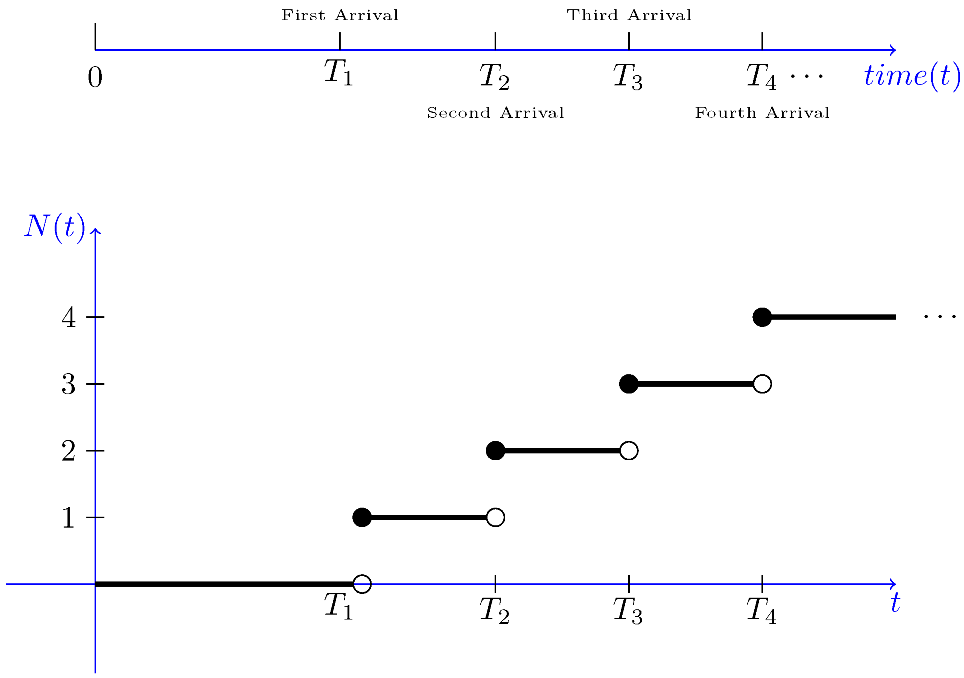


Figure 1.1: A possible realization of a counting process

To begin with, let's first introduce the 'working' space. Let **E** be a locally compact Hausdorff space[1] whose topology has a countable base (LCCB), with Borel $\sigma-$ algebra ϵ. Denote β the algebra (ring) of bounded Borel set.

Let M_E be the space of all bounded finite measures on $\beta(E)$, the random measure maps from probability space $(\Omega, \mathcal{F}, \mathcal{P})$ to the measurable space $(M_E, \beta(M_E))$. Notice that on M_E we usually define the $\sigma-$ ring $\mathcal{M}$ generated by the coordinate mappings:

$$\mu \rightarrow \mu(f) = \int f d\mu$$

where μ is a Borel measure and f ranges over the set C_k of continuous functions whose support is compact.

[1]The Hausdorff space, also known as the separated space, is a topological space where for any two distinct points there exists a neighbourhood of each which is disjoint from the neighbourhood of the other.

Definition A counting process on E is a measurable mapping N of $(\Omega, \mathcal{F})$ into $(M_p, \mathcal{M}_p)$, where $M_p = \{\mu \in M_E : \mu(A) \in \mathbb{N}^+ \text{ for all} A \in \beta\}$, $\mathcal{M}_p = \mathcal{M} \cap M_p$ and $\mathbb{N}^+$ is non-negative nature numbers.

In this paper, we focus on the case where the space E is actually a time space, hence loosely speaking, we can define the point process in a counting process way:

$N_t = \sum_{i=0}^{\infty} \mathbb{I}(t_i \le t)$

Moreover, throughout the whole paper unless other mentioned, we restrict the point process to be simple, i.e., $Pr\{N(\{[t, t + \Delta t]\}) = 0 \text{ or } 1 \text{ for all t}\} = 1$, that is no common jumps are allowed.

Remark With every sample point $\omega \in \Omega$, we associate a particular realization that is a boundedly finite Borel measure on E: it may denoted by $N(\cdot, \omega)$. While for each fixed set A, $N(A, \cdot)$ is a non-negative random variable. In practice, the latter means if we fix a time period $A = [0, t]$, the count data $Y_t = N_t$ is the number of events happened during this period.

We can easily extend the definition to a marked point process, let $N_t = \sum_{i=0}^{\infty} \mathbb{I}(t_i \le t)$ be a point process on E and let $E^{'}$ be a second LCCB. We define the marked point process with underlying process N and realization marks $\{x_i\}_i$ as any point process that,

$$N_t(E^{'}) = \sum_{i=0}^{\infty} \mathbb{I}(t_i \le t, X_i \in E^{'})$$

on $E \times E^{'}$. The random element X_i of $E^{'}$ is called the mark associated with t_i.

1.1.1 Intensity

The intensity λ of a counting process is a measure of the rate of change of its predictable part. Conditional on a time dependent filtration $\mathcal{F}_{t-}$, the intensity is defined as:

$$\lambda(t|\mathcal{F}_{t-}) = \lim_{\Delta t \to 0} \frac{\mathbb{E}(N([t, t + \Delta t])|\mathcal{F}_{t-})}{\Delta t} \tag{1.3}$$

In a similar way, we define the intensity for a marked counting process as:

$$\lambda(t, x|\mathcal{F}_{t-}) = \lim_{\Delta t \to 0, \Delta x \to 0} \frac{\mathbb{E}(N([t, t + \Delta t], [x, x + \Delta x])|\mathcal{F}_{t-})}{\Delta t \Delta x} \tag{1.4}$$

By construction, the counting process $N(t)$ is a sub-martingale, therefore the intensity is non-negative.

The cumulative intensity, as its name suggests, is defined as the integral of an intensity over a period of time, say $[0, t]$:

$$\Lambda(t|\mathcal{F}_{t-}) = \int_0^t \lambda(s|\mathcal{F}_{s-})ds \tag{1.5}$$

The self-exciting process is characterized by its filtration: if the filtration is generated by the counting process itself: $\mathcal{F}_t = \sigma(N(s) : s \leq t)$, we call this counting process a self-exciting process. We dedicate the next subsection to discuss the issue of filtration. Here, to streamline the illustration, we suppress reference $\mathcal{F}_{t-}$ in λ and Λ .

The cumulative intensity and the counting process are connected by the well-known

Doob-Meyer decomposition theorem:

$$N(t) = \Lambda(t) + M(t) \tag{1.6}$$

This theorem states that any counting process can be decomposed as the sum of a cumulative intensity (also known as compensator in some literature) and a martingale $M(t)$. Moreover, the cumulative intensity is predictable and unique. By the martingale property we have,

$$\mathbb{E}(N(t)) = \mathbb{E}\Lambda(t) \tag{1.7}$$

We can interpret the cumulative intensity as the mean of the underlying counting process at $t, \forall t$. From an economic perspective, we may say that the cumulative intensity summarizes all the systematic parts of a counting process, while the martingale accounts for the stochastic part. Notice that the cumulative intensity $\Lambda(t)$ may not be absolutely continuous. One common assumption regarding this (e.g.,Kopperschmidt & Stute (2013)) is to let the process $t \to \Lambda(t)$ be almost surely continuous, while allows unexpected jumps in the intensity function $\lambda(t)$.

The intensity also connects to the probability density of the underlying counting process. Let $U_{n+1} = T_{n+1} - T_n$ be the duration between n^{th} and $n+1^{th}$ arrivals, for each arrival n, let $F_n(du) = Pr\{U_{n+1} \in du\}$ then

$$\Lambda(t) = \Lambda(T_n) + \int_0^{t-T_n} \frac{F_n(dx)}{F_n[x, \infty)}, t \in (T_n, T_{n+1}]$$

where T_i is stopping time [2]. The proof can be found in Karr (1991).

[2] Appendix B contains an introduction to the concept of stopping time.

1.1.2 Filtration and Marks

As mentioned before, in the case of self-exciting, the filtration is generated by the marked counting process itself $\mathcal{F}_t = \sigma(N(s, x_s) : s \leq t, x_s \in R^+)$. Thus, the history, such as the timings of occurrences and their marks, is contained in this filtration. However, the order of timings and marks are different. This difference is crucial to properly understand the filtration $\mathcal{F}_{t-}$, usually referred as the strict history. Intuitively, we will never know the values of marks without the occurrences of events. This suggests that marks are adjuncts to the occurrences.

Notice that the compensator is conditioning on a filtration $\mathcal{F}_{t-}$ interpreted as *strict past* or *strict history*. It is natural to ask 'how to define history' and 'what is filtration'? It turns out that one can use filtration to define history, hence next we are going to carefully define filtration and discuss its structure.

Let the marked point process N_t be defined as before on $\mathbb{N}^+$ with mark space (E', ε), for $B \in \varepsilon$, let $N_t(B) = \sum_{i=1}^{\infty} \mathbb{I}(t_i \leq t, x_i \in B)$. Define the filtration as,

$$\mathcal{F}_t = \sigma(N_s(B) : 0 \leq s \leq t, B \in \varepsilon) \tag{1.8}$$

We term such filtration as the internal history of N. Notice that the filtration defined in this way is right continuous in a sense that for each t, $\mathcal{F}_t = \cap_{h>0} \mathcal{F}_{t+h}$.

To describe the structure of the filtration, we need two more concepts, stopping times and predictability. The former one is a necessary tool to describe a counting process and define history, while the latter is the key ingredient of the Doob-Meyer decomposition, upon which the minimal distance estimation method is based.

Definition A random variable T with values in $[0, \infty]$ is a stopping time if $\{T \leq t\} \in \mathcal{F}_t$ for every t. Intuitively, time T of a random event is a stopping time if at each fixed time t, one can observe whether or not the event has already occurred.

Associated with the stopping time T are the $\sigma-$ algebra,

$$\mathcal{F}_T = \sigma\{\Lambda \in \mathcal{F}_\infty : \Lambda \cap \{T \leq t\} \in \mathcal{F}_t \text{ for all t}\}$$

which sometimes be called 'the past before T' and the $\sigma-$ algebra

$$\mathcal{F}_{T-} = \sigma\{\Lambda \cap \{T > t\} : t \geq 0, \Lambda \in \mathcal{F}_t\} \vee \mathcal{F}_0$$

which comprises the strict past of $\bar{N}$

In the above two equations, $\mathcal{F}_0$ is the filtration that stores some foreknowledge of the marked point process, that is,

$$\mathcal{F}_0 = \liminf_{t>0} \mathcal{F}_t$$

And the the filtration $\mathcal{F}_\infty$ contains the common information among all the filtrations,

$$\mathcal{F}_\infty = \cap_{t>0} \mathcal{F}_t$$

Next, we study the predictability, a key ingredient of Doob-Meyer decomposition. For simplicity, we restrict ourselves to finitely many t's, say $0 = t_0 < t_1 < t_2 < \cdots < t_k$. Put $N_i = N_{t_i}$ for short. At time t_i, only $N_0, N_1, \cdots, N_i$ are known, but not necessarily $N_{i+1}, N_{i+2}, \cdots, N_k$. We are concerned about how to predict future values of N_j given the information at $t = t_i$. For a variable Λ_{i+1} to be a predictor for N_{i+1} at t_i means this Λ_{i+1} needs to be known at t_i. Thus

Definition Let $(\mathcal{F}_i)_{0\leq i\leq k}$ be some increasing filtration, and let $(\Lambda_i)_{0\leq i\leq k}$ be a sequence of random variables. Then we call

$$(\Lambda_i)_i \textbf{ predictable w.r.t}(\mathcal{F}_i)_i \text{ iff } \Lambda_{i+1} \textbf{ is } \mathcal{F}_i \textbf{ -measurable}$$

Now, we can describe the structure of the filtrations $\mathcal{F} = (\mathcal{F}_t)$ as follow:

Proposition For each n, if T_n is a stopping time of $\mathcal{F}$ and $A_t(E')$ is a predictable process, then we have,

$$\mathcal{F}_{T_n} = \sigma((T_1, X_1), \cdots, (T_n, X_n))$$
$$\mathcal{F}_{T_n-} = \sigma((T_1, X_1), \cdots, (T_{n-1}, X_{n-1}), T_n)$$

The proof can be found in Karr (1991) Karr (1991). Intuitively, the accumulating information changes only at the arrival times T_i, whereas $T_i \in \mathcal{F}_{T_i-}$, the mark X_i belongs to $\mathcal{F}_{T_i}$ but not $\mathcal{F}_{T_i-}$ and finally, X_i is the only information in $\mathcal{F}_{T_i}$ not contained in $\mathcal{F}_{T_i-}$.

Once we understand the term 'strict past' and 'filtration', we can re-define the compensator in a more rigorous way.

Definition Assume that $\mathbb{E}(N_t) < \infty$ for every t, then the compensator of a standard point process (absent of marks) N respect a whole history $\mathcal{H}$ such that $\mathcal{F}_t \in \mathcal{H}_t$ for each t is the unique random measure Λ on $\mathbb{R}_+$ such that

- The process (Λ_t) is $\mathcal{H}$-predictable

- For every non-negative $\mathcal{H}$-predictable process C:

$$\mathbb{E}[\int_0^\infty CdN] = \mathbb{E}[\int_0^\infty Cd\Lambda]$$

A similar definition can be applied to the marked point process after adding some appropriated mark spaces.

In many economic studies it is of interest to introduce some external information. In the example of health insurance, we may investigate the impact of income or education on the usage of medical services. We can, in fact, enrich this filtration to include these covariates. Let $\mathcal{H}_{t-} = \mathcal{H}_0 \vee \mathcal{F}_{t-}$ be the conditioned filtration, where $\mathcal{H}_0$ is the $\sigma - algebra$ generated by some external covariates, such as age, sex, race, income, etc. We interpret this filtration as the 'whole history'. Notice that $\mathcal{H}_0$ can also be time-dependent, i.e., $\mathcal{H}_0 = \mathcal{H}_0(t-)$.

The fact that marks are adjuncts to occurrence times inspires us to separate timings and marks. Intuitively we may re-write the intensity for the marked counting process as:

$$\begin{aligned}
\lambda(t, x|\mathcal{H}_{t-}) &= \lim_{|\Delta t \Delta x| \to 0} \frac{Pr\Big(t < T \leq t + \Delta t, x < \mathbf{X} \leq x + \Delta x | \mathcal{H}_{t-}\Big)}{\Delta t \Delta x} \\
&= \lim_{\Delta t \to 0} \frac{Pr\Big(t < T \leq t + \Delta t | \mathcal{H}_{t-}\Big)}{\Delta t} f(x|\mathcal{H}_{t-}, t) \\
&= \lambda_g(t|\mathcal{H}_{t-}) f(x|t, \mathcal{H}_{t-})
\end{aligned} \tag{1.9}$$

where we call $\lambda_g(t|\mathcal{H}_{t-})$ the ground intensity, and it happens to be the intensity of the original marked counting process if marks are ignored, or the ground counting process:

$$N_g(t) = \sum_i \mathbb{I}\{T_i \leq t\} \tag{1.10}$$

The second part $f(x|t, \mathcal{H}_{t-})$ is called the conditioned mark density since it is not only conditioned on the filtration, but also conditioned on the occurrence time t. In appendix C, we use the Janossy measurement language to discuss this separation in a more rigorous way.

1.2 Estimation and Simulation

Since one can separate the marked intensity into a multiplicative form of ground intensity and conditional mark density, it would be straightforward to estimate these two parts separately. In this study, we mainly focus on the estimation of the ground intensity.

In the counting process literature, likelihood based methods are the most commonly used estimation tools, (e.g.,Ogata & Katsura (1988),Zhuang et al. (2002),Aït-Sahalia et al. (2015),Bacry & Muzy (2014) and Mohler et al. (2012)). One requirement of using them is the predictability of the cumulative intensity Λ with respect to the filtration $\sigma(N_g(s) : s \leq t)$. That is, conditional on the filtration, the values of all the explanatory variables at time t should be known and observed just before t. However, as pointed out by Kopperschmidt & Stute (2013), in many complicated economic situations, there is little reason to maintain such an assumption. Instead, the cumulative intensity should respect external shocks or impulses. In that case, the model is most likely not dominated and the likelihood methods are difficult to apply.

In our application, a core task is to update the cumulative individual cost whenever an event occurs. Two sources of cost are considered, the first one comes from the main counting process $N^1(t)$ in which an event is a doctor visit and a mark is the associated individual cost. Another one is the drug purchase cost, represented by a mark linked to a drug purchase counting process $N^2(t)$. As a result, the individual cost coming from the drug purchase serves as an external shock to the main counting process. More precisely, the conditional filtration $\mathcal{H}_{t-}$ in our model is generated not only by the main counting process, but also by the external drug purchase counting process, i.e., $\mathcal{H}_{t-} = \mathcal{H}_0 \vee \mathcal{F}_{t-} \vee \mathcal{G}_{t-}$, where $\mathcal{F}_t = \sigma(N^1(s) : s \leq t), \mathcal{G}_{t-} = \sigma(N^2(s) : s \leq t)$.

Figure 1.2 helps to understand. Here t_i are occurrence times of interested events (in

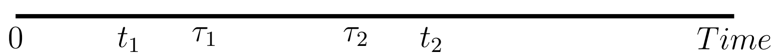


Figure 1.2: A possible realization of event occurrences

our medical utilization application, they are doctor-visiting times) and τ_i are secondary event occurrence times (in our medical utilization application, they are drug purchase times). The interested intensity λ is not predictable with respect to the filtration $\mathcal{F}$ generated only by N^1 since the cumulative individual cost is updated due to drug purchase events. But λ is predictable with respect to $\mathcal{H}$.

1.2.1 Minimum Distance Estimation

To overcome this problem,Kopperschmidt & Stute (2013) develop a parametric minimum distance estimation method. The basic idea consists of using the Doob-Meyer decomposition to minimize the distance between the counting process and its cumulative intensity. This method only requires the observations to be i.i.d. It does not assume the differentiability of the cumulative intensity and allows unexpected jumps in the intensity function.

Formally, let $v_0 \in \Theta \subset \mathbb{R}^d$ be the true parameters, and let $N_{g,1}, ..., N_{g,n}$ be i.i.d copies of n observed ground counting process. For each $1 \leq i \leq n$, let $\mathcal{H}_i(t)$ be an increasing filtration comprising the relevant information about the marked counting process N_i as well as some other external information. Let $\Lambda_{g,v,i}$ with $v \in \Theta \subset \mathbb{R}^d$ be a given class of parametric cumulative ground intensities. Let the true one be $\Lambda_{g,i} = \Lambda_{g,v_0,i}$.

Let,

$$\bar{N}_{g,n} = \frac{1}{n}\sum_{i=1}^{n} N_{g,i}; \bar{\Lambda}_{g,v,n} = \frac{1}{n}\sum_{i=1}^{n} \Lambda_{g,v,i} \tag{1.11}$$

We call the former averaged (ground) point process and the latter averaged cumulative (ground) intensity. Naturally the associated averaged innovation martingale is,

$$d\bar{M}_{g,n} = d\bar{N}_{g,n} - d\bar{\Lambda}_{g,v_0,n} \tag{1.12}$$

The optimization object is:

$$||\bar{N}_{g,n} - \bar{\Lambda}_{g,v,n}||_{\bar{N}_{g,n}}, \tag{1.13}$$

where

$$||f||_\mu = [\textstyle\int_0^T f^2 d\mu]^{1/2}$$

T is a terminating time. This statistic 1.13 is an overall measurement of fitness of $\bar{\Lambda}_{g,v,n}$ to $\bar{N}_{g,n}$. The estimator v_n is computed as,

$$v_n = arg \inf_{v \in \Theta} ||\bar{N}_{g,n} - \bar{\Lambda}_{g,v,n}||_{\bar{N}_{g,n}} \tag{1.14}$$

Kopperschmidt & Stute (2013) show this estimator is consistency and its asymptotic behaviour is

$$\sqrt{n}\Phi_0(v_0)(v_n - v_0) \to N_d(0, C(v_0)) \tag{1.15}$$

where

$$\Phi_0(v) = \frac{\partial}{\partial v}\int_E (\mathbb{E}\Lambda_{g,v}(t) - \mathbb{E}\Lambda_{g,v_0}(t))\mathbb{E}\frac{\partial}{\partial v}\Lambda_{g,v}(t)^T \mathbb{E}\Lambda_{g,v_0}(dt) \tag{1.16}$$

$C(v_0)$ is a $d \times d$ matrix with entries

$$C_{ij}(v_0) = \int_E \phi_i(x)\phi_j(x)\mathbb{E}\Lambda_{g,v_0}(dx) \tag{1.17}$$

and

$$\phi_i(x) = \int_{[x,\bar{t}]} \mathbb{E}\frac{\partial}{\partial v_i}\Lambda_{g,v}(t)\mathbb{E}\Lambda_{g,v_0}(dt) \mid_{v=v_0}, t \leq x \leq \bar{t} \tag{1.18}$$

Remark Let Φ_n be the empirical analog of Φ_0,

$$\Phi_n(v) = \frac{\partial}{\partial v}\int_E (\bar{\Lambda}_{g,v,n}(t) - \bar{\Lambda}_{g,v_0,n}(t))\frac{\partial}{\partial v}\bar{\Lambda}_{g,v,n}(t)^T \bar{\Lambda}_{g,v_0,n}(dt) \tag{1.19}$$

Since all $\bar{\Lambda}_{g,v,n}$ are sample means of i.i.d non-decreasing processes, a Glivenko-Cantelli argument yields, with probability one, uniform convergence of $\bar{\Lambda}_{g,v,n} \to \mathbb{E}\Lambda_{g,v}(t)$ in each t on compact subsets of Θ, we have the expansion,

$$\Phi_n(v) = \Phi_0(v) + op(1) \tag{1.20}$$

Such expansion guarantees that in a finite sample situation, we can replace the unknown matrix $\Phi_0(v_0)$ by $\Phi_n(v_n)$ and $C(v_0)$ by $C^n(v_n)$ without destroying the distributional approximation through $N_d(0, C(v_0))$, where C^n is the sample analog of C. In practice, one needs to plug and replace the true ones with estimators and replace $\mathbb{E}\Lambda_{g,v_0}(dt)$ with its empirical counterpart $\bar{N}_g(dt)$.

1.2.2 Simulation Study

In the original Kopperschmidt & Stute (2013) paper, the authors do not provide a numerical simulation study. Here we provide finite sample evidence by generating a self-exciting process and examining the performance of the minimum distance estimator.

The data generating process we picked is the ETAS (epidemic type aftershock sequence) model. It was first introduced by Ogata & Katsura (1988) and ever since has been widely used in seismology (e.g. Zhuang et al. (2002)). It characterizes earthquake times and magnitudes and belongs to a marked Hawkes process family. The ETAS model has the probabilistic structure we desire: marks are part of the ground intensity and can be separated into ground intensity and conditioned mark density.

The ground intensity of a ETAS model, for its simplest form, could be:

$$\lambda_g(t|\mathcal{F}_{t-}) = \mu + \sum_{i:t_i<t} e^{\alpha x_i} \left(1 + \frac{t - t_i}{c}\right)^{-1} \tag{1.21}$$

where x_i is the magnitude of an earthquake occurring at time t_i, and the mark density, for simplicity, is assumed to be independent:

$$f(x|t, \mathcal{F}_{t-}) = \delta e^{-\delta x} \tag{1.22}$$

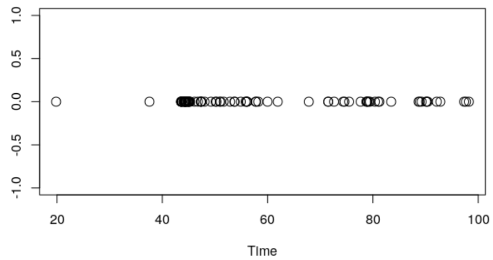
In this simulation study, we focus on the ground intensity since the mark density is the well-known exponential density whose best estimator is trivial: the inverse of sample mean.

We set the true parameters as $\mu = 0.007$, $\alpha = 1.98$, $c = 0.008$ and $\delta = log(10)$. The simulation method we used is called the *thinning method*, introduced by Ogata (1981), Lewis & Shedler (1979). Briefly, this method first calculates an upper bound for the intensity function in a small time interval, simulating a value for the time to the next possible event using this upper bound, and then calculating the intensity at this simulated point. However these 'events' are known to be simulated too frequently (Lewis & Shedler 1979). To overcome this, the method will compare the ratio of the calculated rate with the upper bound to a uniform random number to randomly determine whether the simulated time is treated as an event or not (i.e. thinning). A full description of the algorithm is provided in Appendix D.

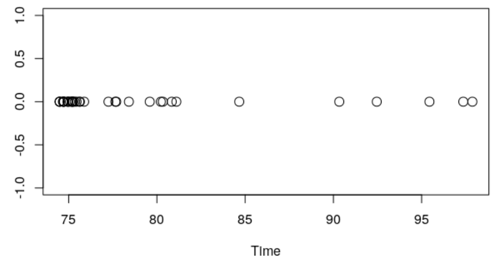
We generate $N = 50, N = 100$ and $N = 200$ individual counting processes for each simulation and in total we have $B = 1000$ replications. The time-intervals are set to be $[0, 3000], [0, 500]$ and $[0, 100]$.

Figure 1.3 presents three quite different individual's event histories simulated by this ETAS DGP using the identical parameter settings as stated before.

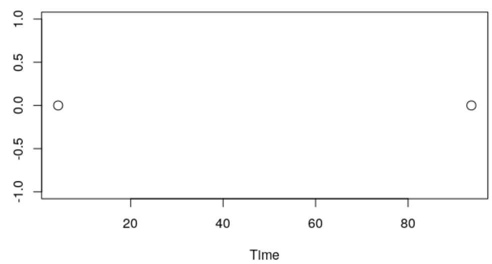
Individual 1 has the most frequent events experience, the total number of events is 92. Individual 2 is somewhat moderate, with 37 events. Individual 3 has the least frequent events with only 2 during the time interval $[0, 100]$. Despite the hugely different behaviors, they are actually governed by the same intensity function. This example demonstrates that a self-exciting process can generate enough heterogeneity without



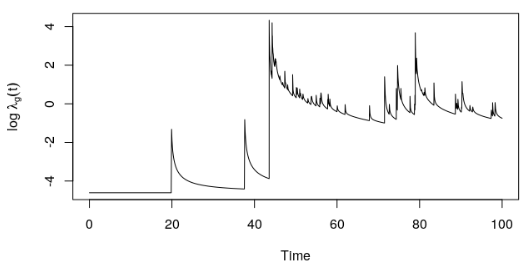
(a) Individual 1, Event Time



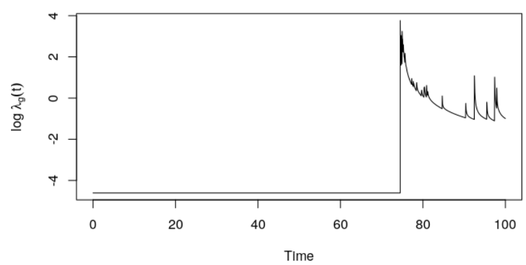
(b) Individual 2, Event Time



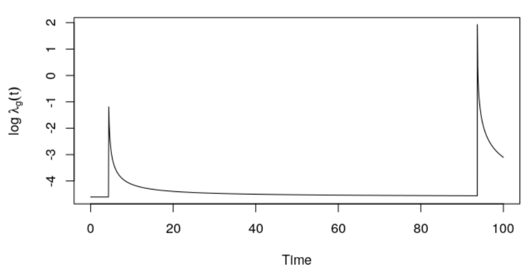
(c) Individual 3, Event Time



(d) Individual 1, log of intenstiy



(e) Individual 2, log of intenstiy



(f) Individual 3, log of intenstiy

Figure 1.3: Three individual's events histories

introducing a latent variable to represent the unobserved heterogeneity.

The estimation results are presented in Tables 1.1-1.3. As the number of observations N increase, the estimators become more stable and their empirical coverage rate gets closer to the theoretical ones. It is also noticeable that the performance of estimators is closely related to the number of events per person. (We increase the length of the time horizon to increase such a number under the same true parameters.)

Table 1.1:: Minimum Distance Estimator Results, with $T = 3000$

$N = 400$	True	MDE	sd	se	CI95	CI90
μ	0.007	0.006957441	0.0005575073	0.0006271522	94.9%	92. 5%
α	1.98	1.978269	0.04350423	0.07331051	93.5%	90.8%
c	0.008	0.008130796	0.001105742	0.001724244	93.9%	91.3%
Distance	1.48622	0.715594				
$N = 200$						
μ	0.007	0.006960397	0.0008548692	0.0008400477	93.6%	88.3%
α	1.98	1.984108	0.08547141	0.1086315	93.2%	90.9%
c	0.008	0.008042743	0.001568495	0.002413533	92.6%	90.3%
Distance	2.183783	1.226474				
$N = 100$						
μ	0.007	0.00684719	0.001007428	0.0011465	93.4%	90.9%
α	1.98	1.964071	0.06827856	0.1654297	92.1%	90.1%
c	0.008	0.008570634	0.001945219	0.003605053	92.3%	90.5%
Distance	3.169824	2.388298				
$N = 50$						
μ	0.007	0.006809876	0.001673515	0.001541488	89.1%	84.9%
α	1.98	1.974604	0.1713249	0.2765146	87.9%	83.7%
c	0.008	0.008979804	0.004142383	0.005475683	86.9%	83.1%
Distance	4.293006	3.474963				

Note: The distance is calculated using the semi-norm **??** with true parameters and the minimum distance estimators, respectively. sd is the standard deviation generated by the Monte Carlo simulation estimates. se is the mean of the standard error of each simulation. CI95(CI90) is the percentage of the 95%(90%) confidence interval generated by se that covers the true parameter.

Table 1.2:: Minimum Distance Estimator Results, with $T = 500$

$N = 400$	True	MDE	sd	se	CI95	CI90
μ	0.007	0.006818704	0.001200632	0.001282338	95.3%	93%
α	1.98	1.985548	0.05759593	0.2580961	96.3%	93.7%
c	0.008	0.008327916	0.002161951	0.005313165	96%	92%
Distance	0.2336264	0.1619424				
$N = 200$						
μ	0.007	0.007056179	0.001633427	0.001783448	92.5%	89.6%
α	1.98	1.977045	0.1709611	0.4486648	91.9%	90.6%
c	0.008	0.009058691	0.004623662	0.008174076	91.5%	89.9%
Distance	0.3579916	0.2119269				
$N = 100$						
μ	0.007	0.00660844	0.003243934	0.002295691	90.1%	86.1%
α	1.98	1.76104	0.4963242	0.85060127	86.6%	83%
c	0.008	0.01662388	0.0151135	0.0174853	86.7%	83.7%
Distance	0.477976	0.4551476				
$N = 50$						
μ	0.007	0.006672302	0.005083828	0.002964079	90.3%	88%
α	1.98	1.761366	0.5886596	2.207844	91.4%	88.9%
c	0.008	0.01808354	0.01897789	0.02508167	90.6%	87.8%
Distance	0.6452985	1.087129				

Note: The distance is calculated using the semi-norm **??** with true parameters and the minimum distance estimators respectively. sd is the standard deviation generated by the Monte Carlo simulation estimates. se is the mean of the standard error of each simulation. CI95(CI90) is the percentage of the 95%(90%) confidence interval generated by se that covers the true parameter.

Table 1.3:: Minimum Distance Estimator Results, with $T = 100$

$N = 400$	True	MDE	sd	se	CI95	CI90
μ	0.007	0.0067466	0.001766587	0.002320197	95.2%	92.9%
α	1.98	1.980313	0.2536825	1.687546	95.1%	94%
c	0.008	0.01027362	0.007087134	0.01646008	95.4%	93.9%
Distance	0.0365799	0.0204637				
$N = 200$						
μ	0.007	0.006614259	0.003093314	0.002845468	93.6%	90.6%
α	1.98	1.91999	0.434588	2.273823	94.5%	93.3%
c	0.008	0.01357907	0.01251991	0.02549106	93.2%	92.1%
Distance	0.05125482	0.03664879				
$N = 100$						
μ	0.007	0.01317505	0.01093067	0.005716749	81.5%	75.7%
α	1.98	1.719879	0.7325846	2.227818	92.2%	89.6%
c	0.008	0.02089188	0.02165684	0.03664059	89%	86.9%
Distance	0.6294044	0.1808156				
$N = 50$						
μ	0.007	0.01273163	0.007051629	0.006974369	85.9%	82.9%
α	1.98	1.87436	0.8308396	3.961052	95.6%	93.5%
c	0.008	0.02130184	0.02805238	0.04548218	89.2%	87.2%
Distance	0.639077	0.1674467				

Note: The distance is calculated using the semi-norm **??** with true parameters and the minimum distance estimators respectively. sd is the standard deviation generated by the Monte Carlo simulation estimates. se is the mean of the standard error of each simulation. CI95(CI90) is the percentage of the 95%(90%) confidence interval generated by se that covers the true parameter.

1.3 Self-Exciting Process as a Complementary Tool to Conventional Methods

In this section, we compare the differences between the self-exciting process and two widely used conventional econometric tools in microdata analysis: count data regression and duration analysis. We argue that many major issues in these two conventional tools can be easily overcome by using a self-exciting process. We also highlight the fact that despite the numerous advantages of using self-exciting process, it can not replace conventional methods completely. Researchers should adopt proper econometric tools to their specific needs.

1.3.1 Compare to the Count Data Regression

Many count data display over-dispersion property: the variance of data exceeds the mean of data. One source of such over-dispersion is excess zeros: the dataset may have more zero observations than is consistent with the basic Poisson model.

Unlike the count data regression, where the discrete counts y is treated as a random variable, in a self-exciting process, the outcome is a time depended counting process $N(t)$. The additional time dimension enables us to generate excess zeros. The intuition of our argument is quite simple: if the terminated time is small (relative to the intensity), we can easily generate a high proportion of zeros. More precisely, we treat the zero event as an end-of-study censoring problem: events will happen in the future, but they are censored due to an end of the study. Also in the generalised count data models (e.g. Zero inflation and Hurdle), zeros and non-zeros (positives) are assumed to come from two different data generating processes (DGPs). Whereas in the self-exciting process, zeros and positives are generated from the same stochastic process.

We use two DGPs to illustrate our argument. The first is the standard Poisson process and the second is a self-exciting process. The Poisson process serves as our baseline model (same as the Poisson regression in count data). Simulations will show that although by setting a small time interval, we can generate a high proportion of zeros, the Poisson process is still equidispersion. The self-exciting process, on the other hand, can mimic the over-dispersion property of data through excess zeros.

Poisson DGP The intensity for a (homogeneous) Poisson process is a constant $\lambda = \mu$. We set $\mu = 5.5$, and let the time interval to be $[0, T^*] = [0, 0.2]$. We run 100 trials of simulation. For each Poisson process, we record its corresponding counts: $Y_i = N_i(T^*), i = 1, 2, \cdots, 100$. The following histogram displays our simulation results.

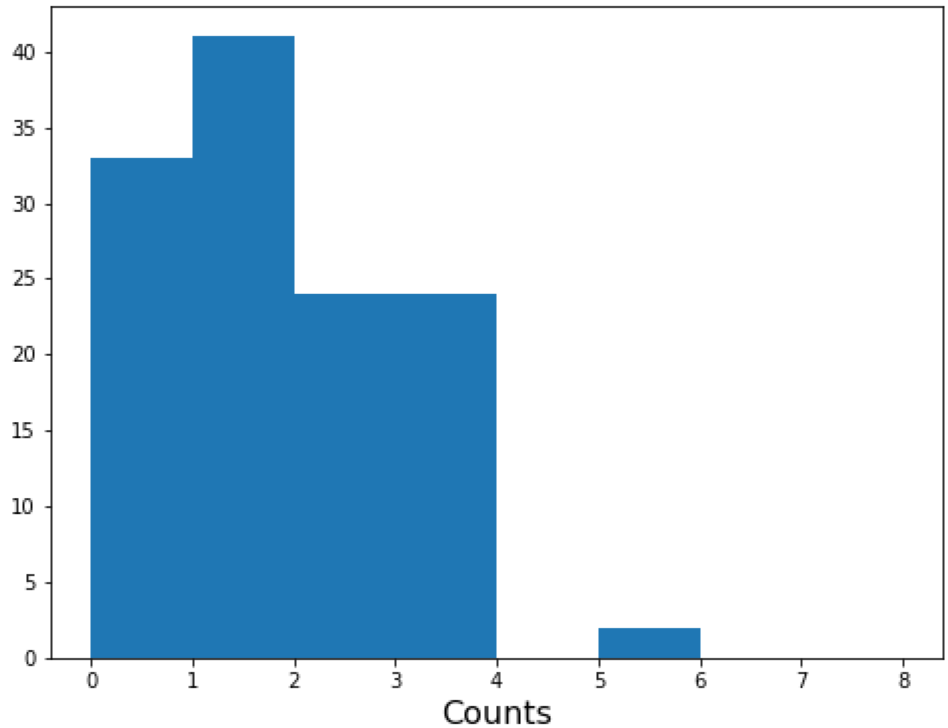


Figure 1.4: Result of Poisson Process Simulation

Self-Exciting DGP The DGP for the self-exciting process is the same as in the simulation study in previous section. Like before, we run 100 simulations and record their corresponding counts at the terminal time T^*. With these parameters, we can generate 44 zero observations out of 100. The largest count is at 92. We plot its histogram as below.

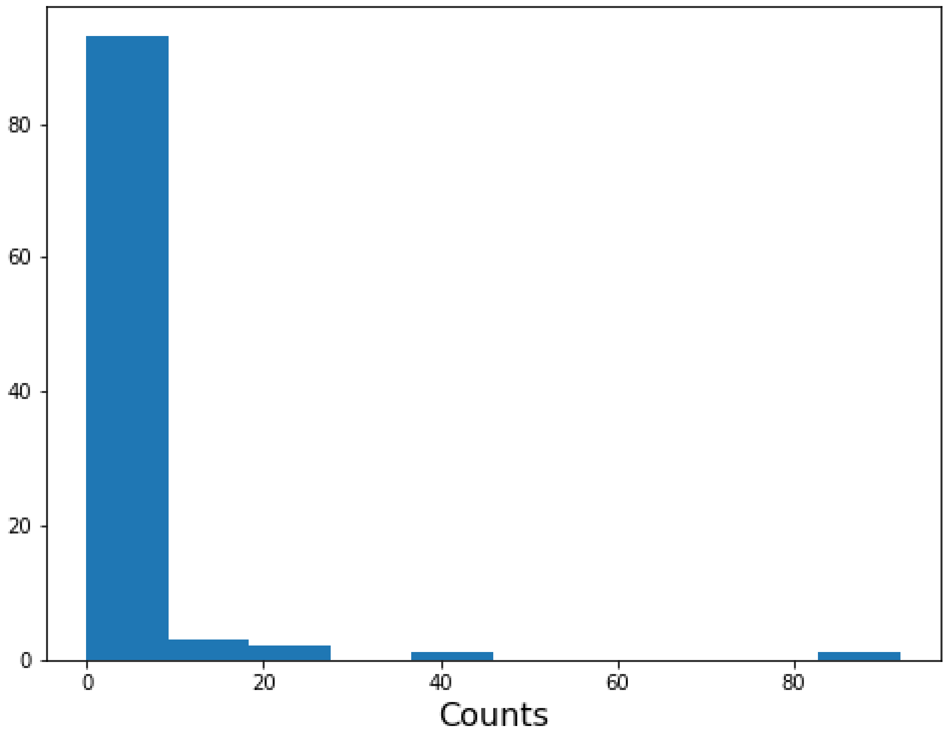


Figure 1.5: Result of Self-Exciting Process Simulation

The self-exciting data exhibits the over-dispersion property: $\bar{Y} = 3.27$, $\hat{V}(Y) = 108.5425$.

1.3.2 Compare to the Duration Analysis

Unlike the counting process where the interested subject is the time stamps of events (by modelling the intensity function), in duration analysis, the subject under investigation is the duration of a default state (by modelling the hazard rate). The intensity function and hazard rate are, in some sense, quite similar but conceptually different.

Consider a self-exciting process, let τ be the time of the last event before time t and

$\mathcal{F}$ be the filtration. Denote the conditional distribution of the time of the next event as:

$$G(t|\mathcal{F}(\tau)) = Pr(T \geq t|\mathcal{F}(\tau))$$

and $g(t|\mathcal{F}(\tau))$ as the corresponding conditional density function. Then from the definition of intensity (equation (**??**)),

$$\lambda(t|\mathcal{F}(\tau)) = \frac{g(t|\mathcal{F}(\tau))}{1 - G(t|\mathcal{F}(\tau))} \tag{1.23}$$

Now, consider a system begins in time 0 and fails at some random time $T > 0$. The hazard rate (or hazard function) $h(t)$ is defined as:

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{Pr\{T \in (t, t + \Delta t)\}}{Pr\{T > t\}\Delta t} \\ &= \frac{f_T(t)}{1 - F_T(t)} \end{aligned} \tag{1.24}$$

Where t here is the duration of a state. The hazard rate tells us the conditional probability of the system failing in the interval $(t, t + \Delta t]$ conditioned on the system is working at time t.

Despite the similarity between (1.23) and (1.24), the intensity and the hazard rate are conceptually different. Intensity deals with recurrent arrivals with a focus on the timing per se, while the hazard rate deals with the duration or the length of only one spell. Most duration analysis can only study the recurrent events with the i.i.d of events assumption holds. It is difficult to employ this method when recurrent events are state dependent. A self-exciting process, on the other hand, is free from these problems since the state dependence is included in the filtration.

Chapter 2

The Cost-Sharing, Shadow Price and Cluster in Medical Care Utilization

Abstract

In this chapter, a self-exciting counting process modelling method is proposed to study the frequency of medical care service utilization under a non-linear budget constraint health insurance policy. This modelling strategy enables researchers to investigate individual's dynamic behavior in a more detailed way. Specifically, for each individual, every doctor visiting record is represented as a point in a self-exciting counting process. Cost associated with such visiting is included in this counting process as a mark. A minimum distance method is employed to find the estimators. Using the Rand Health Insurance Experiment data, we find that individuals respond to a change of shadow price. In addition, we use a matured cluster analysis algorithm to investigate the cluster patterns and discover that compared to free plan, cost-sharing insurance plan with

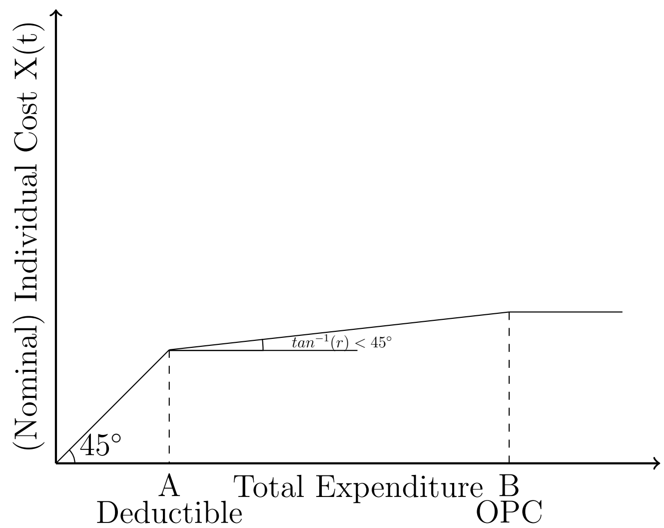
out-of-pocket fees suppress the use of medical services by limiting the number of clusters as well as follow-up visiting within each cluster.

2.1 Introduction

In this paper, we aim to model medical utilization (outpatient only) under a non-linear cost environment using a self-exciting process. Recent studies on health insurance (e.g., Aron-Dine et al. (2012), Einav et al. (2015)) deviate from the classical assumption that individuals only respond to a single linear spot cost[1] and find strong evidence that individuals respond to the dynamic incentives associated with the non-linear nature of a typical health insurance contract. These conclusions suggest that 'it is unlikely that a single elasticity estimate can summarize the spending response to changes in health insurance' and 'such an estimate is not conceptually well defined.' (Aron-Dine et al. 2012).

The driving forces of such a non-linear nature are cost-sharing policies implemented in a health insurance contract. The most common ones are the deductible, the co-insurance rate and the out-of-pocket fee cap (OPC). In a typical setup, individuals need to cover all their medical expenditures below the deductible. Once the threshold is passed, co-insurance is applied, where individuals pay part of the expenditures based on the co-insurance rate. Finally, if the total expenditure paid by the individual passes the OPC, no cost (or very little cost) would be paid by this individual. Figure 2.1 illustrates such a typical non-linear budget constraint.

[1] That is, a linear budget constraint.



The total expenditure is the sum of individual costs and costs paid by the insurance. Points A and B are the deductible threshold and OPC, respectively. When the total expenditure is below A, the co-insurance is 100% (individuals pay all cost) and the slope is 1. Between A and B, a co-insurance rate (the slope) $0 < r < 1$ is applied. Whenever the total expenditure is beyond B, there is no cost for individuals (the slope is 0).

Figure 2.1: Non-linear Individual Cost (Medical Price)

At the heart of this non-linearity is the stochastic cumulative individual cost $X(t)$[2]. Keeler et al. (1977) is the first theoretical paper that studies the consumer's optimal choice under such a non-linear medical price schedule. Using a dynamic programming model, they show that the shadow price of j^{th} episode is a function of demand prior to this episode (hence the cumulative individual cost). One may construct the shadow price (co-insurance rate) as:

$$p^s(t) = 1 - V(X(t))$$

where $0 \leq V(X(t)) \leq 1$ is a bonus that is related to the cumulative individual cost with $V^{'} > 0$. The intuition behind this equation is simple: under the range of deductibles, although individuals need to fully bear the medical cost, each time this person consumes,

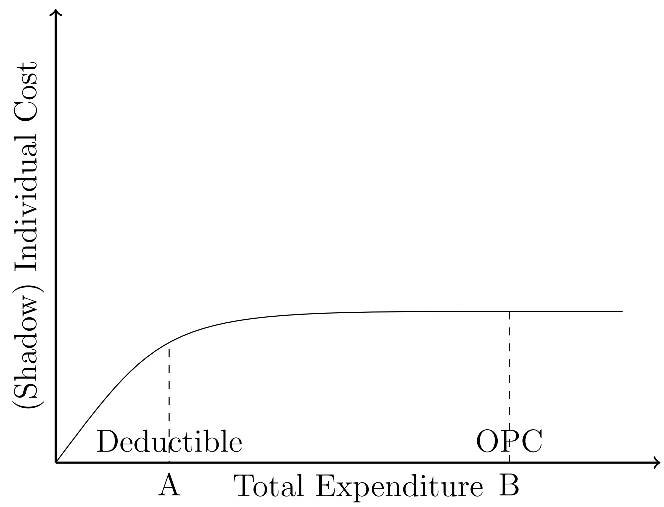
[2]Or equivalently, the total expenditure. Because there is a clear insurance regulation on individual's out-of-pocket cost, cumulative individual cost and total expenditure are one-to-one mapped.

the remaining deductible is reduced and the next instance consumption is more easily to exceed the deductible. As a result, the shadow price for the next purchase is cheaper than the price of the current one (hence the name 'bonus'). Moreover, as the cumulative individual cost gets closer to the deductible, individuals have greater incentive to consume[3]. That is, there should be a positive (negative) relationship between cumulative individual cost $X(t)$ (remaining deductibles) and the probability of medical utilization. For the purpose of self-contain, we review this theory in detail in Appendix A.

The shadow price theory has profound implications on estimating medical demand. First, it suggests one should not use the nominal price. Since the difference between the nominal price and shadow price is not randomly generated, an incorrectly chosen nominal price would lead to a biased estimation. Second, because the shadow co-insurance rate is a function of cumulative individual cost, it implies that individuals will make medical service utilization decisions in a sequential and contingent way. Figure 2.2 illustrates the situation.

Different sources of stochastic disturbances should be distinguished in order to properly model $X(t)$. The first source of randomness is that at any given time $\bar{t}$ within the insurance year, $X(\bar{t})$ is a random variable satisfying $X(\bar{t}) \geq X(s), \forall s \leq \bar{t}$. This non-decreasing random process is difficult to model directly. However, notice that $X(t)$ is a piece-wise constant step function. We may then decompose $X(t)$ as 1) the occurrence time of i^{th} illness episode t_i (the position of i^{th} jump in this step function) and 2) conditional on the occurrence of i^{th} illness, the individual cost $x(t_i)$ for such illness (the size of i^{th} jump). Thus, we could represent the cumulative individual cost as

[3] We assume that medical service is a normal good



Points A and B are the deductible threshold and OPC, respectively. When the total expenditure is below B, the co-insurance rate (the slope) $0 < r(X) < 1$ is a function of cumulative individual cost with $r^{'} < 0$. Whenever the total expenditure is beyond B, there is no cost for individuals.

Figure 2.2: Non-linear Individual Shadow Cost (Medical Price)

a compound counting process: $X(t) = \sum_{i=1}^{\infty} x(t_i)\mathbb{I}\{t_i \leq t\}$. This structure suggests that we could model the time t_i and the cost $x(t_i)$ separately. Medical costs are convenient to assume to be i.i.d and their distribution is well approximated by a log-normal. This is known among literatures (Keeler & Rolph 1988; Handel et al. 2015). Thus the key to model $X(t)$ is to model its occurrence times $\{t_i\}_{i \in \mathcal{N}^+}$.

The second source of randomness comes from other contributions to the individual cost. For example, in this paper, we mainly focus on the outpatient medical utilization. But the costs associated with doctor visits are not the only source of individual cost; other sources could be inpatient expenditures and drug purchase costs. These random costs serve as external shocks to our interested outpatient costs.

The primary goal of this paper is to model $X(t)$, especially the occurrence times of illness episodes. Since the stochastic cumulative individual cost can be represented as a compound counting process, we advocate to use the counting process as our analysis workhorse. In particular, a marked self-exciting process, whose filtration is generated by the counting process itself, is employed to contain the historical information. This information includes both the occurrence time and the non-covered individual charge of each illness episode.

We also aim to model the dependence structure among episodes and to study the effects of cost-sharing policies on episode cluster structures. Episodes are often in a form of a cluster, this is known among medical literatures. A typical example is chronic diseases where patients need to receive treatment periodically.

As will be shown later, the self-exciting process can 1) fully account for different sources of stochastic disturbances of cumulative individual cost $X(t)$, 2) capture the essence of the shadow price: the negative relationship between shadow price and medical utilization (or equivalently, the positive relationship between $X(t)$ and medical utilization) and 3) model the dependence structure among episodes. In comparison, conventional methods such as count data regression and duration models are inadequate to deal with the randomness of individual cost $X(t)$ and the episode cluster structure. As both methods assume events to be i.i.d, which excludes the non-linear price system that we aim to address.

We use the Rand Health Insurance Experiment dataset. Besides it is widely used in the health insurance literature, one advantage of this dataset is that it includes a detailed episode-level claim-by-claim data. We can then update $X(t)$ whenever an event or external shock occurs.

We use a minimum distance method to obtain the estimators. This method is first introduced by Kopperschmidt & Stute (2013) and has the advantage to incorporate external shocks.

This paper contributes three strands of literatures. First, we enrich the ever-expanding literatures that aim to study individual response to a non-linear budget constraint. Second, we introduce a new econometric tool that can be applied beyond health insurance studies. Potential applications include but not limited to labour economics (studies of multiple unemployment, work absences), industry organization (sequential entry games) and criminology etc. Last, we provide a simulation study to exam the performance of this new minimum distance estimation method.

The paper is constructed as follow. Section 2.2 provids a brief literature review on non-linear budget constraint problem in health insurance. Section Section 2.3 discuss the issue of heterogeneity and Section 2.4 introduces the dataset. Section 2.5 presents our model, in which the stochastic property of cumulative individual cost, the effect of cost-sharing policy and the dependence structure of episodes are fully considered. Section 2.6 presents estimation results of the model. We also use a mature machine learning algorithm to analyse the cluster structure of episodes in this section. In section 2.7 we discuss the advantages of our modeling strategy over other conventional reduced form or structural form methods. Section 2.8 concludes the paper.

2.2 Literature Review

We briefly review some literatures that try to include the non-linear budget constraint in their models.

Aron-Dine et al. (2012) construct a future price $p^f = 1 - Pr(X(T) \geq \tilde{X})$ in order to reject the null hypothesis that individuals only respond to a single spot price. Here $X(T)$ is the cumulative individual cost on the last day of an insurance contract year, and $\tilde{X}$ is the deductible. They find a negative relationship between the future price and the initial medical use. Notice that in their construction of future price, only $X(T)$ is used, the rest of $X(t), \forall t < T$ is ignored. In principle, one could construct future price as a function of time using the same method: $p^f(t) = 1 - Pr(X(t) \geq \tilde{X})$. But in practice this would lead to a complicated procedure as one needs to use simulated future price to instrument the future price to correct the estimation bias, see Aron-Dine et al. (2012) for details.

Brot-Goldberg et al. (2017) define their shadow expected marginal end-of-year price at month m as a conditional expectation: $p^e_m = \mathbb{E}(r_{EOY}|X(m), Z, H)$ where r_{EOY} is the end-of-year co-insurance rate, Z is a vector of covariates and H is a measurement of health stock. They non-parametrically estimate the probability density function on cells of equivalent consumers using triple $(X(m), Z, H)$. In practice, they only use age as their sole explanatory variable. Their results suggest that shadow price have a limited impact on spending reduction.

Einav et al. (2015) construct and estimate a dynamic economic model to study individual's drug purchase behavior. In each period, the cumulative individual cost is updated by: $X(t) = X(t-1) + x(t)$, where $x(t)$ is the aggregate individual cost in the current period. Thus $X(t)$ here is not 'totally' stochastic: the occurrence time of illness is ignored and $x(t) = \sum_{i:t_i \in \text{current period}} x(t_i)$ is an aggregate random variable. Moreover, one may find difficulties to model the shadow price in a structural model, since the shadow budget constraint (as illustrated in figure 2) is actually unobserved by researchers.

The $X(t)$ is a compound process and is difficult to analysis. Luckily, one can actually identify it with the marked counting process (see Karr (1991)). One advantage to do so is that the intensity of a marked counting process can be separated as a ground intensity and a conditional mark density. This separation allows us to concentrate on the occurrence times of events.

To the best of our knowledge, no literature has ever explicitly taken the sources of stochastic disturbances mentioned before into consideration. In addition, no literature has ever measured the individuals' responds to the shadow price on a episode level. The main contribution of this paper is to fill this gap from a self-exciting process perspective.

2.3 Heterogeneity in a Self-Exciting Process

In a variety of contexts, it is often noticed that individuals who have experienced an event in the past are more likely to experience the event again in the future than are individuals who have not experienced the event (Heckman 1981). One explanation, best known as the unobserved heterogeneity, is that in addition to the observed variables, there are other relevant variables that are unobserved but correlated with the observed ones. Unobserved heterogeneity (UH) is an important issue in health insurance literature. This is because prices are endogenous: they are lower on average for those who tend to have more episodes. Sickly individuals tend to consume more care services and hence are likely to exceed their OPCs. Therefore, it is crucial to separate the sickliness effect from the shadow price effect. In models like count data regression and duration analysis, the most common way to characterize UH is through the random effect: integrate out the UH term to obtain a marginal distribution.

This strategy faces two problems: 1) A lack of economic theory supporting the choice of the unobserved heterogeneity distribution G. In most duration analysis literature, G is chosen to be Gamma, it is more mathematical convenience than anything else, since by doing so, one can have a closed form of the marginal distribution. 2) It is well known among literatures that there is severe damage of misspecification of G. Van den Berg (2001) provides a theoretical example, and Heckman & Singer (1984) use real application data and various G distributions (Normal, Log Normal and Gamma) to demonstrate that the estimation results tend to be unstable.

Another explanation of heterogeneity is related to state dependence (SD). This concept says that past experience has a genuine effect on future events in a sense that an otherwise identical individual who did not experience the event would behave differently in the future. The definition of the self-exciting process naturally includes the idea of state dependence. Thus in this study, we advocate to assume the state dependence as the source of individual heterogeneity. Specific to our application, individuals are assumed to have no knowledge about their health status at first but gradually update their awareness as episodes and medical utilization are experienced. Different past experiences will generate different behaviors even if the underlying intensity is the same. In the next section, we will demonstrate this fact by a simulation.

SD shares many common properties with random effect. For example, in the mixed proportional hazard model, it is well known that unobserved heterogeneity leads to a 'weeding out' or 'sorting' phenomenon (Van den Berg 2001). That is, individuals with the highest values of UH term leave the default state quickest on average, and the individuals who are still in the state tend to have lower values of UH term. In our medical utilization application, the state of interest is the duration of keeping healthy.

A high value of UH term means a bad health condition (or high level of sickness), and on average it means the duration of being health is short. If only the outpatient is interested, this means on average, we would observe more doctor visiting histories among these individuals on a defined time period. Thus, simply by counting the number of past events, we can capture this 'weeding out' phenomenon.

The SD term can be flexibly modelled. In this study, we suggest the form

$$\sum_{i:t_i<t} K(t_i, t) \tag{2.1}$$

and include it in the intensity function.

There are three advantages to do so. First, the SD term is time dependent, which means that it can be updated. Compared to the conventional time-invariant unobserved heterogeneity term, we believe such a modelling method is more realistic in our empirical study.

Second, the choice of $K(t_i, t)$ is flexible and can be consistent with economic theory. For example, we may capture the seasonality effect by setting $K(t_i, t) = \alpha sin(\beta(t - t_i) + \gamma) + \delta$, or in our application, as explained later, we may study the cluster phenomenon of medical care utilization by letting $K(t_i, t) = \mu exp(-\mu(t - t_i)), \mu > 0$.

2.4 The Data

The data we used come from the well-known RAND Health Insurance Experiment (RAND HIE), one of the most important health insurance studies ever conducted. It addressed two key questions in health care financing:

1. How much more medical care will people use if it is provided free of charge?

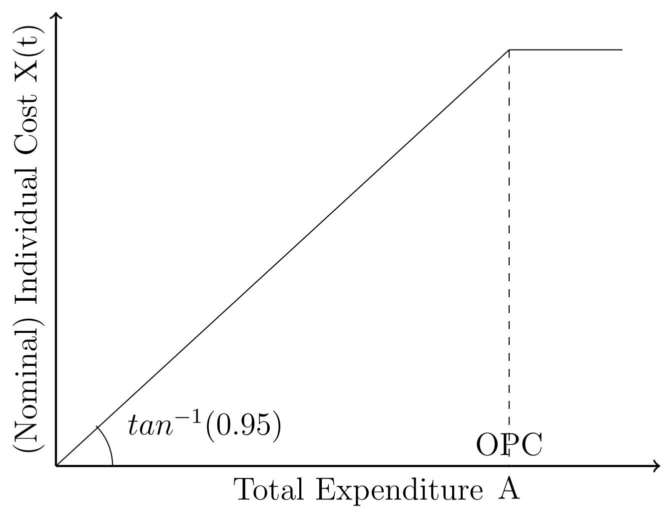
2. What are the consequences for their health?

The HIE project was started in 1971 and was funded by the Department of Health, Education, and Welfare. The company randomly assigned 5809 people to insurance plans that either had no cost-sharing, 25%, 50% or 95% coinsurance rates. The out-of-pocket cap varied among different plans. The HIE was conducted from 1974 to 1982 in six sites across the USA: Dayton, Ohio, Seattle, Washington, Fitchburg-Leominster and Franklin County, Massachusetts, and Charleston and Georgetown County, South Carolina. These sites represent four census regions (Midwest, West, Northeast, and South), as well as urban and rural areas.

Early literatures that use this data usually avoid the problem of non-linear budget constraint by assuming that individuals only respond to one price system. Typical econometric tools involved are the linear regression (after aggregating the data), the count data regression and the duration analysis. None of them is capable to fully model the stochastic structure of cumulative individual cost $X(t)$.

Because the complicated structure of our self-exciting process, to ease the burden of computation, we only use data from Seattle, which has the largest medical claim records available. We separate the data according to two different insurance plans: zero coinsurance rate plan (free plan, denoted as P0), in which the patient does not pay anything; and a cost-sharing plan (denoted as P95) in which a coinsurance rate of 95%

applied and OPC is 150 USD per person or 450 USD per family[4](i.e., before exceeding the OPC, individuals need to pay 95% of the medical care cost, once the OPC is reached, all the cost is paid by the insurance.). The OPC and coinsurance rate in this plan only applied to ambulatory services; inpatient services were free. Both plans covered a wide range of services. Medical expenses included services provided by non-physicians such as chiropractors and optometrists, and prescription drugs and supplies. There is no deductible in this insurance contract. The following figure summarizes the P95 contract design.



Point A is the OPC . When the total expenditure is below A, the co-insurance (slope) $r = 0.95$ is applied. Whenever the total expenditure is beyond A, there is no cost for individuals.

Figure 2.3: Contract design for P95

We also include the data of drug purchase records with information such as the purchase dates and the values of non-covered charges. As discussed in the previous section, we may treat the drug purchase as another counting process and as an external shock to our primary one (doctor-visiting counting process). In the original dataset,

[4]In 1973 dollars.

one individual may have several claims in one day, and we combine all claims with an identical date into one and sum the non-covered charges.

The occurrence time stamp is defined as the annual duration between the beginning date of the insurance policy and the date this person visited a health care institution. For example, if the insurance begins on Jan-01-1977 and the date of a doctor visit is Oct-01-1977, the time stamp is then 0.748 (years). When preparing the dataset, we delete all the records with missing duration information. (Hence we exclude the cases of censoring.)

When analyzing the cost-sharing plan, we restrict our dataset within the contract year 1977-1978 since the cost-sharing policies are renewed annually. But such restriction is not needed for the free plan since there is no within-year cost sharing policy. For this plan, the time horizon ranges from 1975 to 1980. When the individual cost information is missing, we replace it with zero. In the end, we have 243 individuals in the free plan with 7638 claims over the years and 131 individuals in the cost-sharing plan and the total number of claims is 1103 within the 1977-1978 contract year.

We also include some demographic covariates: age, sex, education (in terms of schooling years) and log-income. For simplicity, we fixed all ages at the enrolment time. Thus all covariates are time-independent. More covariates can be added, but we are limited by computation capacity.

2.5 The Model

As discussed before, the focal point of the self-exciting counting process approach is to model the ground (cumulative) intensity function. We construct the intensities by explicitly taking different randomness sources of cumulative individual cost $X(t)$ and episode dependence structure into consideration.

2.5.1 Free Insurance Plan

Our ground intensity $\lambda_g(t)$ for each individual[5] who belongs to the free insurance plan consists of two parts:$\lambda_g(t) = \lambda_1\lambda_2(t)$. λ_1 deals with the covariates effect, while λ_2 is the SD term discussed above.

Like many count data regression and duration models, the covariates effect is presented as an exponential function:

$$\lambda_1(Z) = exp(\gamma^T Z) \tag{2.2}$$

where Z is a vector of individual characteristics including age, sex, education and log-income, etc.

The SD term is specified as:

$$\lambda_2(t) = \sum_{i=1}^{N(t-)} \mu \cdot exp(-\mu(t - t_i)), \mu > 0 \tag{2.3}$$

[5] Therefore, we ignore the individual subscript.

When no event occurs before time t, we normalize the SD term to unit. That is $\lambda_2(t) = 1$ if $N(t-) = 0$. The ground intensity is then degenerated to $\lambda_g(t) = exp(\gamma^T Z)$. Implicitly, we assume that before the first doctor visit, the individuals do not understand their own health status and therefore, there is no individual heterogeneity.

The 'kernel' $\mu \cdot exp(\cdot)$ characterizes the episode dependence structure. More specifically, the propensity of a follow-up visit is governed by such a 'kernel': the intensity is high when the elapsed time is short and will gradually decrease as time goes by. We will argue such an assumption is reasonable: the individual is vulnerable when she just receives the treatment and is more likely to be sick again, but she will gradually recover as time goes by and will be less likely to experience sickness. The summation over these 'kernels' means we take all the past episodes into consideration. But the weight for each episode is different. By construction, the effects of far away past experiences will deteriorate, but the latest ones have the most important influences.

The usual method to model such phenomena in a structural form model is to assume health events arrive periodically with a probability $S^{'}$, which is drawn from $F(S^{'}|S)$ where S is the arrival probability from a previous period. Einav et al. (2015) further simplify this assumption by letting S take one of two values, S^L and S^H (with $S^L < S^H$), and that $Pr(S^{'} = S^J|S = S^J) \geq 0.5, J \in \{L, H\}$, so there is weakly positive serial correlation. This exceedingly simplified assumption is made mainly for computational reasons. And the above Markov process is most likely inadequate to model the episode cluster structure. We conclude that our cluster set up is more realistic and is quite difficult, if not impossible, to build within the conventional econometric models.

To sum up, for the free insurance plan, P0, the intensity is expressed as:

$$\lambda_{P0}(t) = \lambda_1(Z)\lambda_2(t) \tag{2.4}$$

2.5.2 Cost-Sharing Insurance Plan

As for the cost-sharing plan, λ_1 does not change. The cost sharing policy has two hypothetical effects: 1) The late year effect, that is when the contract year is near the end, individuals, especially those who have already exceeded the OPC may use the medical service more frequently than before (cash-in effect) since the cost-sharing policy will be set to default next year and the shadow co-insurance rate would be expensive once again. 2) The shadow price effect discussed in the introduction section. We update the cumulative individual cost whenever an event occurs. To account for the cost-sharing effects, we modify λ_2 as follows:

$$\lambda_2^*(t) = \beta_1 exp(\beta_1 t) + \sum_{i=1}^{N_g(t-)} b\exp(\beta_2 X(t_i))\mu exp(-\mu(t-t_i)) \tag{2.5}$$

here $X(t)$ is the cumulative individual cost at time t. It includes the non-covered charge from outpatient medical utilization as well as drug purchase:

$$X(t) = \sum_{i=1}^{N_g^1(t-)} x_i + \sum_{i=1}^{N_g^2(t-)} y_i \tag{2.6}$$

where x_i is the non-covered charge for i^{th} doctor visiting, $N_g^1(t)$ is the associated ground counting process. y_i is the non-covered charge for i^{th} drug purchase and $N_g^2(t)$ is the drug purchase ground counting process. The construction of $X(t)$ essentially follows the definition of cumulative individual cost mentioned in the introduction section. Recall the shadow price is defined as $1 - V(X(t))$, where $V(X(t))$ is the bonus which depends on the cumulative individual cost. If $V(X(t)) \propto \exp(\beta_2 X(t))$, then the term $b \exp(\beta_2 X(t))$ can be thought of as a measure of medical utilization bonus. We would expect $\beta_2 > 0$ to be significant if individuals do respond to shadow price.

We use the term $\beta_1 exp(\beta_1 t)$ to model the late year effect: we would observe β_1 significantly greater than zero if such an effect is true.

To summarize, the ground intensity for the cost-sharing plan is:

$$\lambda_{P95}(t) = \lambda_1(Z)\lambda_2^*(t) \tag{2.7}$$

There are several pieces to put together in order to estimate the parameters of the cost-sharing effects model. As Keeler & Rolph (1988), we assume that there are no interactions between within-year cost sharing effects and the effects of other explanatory variables, so that all the effects of explanatory variables other than cost sharing on frequencies of episodes are summarized in $\lambda_1(Z)$ and all episode dependence structure is captured by $\lambda_2(t)$ $(\lambda_2^{'}(t))$. We first estimate the free plan by minimizing

$$||\bar{N}_g^{P0} - \lambda_1(Z) \int_0^T \lambda_2(t)dt||_{\bar{N}_g^{P0}}$$

thus, the individual heterogeneity and the episode dependence structure of the intensity are estimated by $\hat{\lambda}_1(Z)$ and $\hat{\lambda}_2(t)$. When estimating the cost-sharing plan, these two parts are then treated as fixed, which leaves us with only cost-sharing effect parameters (i.e.,β_1, β_2 and b) to be estimated[6]. Thus the minimization object is:

$$||\bar{N}_g^{P95} - \hat{\lambda}_1(Z) \int_0^T \big(\beta_1 exp(\beta_1 t) + \sum_{i=1}^{N_g(t-)} b \exp(\beta_2 X(t_i))\hat{\mu} exp(-\hat{\mu}(t-t_i))\big) dt||_{\bar{N}_g^{P95}}$$

2.6 Main Results

The main results are presented in Table 2.2. Some words on the numerical optimization are in order. Because of the non-linear nature and complexity of the cumulative intensity function, it is impossible to write down a closed form solution for this minimization problem. And some numerical algorithms are implemented to find the optimization. For the free plan, we first run a simulated annealing (SA) routine to assess reasonable ranges for all the parameters, then a down-hill (Nelder-Mead) optimization algorithm is employed to refine the results. Similar steps are used for the cost-sharing plan with individual specific parts of cumulative intensity fixed. We manually stop the SA algorithm after 24 hours but do not intervene with the down-hill algorithm until it reports success.

2.6.1 Interpreting the Covariates

The interpretation of coefficients is not as straightforward as in linear regression. However, we may fix a time period and treat the counting process as count data. The interpretation is then identical to that of a count data regression analysis. Formally,

[6]We exploit the fact that all individuals are assigned to different plans randomly. By plugging the individual specific estimators from the free plan into the cost-sharing plan, we can still have consistent estimators.

Table 2.1:: Basic Results

	Estimator	*Description*
μ	27.87126321*** (8.39814247)	coefficient of the episode dependent structure
age	-0.13394806*** (0.03122223)	
age2	0.15470947*** (0.04225468)	$(age)^2/100$
male	-0.71703944 (0.4738397)	
edu	-0.35495029*** (0.0858603)	
edu2	0.99428386*** (0.3361197)	$(edu)^2/100$
log income	0.59265516*** (0.03643453)	
b	0.65898635*** (0.0580308)	
β_1	0.1068388 (0.27547018)	coefficient of late year effect
β_2	0.00383393*** (0.00061892885)	coefficient of non-covered charge
Distance	0.898473	Free Plan
Distance	1.10226	Cost-Sharing Plan

Note: standard errors in brackets, *p<0.1; **p<0.05; ***p<0.01

recall the Doob-Meyer decomposition, for a fixed time period $[0, t], \forall t \in [t, \bar{t}]$, we have

$$\mathbb{E}(\Lambda_g(t|Z)) = \mathbb{E}(N_g(t)|Z) = \mathbb{E}(Y_t|Z)$$

The count data Y_t is the number of events occurring during this time period. Let scalar z_j denote the j^{th} covariate. Differentiating

$$\frac{\partial \mathbb{E}(Y_t|Z)}{\partial z_j} = \gamma_j \mathbb{E}(\Lambda_g(t|Z))$$

by the exponential structure of $\lambda_1(Z)$. That is, for example, if $\hat{\gamma}_j = 0.2$, $\bar{\Lambda}_n(t|Z) = 2.5$, then one-unit change in the j^{th} covariate increases the expectation of Y_t by 0.5 units.

Two remarks are in order. First, notice that the sign of the response $\partial \mathbb{E}(Y_t|Z)/\partial z_j$ is given by the sign of γ_j since the accumulated intensity Λ_t is always positive. Second, if one covariate coefficient is twice as large as another, then the effect of a one-unit change of the associated covariate is double that of the other. This result follows from

$$\frac{\partial \mathbb{E}(Y_t|Z)/\partial z_j}{\partial \mathbb{E}(Y_t|Z)/\partial z_i} = \frac{\gamma_j \mathbb{E}(\Lambda_g(t|Z))}{\gamma_i \mathbb{E}(\Lambda_g(t|Z))} = \frac{\gamma_j}{\gamma_i}$$

With these in mind, we can interpret our results. *Age.* The overall effect for age is as follows: at first, the intensity will decrease as age increases, after one passes the age of 43, the intensity and age are positively correlated. It is well-known that the youngsters are more risky compared to their mid-age counterparts. While as individuals begin to age, they become physically weaker and more prone to sickness.

Sex. Females seem to be more likely to go the doctor, but the result is not significant.

Education. There are two explanations about the correlation between education and frequency of doctor visits. One is that individuals with a higher level of education are positioned in more important jobs and their absence from work may damage not only their output but also that of their peers', thus the potential cost of going to hospital is much higher which leads to a negative correlation. The other explanation says that with higher education, people are more aware of the importance of good health and are willing to go to the doctor more frequently, besides, education and income are known to be positively linked, thus education should be correlated to intensity positively. Our results suggest that with education of less than 17 years (roughly equivalent to a Master's degree), the overall effect favors the first explanation. But with a higher education level (Master and above) the overall effect favors the second explanation.

Income. Income is positively related to the use of medical services, which is not surprising. A higher income gives individuals the ability to cover the opportunity cost related to absence from work (to visit a doctor).

2.6.2 Cost-Sharing Effects

There is weak evidence supporting the existence of the late year effect (t-ratio of 0.387842).

The shadow price effect is captured by $b\exp(\beta_2 R(t))$. The most important parameter here is β_2. If β_2 is close enough to zero, we may observe a flat, almost linear curve, which indicates that individuals only respond to one price system (the spot price system). However, if β_2 is positively away from zero, we can safely claim that individuals do understand the design of the insurance policy and take advantage of the shadow price. Our result provides strong evidence for the shadow price effect and we are confident to reject the null hypothesis: $\beta_2 = 0$. Figure 2.4 shows the graphical result.

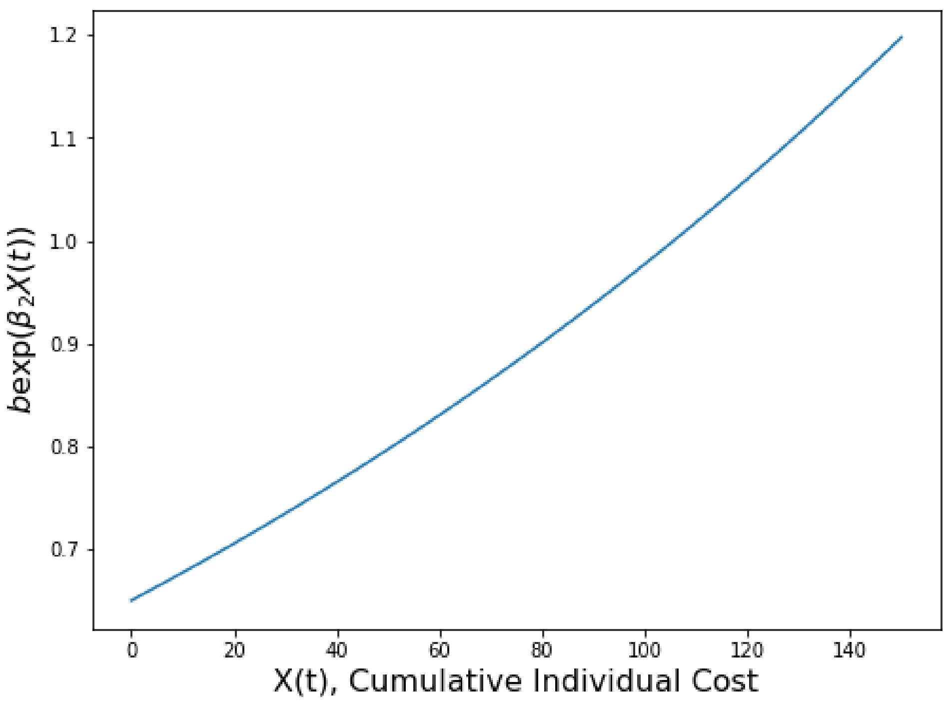
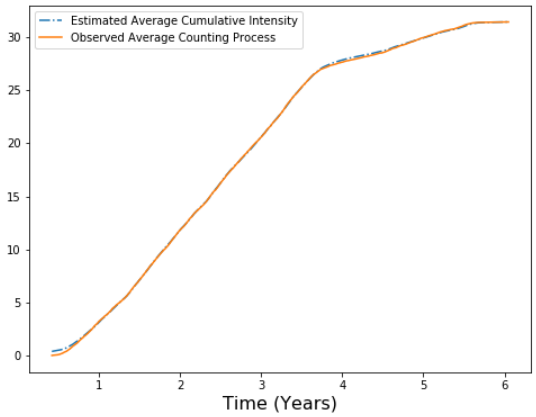
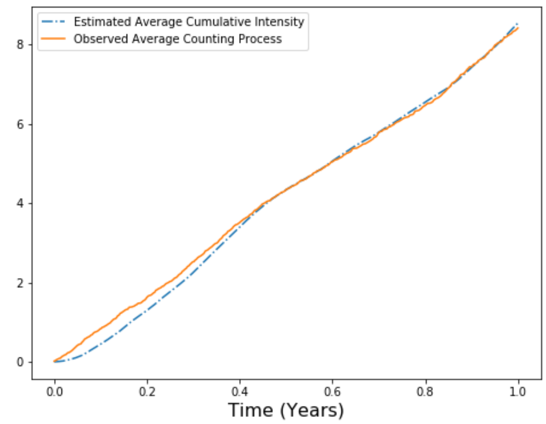


Figure 2.4: The shadow price effect

Overall, the model fits the data well. To assess the goodness of fit, we generate estimated (averaged) cumulative intensity against the observed (averaged) counting process. Figure 3.1 presents the results. The patterns of estimated and observed are quite similar in both free plan and cost-sharing plan.



(a) free plan



(b) cost-sharing plan

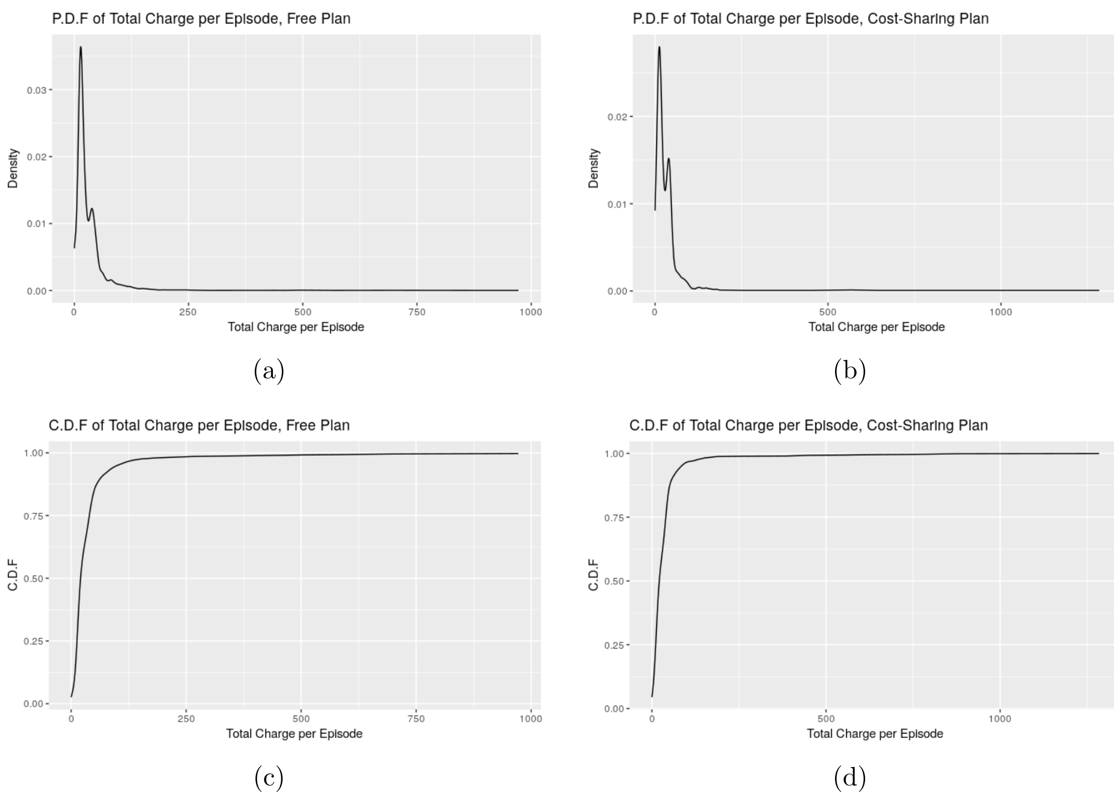
Figure 2.5: Goodness of fit

2.6.3 Mark Density

To complete our model, we will non-parametrically estimate the conditional mark density. The interested mark in this application should be the non-covered charge. However, in the free plan, such marks are zero in most records. Because of this, and for the purpose of comparison between these two plans, we instead use total charge as our marks. Since insurance plans have clear regulation on the cost-sharing policies, once we observe the total charge, there is no ambiguity in knowing the non-covered charge.

We assume the nominal price are i.i.d distributed. Thus we have $f(x|t, \mathcal{H}_{t-}) = f(x)$. A standard kernel density estimation is used to analyse this mark density. The bandwidth selection results are 3.84075 and 5.79464 for cost-sharing and free plans respectively. We plot the densities and distributions in Figure 2.6.

It is not hard to tell that the densities in the two different plans are similar, indicating the charge per episode is stable across various insurance plans. This result is consistent with previous studies (e.g. Keeler & Rolph (1988))



(a) (b)

(c) (d)

Figure 2.6: Plots of total charge empirical distributions with a Gaussian kernel

2.6.4 Cluster Analysis

The episodes tend to be clustered or grouped together (i.e., we are rejecting the assumption that episodes are independent). One reason is because of the nature of chronic diseases: regular or frequent treatments are needed to ease or eliminate the pain. Another explanation is because one disease may trigger the occurrence of another one in the short term.

As mentioned before, the dependent structure (or the cluster structure) is governed by $\exp(-\mu t)$. Figure 2.7 presents such a structure using our estimator $\hat{\mu}$.

It is not hard to tell that the densities in the two different plans are similar, indicating the charge per episode is stable across various insurance plans. This result is consistent with previous studies (e.g. Keeler & Rolph (1988))

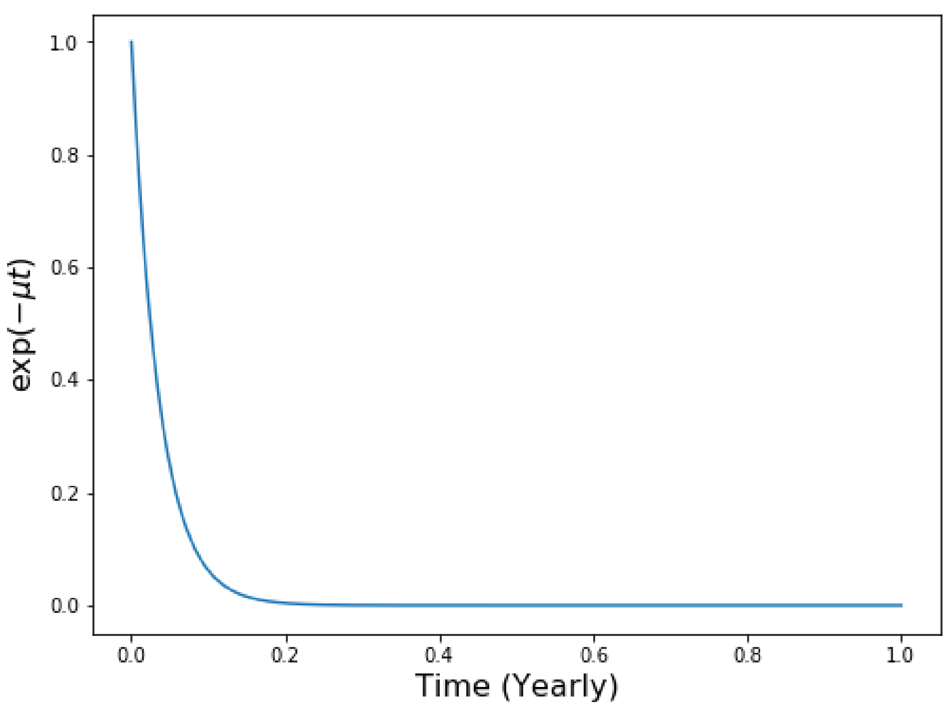


Figure 2.7: The cluster structure among episodes

The likelihood of follow-up visiting is high at the beginning and then decreases as time goes by. After roughly 3 weeks to one month, the likelihood is small enough to ignore.

Literatures have documented that cost-sharing policies reduce the frequency of medical utilization (e.g. Keeler & Rolph (1988); Aron-Dine et al. (2015)). Our question is how do these policies affect the cluster structure among doctor visiting? Will they reduce the average number of clusters per person? Will they reduce the average follow-up visits inside a cluster? To the best of our knowledge, few literature has considered these issues, since most of them use the hypothesis that episodes are independent.

Here we use a cluster analysis algorithm called DBSCAN (Density-based spatial clustering of applications with noise) that is widely used in computer science and statistical learning.

For this algorithm, there are two inputs: Eps, the radius of one density region, and minPts, the minimum number of points required to form a dense region. For the purpose of DBSCAN clustering, all points are classified as core points, border points and noise points. Core and border points form a cluster via different definitions of 'reachable'. Noise points are the points that do not belong to any cluster. We provide details of this algorithm and the definition of a cluster in Appendix.

The ability of this algorithm to identify 'noise' points is particularly appealing to us. This is because some acute episodes are small in scale and only need one doctor visit to fully recover. They are not linked to the rest of episodes.

Based on the estimation of the SD term, we set $Eps = 21$ days and as a rule of thumb $minPts = 2$[7]. For the purpose of comparison, we restrict the time horizon in both plans to 1977-1978 insurance year. For each individual (both free plan and cost-sharing plan), we run the DBSCAN algorithm, document the number of clusters, the average number of instances per cluster and the number of noise points. For each insurance plan, we then compute the average number of clusters per person, the average number of instances per cluster per person and the average noise points per person. Table 2.2 summarizes the results.

Table 2.2:: Cluster Analysis

	avg cluster number	avg cluster members	avg noise points
free plan	1.2287	4.55187	1.62332
Cost-sharing plan	0.862595	3.3625	1.47328

The effects of cost-sharing policies on cluster structure are threefold. First, they reduce the average number of clusters per person. That means for the initial episode, the cost-sharing policies suppress the first doctor visiting behaviors. Second, within each cluster, they reduce the number of follow-up visits. Third, cost-sharing policies reduce the average number of noise points per person, i.e., they discourage individuals to use medical services when they have small episodes like minor injuries.

2.7 Discussion

Our model characterizes the individual's decision making process in both free and cost-sharing plans. The results show that individuals react to shadow price systems.

[7] The rule of thumb is minPts = dimension +1

The method we adopted is different than conventional reduced form or structural form econometrics tools used in non-linear budget constraint studies.

In the literature of medical utilization, how to quantify the response of medical spending with respect to its price is at the core of the debate. Cutler & Zeckhauser (2000) summarize a long list of literature all reporting an estimate of a single price elasticity of demand for medical care with respect to the out-of-pocket price. Particular to the data we used, the RAND HIE, such an estimate is -0.2(Manning et al. 1987),(Keeler & Rolph 1988)). Most of these literatures obtain the estimate of such single elasticity by assuming individuals only respond to the spot (out-of-pocket) price. Recent literatures (e.g. Cardon & Hendel (2001), Dalton (2014), Kowalski (2015),Aron-Dine et al. (2015) and Einav et al. (2015)) deviate from this assumption and find strong evidence for such deviation.

Aron-Dine et al. (2015) proposes to use two different elasticities in a classical reduced form: one with respect to spot price and the other with respect to future price. They define the future price as the expected end-of-year price, with expectations taken over all individuals in the same insurance plan. Such future price depends on three elements: the cost-sharing features of the insurance plan, the duration of the insurance plan and the expected spending of individuals. Thus, essentially, they impose a strong assumption that individuals have no private information about their health conditions and health shocks. Although using a firm-level data, they find strong evidence rejecting the null hypothesis that individuals only respond to spot price, they fail to find similar conclusions using the Rand HIE data. The explanation they give for this result is the relatively small sample size.

Our method, on the other hand, tackles the problem of quantifying the response of medical spending from a different perspective: rather than estimating the elasticities directly, we try to measure the bonus $V(X)$ as a function of cumulative individual cost. By doing so, we avoid unrealistic assumptions made by Aron-Dine et al. (2015) but at the same time manage to keep the non-linear property. In addition, we exploit the data in its maximum capacity. Instead of just looking $X(T)$ like Aron-Dine et al. (2015), we use all the cumulative individual cost information: $X(t) : t \leq T$. Thus, in our method, a more direct measurement of responses to the out-of-pocket spending up until now (β in our model) is estimated instead of the elasticities. We believe such features explain our success in finding evidence supporting the non-linear budget constraint in the Rand HIE data.

In a structural paper,Einav, Finkelstein, & Schrimpf (2015) build and estimate a simple dynamic model of optimizing agent's drug utilization decisions given a non-linear insurance contract design. They also find evidence supporting the hypothesis that individuals take into account the dynamic incentives by showing the discount factor of their value function is non-zero. The shortcoming of this structural form is that it fails to address the stochastic nature of the occurrences. What structural models do is to divide the time line into several equispaced cells, and within each cell, one only needs to concern whether an event is occurred or not. As for how many events occurred in that cell and more importantly the timings of occurrences, they are irrelevant. The self-exciting approach, on the other hand, is more adequate to model the stochastic properties of individual cost $X(t)$: the events' occurrence times $\{t_i\}_{i\in\mathcal{N}^+}$ are at the central of our model objectives. In addition, compared to the structural form, our method is much simpler and only requires minimal assumptions. For example, it is a common practice in structural form to use a Markov transition probability to describe the shock dependence. Einav et al. (2015) further simplify it by restricting the shocks to be either big or small. These assumptions are mainly made for computational purpose. There is littler reason to believe that the real mechanism would follow such restriction. In comparison, our approach allows all the past experiences contribute to future randomness in a form of state dependence. This is a much less restricted and more realistic and complicated assumption. Despite that, our model is still computationally easier than a typical structural form econometric model.

2.8 Conclusion

In this paper, we provide a methodology to construct a behavioral model of medical care utilization. At the core of this method is the self-exciting counting process. It allows researchers to take historical information into the model. A minimum distance estimation is advocated instead of the conventional likelihood-based methods. By doing so, one may introduce external shocks to the self-exciting process. This enables researchers to use more realistic model settings. We use such a methodology to build a decision making process model of medical care utilization and find that individuals are responsive to shadow price and take into account the dynamic incentives. Furthermore, we use a matured statistical learning algorithm to analyze the cluster structure of doctor visiting behaviors. We find cost-sharing policies do affect the clusters in numerous ways.

Appendix

2.A The KNP Theorem

Keeler,Newhouse and Phelps (1977) Keeler et al. (1977) considered the theory of a consumer facing a deductible health insurance contract. There are two differences from the usual consumer theory: 1) the price per unit in a specified amount of medical care services bought within a specified period is changeable and 2) illness uncertainty is present regarding the future medical service demand. When uncertainly is present, any cost below the range of deductible limit has the bonus of reducing the remaining deductible, and hence reducing the future cost. That is the greater the chance that future expenditures will exceed the deductible, the cheaper today's purchase of medical care service. They argue that when analysing the reaction to the deductible of a consumer, one needs to take account of the sequential decision problem.

Formally, they assume a rational agent whose object is to maximize a sequence of utilities $(U(e_t, H_t))_t$ under illness uncertainty, where, after ignoring the time subscription, e is the flow of other consumption goods and H, in terms of dollar, is the stock of health which is related to the medical care service h. The evolution of the perceived health stock is $H_t = H_{t-1} - l_t + g(h_t, l_t)$, where l, in terms of dollar, is the random loss of health from illness and $g(h, l)$ is the production function of the stock of health. Before exceeding the deductible, denote $o_t = p_h \times h_t$ as the out-of-pocket payments for medical care at time t where p_h is the medical service price, after reaching the deductible, the co-insurance rate C would apply to the consumers, we normalize the price of other goods to 1. Let $(d_t)_t$ be a sequence of the remaining deductibles.

The one-period demand case is shown in the figure 2.A.1

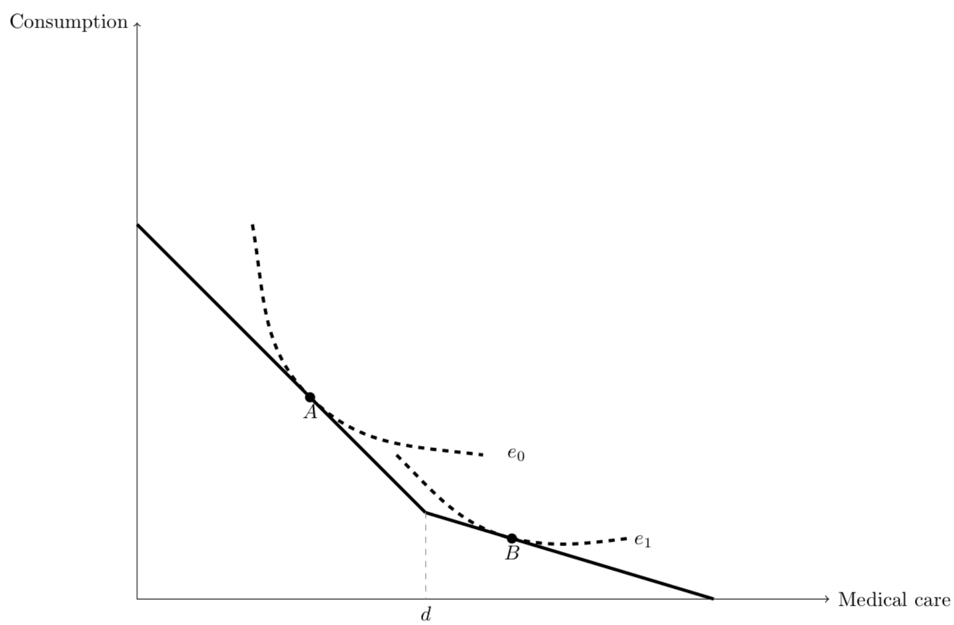


Figure 2.A.1: An insurance policy with a deductible equal to d

They have shown that in this one-decision case that consumers with smooth indifference curves would never be observed purchasing exactly d units of medical care and, in general, would not be near the kink in the budget line. Furthermore, with a normal utility function, there exists a critical level of illness l_d. All losses smaller than l_d will produce purchases of h less than d, and losses greater than l_d will lead to $h > d$.

Suppose $U(X, H)$ applies only to the stocks of other goods and health at the end of the deductible period. The multi-periods problem can be represented formally as a consumer trying to maximize the expected value of:

$$\sum U_t(e_t, H_t) + U(X, H) \tag{2.8}$$

subject to the budget constraint:

$$X + \sum (e_t + o_t) = Income \tag{2.9}$$

Notice that it is easy to see that the medical service needed to cure the illness is related the scale of illness l, hence we should let the other consumption and the medical service be functions of l: $e(l), h(l)$.

Such problem can be formulated as a dynamic program. Let $W(x_t, H_t, d_t, t)$ be the expected sum of utility from t to the end of the deductible period, for an agent who at the start of period t has wealth x_t, health H_t and remaining deductibles d_t. We suppose that the distribution of illness, $f(l, H_t)$, also depends on H_t. Then:

$$W(x_t, H_t, d_t, t) = \int \max_{e(l),h(l)} U\{e(l), H_t - l + g[h(l)]\} + W(x_t - e(l) - p_h h(l), H_{t+1}, d_{t+1}, t+1) f(l, H_t) dl \tag{2.10}$$

where $d_{t+1} = max(0, d_t - o_t)$. The formula for the remaining deductibles means one needs to take this optimization problem in a sequential and contingent way: all the history deductible information should be included to calculate the current remaining deductible. W can be calculated backwards from the end of the period to his present position.

Such results have important implications for estimating demand curves for medical services. For a person's j^{th} illness, the demand can be expressed as:

$$q_j = f(p_j, Z) + \epsilon_j$$

where q_j is the quantity demanded by this individual in the j^{th} episode, p_j is the shadow price for j^{th} episode, Z is vector of other variables that may affect the demand, and ϵ_j is a random error term with $\mathbb{E}(\epsilon_j) = 0$, $var(\epsilon_j) < \infty$.

Given the existence of remaining deductible, the model suggests that each p_j is a function of demands (hence out-of-pocket fees) prior to the j^{th} episode and of time remaining in the period,

$$p_j = g\big(\sum_{k=1}^{j} q_k, t_j\big)$$

Suppose T is the number of episodes occurring during one year and there is no variation in T across individuals, then the average effective price is

$$p = \sum_{j=1}^{T} p_j / T \tag{2.11}$$

Most of the literature did not use the p as the measurement of price; instead, they use the fraction of annual expenditures paid out of pocket $\bar{p}$. However, $\bar{p}$ measures the effective price with error: $(\bar{p} - p) = \delta$. The authors argue that since the price measurement error δ is not random but is generally correlated with the true price, the estimated price coefficient could be inconsistent.

2.B Janossy representation of generic probabilistic structure of marked point process

Definition The Janossy measures are non-probability measures for point process N and are defined as the measures satisfying,

$$J_n(A_1 \times, \cdots, \times A_n) = p_n \sum_{perm} \Pi_n(A_{i_1} \times, \cdots, \times A_{i_n}) \tag{2.12}$$

For marked point process $\bar{N}$ they are,

$$J_n(A_1' \times, \cdots, \times A_n') = p_n \sum_{perm} \Pi_n(A_{i_1}' \times, \cdots, \times A_{i_n}') \tag{2.13}$$

Where $A_i' = A_i \times B_i$, B_i are the mark spaces.

The Janossy measure has a nice interpretation, if $E = R$ and $j_n(t_1, \cdots, t_n)$ denotes the density of $J_n(\cdot)$ with respect to a Lebesgue measure on $(R)^n$ with $t_i \neq t_j$ if $i \neq j$, then $j_n(t_1, \cdots, t_n)dt_1 \cdots dt_n = Pr\{$ there are exactly n points in the process one in each of the n distinct infinitesimal regions $(t_i, t_i + dt_i)$ $\}$. Its marked counterpart density is defined as

$$j_n(t_1, \cdots, t_n, x_1, \cdots, x_n)dt_1 \cdots dt_n dx_1 \cdots dx_n \tag{2.14}$$

with interpretation as $Pr\{$ points around $\{t_1,\cdots,t_n\}$ with marks around $(x_1, \cdots, x_n)$ $\}$. These interpretations, are in fact indicating that the Janossy density is nothing but the likelihood of a point process.

Definition The likelihood of a realization of a point process N on a bounded Borel set $E \subseteq R^d$, when $n = N(E)$, is the local Janossy density

$$L_E(t_1, \cdots, t_n) = j_n(t_1, \cdots, t_n|E) \tag{2.15}$$

In the formation of Janossy measure, the condition $\sum p_n = 1$ can take the form of

$$\textstyle\sum_{n=0}^{\infty}(n!)^{-1}J_n(E^{(n)}) = 1$$

It turns out that the Janossy measures can uniquely determine the point process, one can find the proof as well as a more detailed introduction of Janossy measure in Chapter

5,7 of Daley et al (2007)Daley & Vere-Jones (2007)

We now use the Janossy measure to describe the probability of the marked point process. Let $S_n(t|\mathcal{F}_{t-}) = 1 - \int_{t_{n-1}}^{t} p_n(u|(t_1, x_1). \cdots, (t_{n-1}, x_{n-1}))du$ be the conditional survival function. The probability structure of the marked point process can be viewed as

$$
\begin{aligned}
J_0(T) &= S_1(T), \\
j_1(t_1, x_1|T) &= p_1(t_1)f_1(x_1|t_1)S_2(T|(t_1, x_1)) \text{ as } (0 < t_1 < T) \\
j_2(t_1, t_2, x_1, x_2|T) &= p_1(t_1)f_1(x_1|t_1)p_2(t_2|(t_1, x_1))f_2(x_2|(t_1, x_1), t_2)S_3(T|(t_1, t_2), (x_1, x_2)) \\
&\quad \text{as } (0 < t_1 < t_2 < T)
\end{aligned}
\tag{2.16}
$$

where T is the endpoint (length of the time interval), p_i are the densities, suitably conditioned, for the locations in the ground process, and the $f_i(\cdot)$ refer to the densities, again suitably conditioned, for the marks.

The interpretation is just like the one mentioned at very beginning of this subsection, take $j_1(t_1, x_1|T)$ as an example, it says the likelihood of there is one point locates in the time interval $[0, T]$ with mark value x_1 is equal to the probability of only one event happened times the density of the mark of being x_1 conditional on such event happened at time t_1, also we need to multiple the survival function conditional on the first event information.

We now make a critical shift of view. Instead of specifying the conditional density p_n, we express them in terms of conditional hazard functions,

$$h_n(t|t_1, \cdots, t_{n-1}) = \frac{p_n(t|t_1, \cdots, t_{n-1})}{S_n(t|t_1, \cdots, t_{n-1})} \tag{2.17}$$

Equivalently, $p_n(t|t_1, \cdots, t_{n-1}) = h_n(t|t_1, \cdots, t_{n-1}) \exp(-\int_{t_{n-1}}^{t} h_n(u|t_1, \cdots, t_{n-1})du)$. Plug it into Equation 2.16, and recall from Equation **??** the likelihood of a counting process is

$$L(\cdot) = \Pi_i \lambda_{t_i}(B) \exp\Big(-\int \lambda_u(B)du\Big)$$

we have,

$$\begin{aligned}
j_n(t_1, \cdots, t_n, x_1, \cdots, x_n|T) &= h_1(t_1) \cdots h_n(t_n|t_1, \cdots, t_{n-1}, x_1, \cdots, x_{n-1}) \times \\
&\quad f_1(x_1|t_1) \cdots f_n(x_n|t_1, \cdots, t_n, x_1, \cdots, x_{n-1}) \times \\
&\quad \exp\Big(-\int_0^{t_1} h_1(u)du\Big) \cdots \exp\Big(-\int_{t_n}^{T} h_{n+1}(T|t_1, \cdots, t_n, x_1, \cdots, x_n)\Big) \\
&= L(\cdot)
\end{aligned} \tag{2.18}$$

Thus, the densities for the locations can be expressed in terms of corresponding hazard functions. And the conditioning in the hazard functions now can include the values of the preceding marks as well as the length of the current and preceding intervals.

$$\lambda(t,x|\mathcal{F}_{t-}) = \begin{cases} h_1(t)f_1(x|t) \\ \vdots \\ h_n(t|(t_1,x_1),\cdots,(t_{n-1},x_{n-1}) \times f_n(x|(t_1,x_1),\cdots,(t_{n-1},x_{n-1},t),(t_{n-1} < t \leq t_n, n \geq 2) \\ \vdots \end{cases} \tag{2.19}$$

Where $\mathcal{F}_{t-}$ again is the internal history of the marked point process and $h_i(\cdot)$ are the corresponding hazard functions,suitably conditioned. Notice that an generalization from internal history to a whole history $\mathcal{H}_{t-}$ can be done quite easily, we just need to let the Janossy density to be conditional on the external history information. We can rewrite the above as

$$\lambda(t,x|\mathcal{F}_{t-}) = \lambda_g(t,x|\mathcal{F}_{t-})f(x|t,\mathcal{F}_{t-}) \tag{2.20}$$

where λ_g is the ground intensity. When the mark space is continuous, we have $\lambda(t,x|\mathcal{F}_{t-})dtdx = \mathbb{E}[\bar{N}(dt \times dx)|\mathcal{F}_{t-}] = \lambda_g(t,x|\mathcal{F}_{t-})f(x|t,\mathcal{F}_{t-})dtdx.$

2.C The Thinning Method for Simulation

The detailed thinning method steps can be summarised as:

1. Let τ be the start point of a small simulation interval
2. Take a small interval $(\tau, \tau + \delta)$

3. Calculate the maximum of $\lambda_g(t|\mathcal{F}_{t-})$ in the interval as

$$\lambda_{max} = \max_{t\in(\tau,\tau+\delta)} \lambda_g(t|\mathcal{F}_{t-})$$

4. Simulate an exponential random number ξ with rate λ_{max}

5. if

$$\tfrac{\lambda_g(\tau+\xi|\mathcal{F}_{t-})}{\lambda_{max}} < 1$$

   go to step 6.

   Else no events occurred in interval $(\tau, \tau + \delta)$, and set the start point at $\tau \leftarrow \tau + \delta$ and return to step 2

6. Simulate a uniform random number U on the interval $(0, 1)$

7. If

$$U \leq \tfrac{\lambda_g(\tau+\xi|\mathcal{F}_{t-})}{\lambda_{max}}$$

   then a new 'event' occurs at time $t_i = \tau + \xi$. Simulate the associated marks for this new event.

8. Increase $\tau \leftarrow \tau + \xi$ for the next event simulation

9. Return to step 2

2.D DBSCAN Cluster Analysis

The DBSCAN algorithm classified all points into three: core points, border points and noise points. We start by defining these points. For a set of points $X = \{x_1, x_2, \cdots, x_N\}$.

Definition ϵ neighbourhood of a point x, denoted by $N_\epsilon(x)$ is defined by $N_\epsilon(x) = \{y \in X : d(y, x) \leq \epsilon\}$. Where $d()$ is a metric.

Definition Density is defined as $\rho(x) = |N_\epsilon(x)|$, the number of points in a ϵ neighbourhood.

Definition Core point: let $x \in X$, if $\rho(x) \geq minPts$, then we call x a core point. The set of all core points is denoted as X_c, let $X_{nc} = X \setminus X_c$ be the set of all non-core points.

Definition Border point: if $x \in X_{nc}$ and $\exists y \in X$ such that $y \in N_\epsilon(x) \cap X_c$, then x is called a border point. Let X_{bd} be the set of all border points.

Definition Noise point: let $X_{noise} = X \setminus (X_c \cup X_{bd})$, if $x \in X_{noise}$, then we call x is a noise point.

To define what is a cluster under the DBSCAN setting, we need a few more definitions about 'reachable'.

Definition Directly density-reachable: if $x \in X_c$ and $y \in N_\epsilon(x)$, we may say y is directly reachable from x.

Definition Density-reachable: let $x_1, x_2, \cdots, x_m \in X, m \geq 2$. If x_{i+1} is directly density-reachable from x_i, $i = 1, 2, \cdots, m-1$. We call x_m is density-reachable from x_1.

Definition Density-connected: a point x is density connected to a point y if there exists another point $z \in X$ such that both y and x are density-reachable from z.

Definition Cluster: a non-empty subset C of X is called cluster if it satisfies:

- (Maximality) $\forall x, y$: if $x \in C$ and y is density-reachable from x, then $y \in C$.
- (Connectivity) $\forall x, y \in C$: x is density-reachable to y.

For a detailed algorithm description, we refer to the original Ester et al.(1996)Ester et al. (1996) paper.

Chapter 3

The Dynamic Behavior in Work Absence

Abstract

We use the self-exciting processes to study individuals' absence behaviors. Such behaviors are dynamic because of the firm's absence regulation, where a worker's absence records determine her absence benefit. The self-exciting process is state-dependent and enables us to include the individual's absence records into the model. We decompose an absence into an incidence event ('asking for absence') and a recovery event ('returning to work'). For each absence, we also distinguish short-term from long-term. Using firm-level data, we find that workers do consider absence records when they make short-term incidence and recovery decisions, but this is not the case for long-term events. Inspired by the empirical results, we build a simple economic model.

3.1 Introduction

The primary purpose of this chapter is to investigate the dynamic behaviors in work absences, where individuals may take past experiences into absence decision-making considerations. Dynamic here means preferences or constraints to future choices are altered by past experiences. A typical source of such behavior is the absence score that most firms employed in their work absence regulations. Absence scores are accumulated over time and are based on individuals' absence records. In principle, higher absence scores lead to more severe penalties (e.g., fall of income or even possible layout) and vice verse. Individuals then have to dynamically make their absence decisions. In this paper, we use the self-exciting process, a special counting process, to model the dynamic and state-dependent absence decision-making processes.

Work absence is not uncommon among both developed and developing countries. U.S Bureau of Labour 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, labour costs are the single most considerable budgetary expense. Fister-Gale (2003) cites research showing that absenteeism costs in one survey population accounted as high as 14.3% of total payroll.

Despite the sizable number of work-hours involved and impacts on productivity, economists have paid little attention to the issue of absenteeism. 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 filed. Markussen et al. (2011) show that employee heterogeneity drives most of the 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 in the work absenteeism literature. In general, psychological literature argues, according to Steers & Rhodes (1978), that the 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 literature argue that absence scores is a significant work absence decision-making factor. However, no satisfied empirical work has been done to support this claim.

In this paper, we aim to provide empirical evidence on the existence of this trigger absence behavior. The inclusion of absence score creates a state-dependent structure in the econometric models. We use self-exciting processes to incorporate such structure. The self-exciting process is a counting process whose filtration is generated by the counting process itself. Thus, a self-exciting process is state dependent: past experience has effects on the future events. Throughout the analysis, we try to avoid making further assumptions other than the independent, identical distributed individuals. However, within a single individual, absences are not independent.

To fully explore the features of the self-exciting process, we decompose a work absence into two decision making processes: an incidence process and a recovery process. An event in the incidence process is defined as a worker asking for an absence. An event in the recovery process happens when a worker decides to return to work from the absence. Incidence processes and recovery processes are different. Individuals may encounter shocks (e.g. illness) and seek absences, but at first, they do not have full information about the shock sizes. However, such information is available when individuals make the returning decisions. The stochastic interpretation is also different. For example, the likelihood for a hard-working individual to ask for a leave is low, while the likelihood for the same individual to return to work quickly is high. To this end, we will model these two processes separately.

We also distinguish between the short-term and the long-term absences. Short-term absences are more voluntary than the long ones. The motivation for such absences can be interpreted as maximising one's leisure time under a reasonable budget constraint. The long-term absences, on the other hand, are often related to 'involuntary' causes. A typical example is the sick-leave. Thus, in total, we will construct four types of models: short-term incidence, short-term recovery, long-term incidence and long-term recovery.

Comparing to the conventional econometric tools used in work absenteeism: the count data regressions (Delgado & Kniesner 1997) and the duration models (Barmby et al. 1991; Markussen et al. 2011; Fevang et al. 2014), the self-exciting process has the advantage of modelling the dynamic decision-making process. In the count data regression literature, the study subject is the counts of events during a period. Thus, count data models lose the dynamic information by aggregating the absence records over the defined period. Duration models often assume that absence durations are i.i.d, which is incompatible with the state-dependent setting. Lagged duration models (Honoré 1993) do exist, but they are difficult to apply.

Structural econometric models can be used to include the state-dependence. However, it would be quite complicated when one tries to model four decision making processes (short (long)-term ask for leaves, short (long)-term return to work) with only one economic model.

The modelling strategy we used (i.e., the separation of incidence and recovery decision making processes and the distinction between short-term and long-term absences) requires a raw absence records dataset, in which the researchers should have access to the details of each absence, including the beginning and ending dates as well as necessary individual demographic information. In our empirical study, a firm-level administration dataset is used. We will formally introduce the data in the later section.

This paper contributes to two strands of literature. First, we provide substantial evidence on the existence of dynamic behavior in the work absences. Specifically, we observe workers dynamically make absence decisions in short-term absences in both incidence and recovery processes. While in the long-term absences, such dynamic behavior plays an insignificant role in the decision making processes.

Second, we provide a modelling method that complements to the conventional methods (structural models, count data regression and duration analysis models.) Instead of focusing on the events per se, we try to model the whole behavior process. Thus we require i.i.d of individuals, but not the events.

The paper is structured as follow. Section 2 introduces the data and provides some preliminary results based on conventional count data regression and duration models. The aims of these preliminary results are mainly to show the existence of strategic behavior in work absences, to highlight the incompatibility of the conventional methods and to illustrate the nature of the problem. In section 3, we first introduce some notations and basics about the self-exciting process, followed by the presentation of our model. We also discuss the difficulties to include the unobserved heterogeneity in the model and some workarounds. In section 4 the estimating results are presented and discussed. Based on the empirical findings, we develop a simple economic model in section 5. Section 6 compares the self-exciting process to count data regression and duration analysis models. Finally, section 7 concludes the whole paper.

3.2 Data and Preliminary Results

In this section, we briefly introduce the data and present some preliminary results based on conventional count data regression and duration models. Following the procedure proposed by Heckman (1981), we also provide some evidence that support the existence of state dependence in the data. At the end of this section, we will illustrate the nature of the work absenteeism problem.

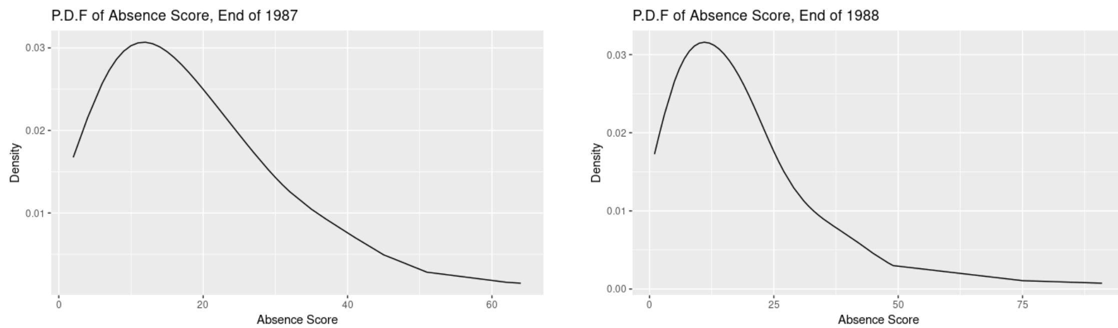
3.2.1 The Data

The data we used come from a UK based manufacturing firm, which produces a homogeneous product using production lines. Other publications that use the same data (or a subset of the data) are Barmby et al. (1991), Barmby et al. (1995), etc. In 1983, the firm introduced an experience rated sick-pay scheme where workers with less cumulative absence scores receive a better sick-pay benefit. More specifically, the scheme provides the sick-pay benefit at three levels: Grade A workers are paid with their full normal wage including bonuses less the statutory sick-pay (SSP) of the UK social security; Grade B workers are paid with their basic wages less SSP; Grade C workers receive no benefits from the firm. All the workers are eligible to the SSP.

To be eligible to the SSP, workers should be absent from work for more than three consecutive days. Because of this requirement, we define the short-term absences as the ones whose duration are less or equal to 3 days, and the others are categorised as long-term.

Workers are categorised into these three grades based on the absence records over the previous two years: at any given time, individuals need to consider both last year's and current year's absence scores, since these scores will decide next year's benefit. Each day of absence attracts a certain number of 'points', mostly 1 point, depending on the cause of this absence. To simplify our analysis, we assume that one day off is 1 point of absence score. 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.

We believe there is no abnormal behavior occurs around the cut-off points 21 and 41. To show it, we non-parametrically estimate the absence score density function at the end of the year 1987 and 1988. Figure 4.1 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, the last year's absence records (1988) to determine 1989's sick pay benefit is also the middle year to determine the benefit for 1990, hence the absence score is updated in a 'smooth' way, and 3) the absence score will only affect the sick-pay benefit (which is stochastic: only receive the benefit when ill) not the salary (which is deterministic), hence the incentive to 'control' the absence scores around the cut-off points are not strong.



(a) At the end of 1987 (b) At the end of 1988

Figure 3.1: Non-parametric P.D.F of absence scores

A worker's decision to be absent will not only lead to a loss of earnings[1] but also affect the eligibility for the sick-pay at some point in future, usually in a stochastic fashion. The incentives to take a leave and to return to work from an absence created by this scheme are complex and raise challenges for econometric modeling.

The data consists of detailed absence records: the beginning and ending dates of absences, type of absences (sick-leave, maternity release, jury service, work accident etc.) as well as individual characteristics such as age, gender, contract type, etc. In this paper, we will deal with the 'working age', which is the real age subtracts the legal working age (16 in the UK). Some common covariates such as education, wage and job hierarchy are not included in this dataset. However, we do not think these covariates could play significant roles: most workers are blue-colour, who have similar education backgrounds,

[1] That is not the case for class A workers whose benefit will not be affected during an absence. However, for the other two classes, some loss of income is a certain.

receive similar wages and their job levels are more or less the same. We use the data from calendar year 1987 to 1988. In total, we have 878 workers with 5718 absence records.

Figure 3.2 shows the histogram distribution of the length of absences. Among all the absences, 1-day off leaves account for more than half. Around 78.1% are short-term absences. Long-term absences, especially those longer than ten days are rare.

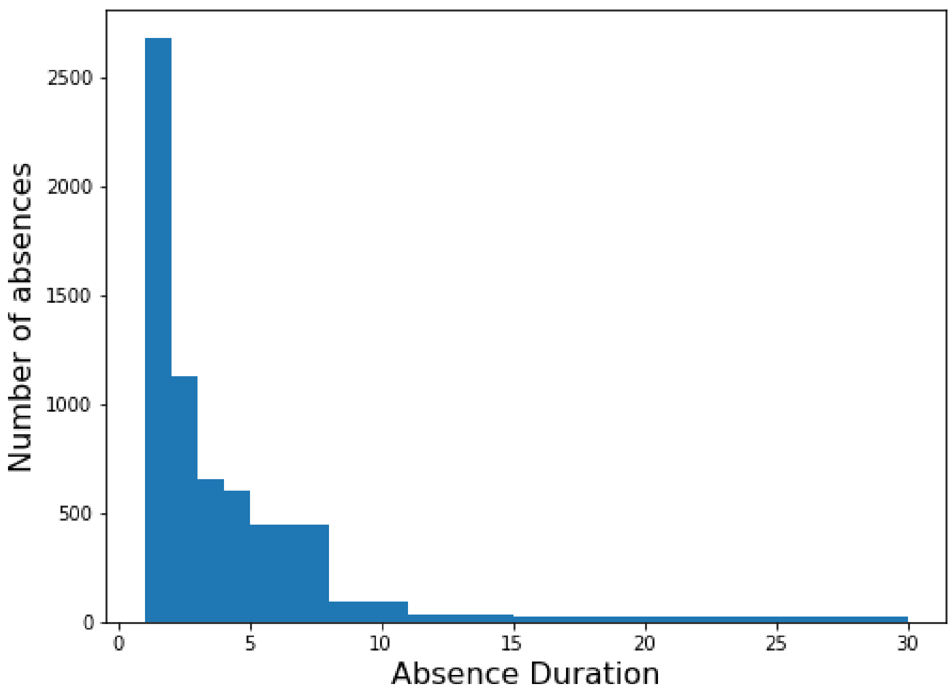


Figure 3.2: Most frequent absence durations

3.2.2 Preliminary Results

Conventionally, count data regression (Delgado & Kniesner 1997) and duration models (Markussen et al. 2011) are commonly used in the analysis of sickness absences. In this subsection, we provide some preliminary results using these methods.

The subject under study in the count data regressions is the counts of occurred events over a period. In our application, this subject would be the number of absence records in the year 1988. We use four count data regressions: the Poisson, the negative binomial, the zero-inflation and the hurdle models. The Poisson regression is the basic model for count data analysis. One restriction to this model is the equidispersion assumption: the mean of the counts must be equal to the variance. To overcome this restriction, researchers have proposed more general over-dispersion model. Negative binomial model is particularly popular. One source of over-dispersion is the excess of zeros. Two models are often used to deal with this property: zero-inflation and hurdle models. The general idea is first to use a binomial distribution describing the zeros and then to use another probability distribution to describe positives. In the zero-inflation model, the second probability distribution can generate both zeros and positives. While in the hurdle model, this probability distribution is truncated at zero. We left the technical description of these count data regression models in the Appendix. One important trait that we include in the models are the absence counts in the previous year (1987). The goal of this trait is to obtain some insights on how past experiences could alter future decisions.

Table 3.1 summarizes our count data regression results. One crucial explanatory variable is the number of times of absences in 1987, which are used as an approximation of heterogeneities of individuals. The results are quite similar across different models. This conclusion is consistent with previous literature (Delgado & Kniesner 1997).

Another commonly used tool is the duration analysis. Here, we study the duration of attendance until the first absence in a year. The workhorse in the duration analysis is the hazard rate, which is the ratio of the probability density function to the survival

Table 3.1:: Count Data Regression Results

	Dependent variable:			
	count88			
	Poisson	*negative binomial*	*hurdle count part*	*zero-inflated count part*
	(1)	(2)	(3)	(4)
age	-0.005	-0.006	-0.017	-0.005
	(0.011)	(0.016)	(0.012)	(0.012)
age2	0.007	0.008	0.015	0.0002
	(0.014)	(0.019)	(0.014)	(0.016)
male	-0.249***	-0.224***	-0.230***	-0.236***
	(0.045)	(0.065)	(0.046)	(0.048)
full	0.104**	0.115	0.094*	0.115**
	(0.049)	(0.074)	(0.050)	(0.052)
marriage	-0.066	-0.076	0.002	-0.011
	(0.052)	(0.075)	(0.056)	(0.059)
count87	0.131***	0.156***	0.086***	0.101***
	(0.005)	(0.008)	(0.006)	(0.006)
Constant	0.944***	0.866***	1.565***	1.234***
	(0.193)	(0.284)	(0.205)	(0.220)
Observations	874	874	874	874
Log Likelihood	-1,991.314	-1,878.365	-1,965.877	-1,940.922
θ		3.445*** (0.383)		
Akaike Inf. Crit.	3,996.627	3,770.731		

$age2 = age^2/100$. *p<0.1; **p<0.05; ***p<0.01. This table presents the four counting data regression results. For the zero-inflation and hurdle models, we only present the count parts. The dependent variables are the counts of absences in the year 1988. One important trait is the counts of absences in the previous year, which is positively correlated with the dependent variable. It would be wrong to interpret the results as causal, since otherwise it implies the work discipline regulation play exactly the opposite role: encourage more absences. Instead, this trait should be interpreted as the approximation of the unobserved heterogeneity.

function. It can be interpreted as the failure rate or the force of mortality. We study a baseline duration model, where the hazard rate is constant over time and no presence of the unobserved heterogeneity. Appendix documents the details of this model. The first column in Table 3.2 reports the estimation results of this standard duration model.

We also study a more commonly used duration model where the unobserved heterogeneity is introduced. This hazard rate has a multiplicative form of the unobserved heterogeneity term, a random variable, and the remaining part. As proposed by Heckman & Singer (1984), we use discrete distribution to approximate the true random variable distribution, and obtain the non-parametric maximum likelihood estimator (NPMLE). Detailed description about this extension model as well as the NPMLE can also be found in Appendix. Note that one requirement to use the NPMLE is the independent of the unobserved heterogeneity with all other covariates. In our application, this is clearly violated, as the absence counts in the previous year is correlated with the heterogeneity term. Nevertheless, we still present the results.

Column 2 of Table 3.2 presents the estimation results for this model. The log-likelihood value for two mass points and three mass points are almost the same and the probability associated with the third mass point is close to zero. Based on these information, we believe, two mass points would be good enough.

Count data regressions and duration models are incapable of studying the strategic behavior. For count data regressions, the information is aggregated at the end of one year. Hence the dependent structures among events are lost. For duration models, one needs to maintain the events independence assumption. Thus, by design, the duration models assume that past events are uncorrelated with future ones. Notice some

Table 3.2:: Duration Analysis Results

	Dependent variable:	
	duration	
	Standard	*Heckman & Singer*
age	-0.028***	-0.068***
	(0.006)	(0.026)
age2	0.047***	0.094***
	(0.010)	(0.032)
male	-0.115	-0.133
	(0.093)	(0.113)
full	0.163	0.147
	(0.105)	(0.133)
marriage	0.076	0.119
	(0.099)	(0.127)
count87	0.254***	0.264***
	(0.010)	(0.013)
Observations	878	878
Log Likelihood	−248.668	-224.8397
χ^2	576.961*** (df = 5)	
Number of Mass Points		2

Note: This table presents the duration analysis results. The subject under study is the attendance duration before 1988's first absence. No short and long-term absence distinguishing in this table. Heckman and Singer's NPMLE is employed to approximate the distribution of unobserved heterogeneity. We found 2 mass points are good enough. The counts in 1987 is positive, indicating that the more absences in the previous year, the higher the likelihood to ask leaves. This result can not be interpreted as casual, instead, it approximate the unobserved heterogeneity. *p<0.1; **p<0.05; ***p<0.01

multiple-spell models break the independence assumption and allow lagged duration dependence (Honoré 1993). However, this lagged duration model is in a panel setting, and its hazard rate can be very difficult to study, since one needs to separate the state dependence from the unobserved heterogeneity. This difficulty is even more intimidating when one distinguishes short-term and long-term absences, as these two panels are shocks for each other.

From the preliminary results of both count data regression and duration analysis, we have seen that the counts of previous year's absences are positively correlated with the dependent variable (count data regression) and the hazard rate. It would be wired to interpret these results as causal since it implies the more absences one took last year, the more absences one would ask for in this year; or the more absences one took last year, the higher hazard (hence, the shorter the attendance duration) one would have. This interpretation contradicts the intention of the firm's absence benefit program. The proper interpretation of this trait should be an approximation of heterogeneities of individuals: frequent-absence workers tend to have more absences all the time, while less frequent workers should have fewer absence records in the future.

3.2.3 The Nature of the Problem

In this subsection, we try to illustrate the econometric challenges when modelling the strategic behavior. In order to provide a better understanding, consider Figure 3.3, which shows a possible realisation of work absences. The dash lines here are absence periods, and the solid lines are the attendance periods. Lower case 's' and 'r' are the starting and recovery dates of a short-term absence respectively, and upper case 'S' and

'R' are the starting and recovery dates for a long-term absence. In this example, we have two short-term absences before a long-term absence.

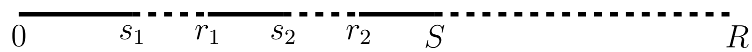


Figure 3.3: A possible realization of absences

Suppose now we are at $t \in [r_1, s_2)$ and the goal consists of investigating how likely is next absence at time $t + dt$. To account for the strategic behavior, one needs to include previous absence records. In our application, it is the cumulative absence time that matters most. Hence $d_1 = r_1 - s_1$ should be in the model. We would expect the coefficient of the cumulative absence time is negative if the firm's absence benefit program is working. That is, larger cumulative absence time will discourage any further absence behavior.

Besides, we would also like to investigate whether asking for leaves is duration dependent. Therefore, we need to include a time dependence term $t - r_1$.

The distinction between short term and long term absences also creates a challenge. Since the economic motivations behind these two different absences are disparate, one should model them separately. However, the cumulative absence time is the summation of these two. Thus these two models depend on each other, creating a quasi-simultaneous equation system.

At this moment, it is evident that conventional micro-econometric tools such as count data regression and duration analysis offer no satisfying solution to this problem. The nature of this strategic behavior problem is the state dependence. In the next section, we are going to introduce the self-exciting process, that by definition is state dependent and can model the strategic behavior among work absences.

3.3 Econometric Models for Work Absenteeism

3.3.1 The Work Absenteeism Models

For each individual i, define an incidence counting process that records all the starting dates of absences:

$$N_i^1(t) = \sum_{j=1}^{\infty} I\{t_j \leq t\} \tag{3.1}$$

define a recovery counting process that stores all the ending dates of absences:

$$N_i^2(\tau) = \sum_{j=1}^{\infty} I\{\tau_j \leq \tau\} \tag{3.2}$$

We use t and τ to distinguish the times of beginning and ending dates of absences. Thus the i^{th} absence duration is just $d_i = \tau_i - t_i$. The time here is in terms of years: $t \in [0, 2]$ and $\tau \in [0, 2]$ (Two years data).

Recall that by the eligible condition for the SSP, we category any absences that is less or equal to 3 days as short-term and other absences as long term. Define three alternative states, $k = 1, 2, 3$, that an individual can occupy in our model: attendance ($k = 1$), short-term absence ($k = 2$) and long-term absence ($k = 3$). Thus $\lambda_{12}(t)$ is the short-term incidence intensity function (from attendance state to short-term absence), and $\lambda_{21}(t)$ is the short-term recovery intensity function, other two long-term intensity functions follow the same index rule.

Furthermore, we define attendance periods as any time intervals between the last recovery dates and the next starting dates of absences. Define absence periods as any time intervals between the starting dates and the recovery dates of absences. Figure 3.3 in section 2 describes the situation. We assume that a new absence cannot occur without the end of current absence. That is the incidence intensity is zero in the absence period. Similarly, recovery events cannot occur before any absences ever started: the recovery intensity is zero in attendance period.

A Model for Incidence Processes

The self-exciting has the advantage of including history information, but if there is no previous absence records (both long term and short term), one might instead use the duration analysis and study this constant hazard rate:

$$h(X_i, \nu_i) = \exp(\nu_i) exp(\mathbf{X}_i' \gamma_{1k}) \tag{3.3}$$

where $\mathbf{X}_i$ is a vector of covariates, and the random variable ν is used to represent the heterogeneity in the intensity. As usual, we will use the Heckman and Singer's NPMLE to handle the random effect term. Notice that for individuals that have no absence records during the investigation period, we may treat them as censoring individuals in this duration analysis.

If previous absence records exist, for individual i, the overall incidence intensity is specified as:

$$\lambda_{i,1k}(t) = \begin{cases} \lambda_{1,k}(\mathbf{X}_i)\lambda_{2,k}(t)\Big(\lambda_{3,k}(t) + \lambda_{4,k}(t)\Big), t \in \text{attendance period} \\ 0, t \in \text{absence period} \end{cases} \tag{3.4}$$

Notice here we allow both short-term and long-term incidence intensities to be positive during the same attendance period. Furthermore, we assume that conditional on current common filtration, the occurrences of short-term and long-term incidences are independent. That is when studying the short-term (long-term) incidence process, we consider absences are due to short-term (long-term) causes and long-term (short-term) causes are independent censoring. This assumption resembles the cause-specific hazard model rather than the competing risk model in the duration analysis.

The four components are:

$$\lambda_{1,k}(\mathbf{X}_i) = exp(\mathbf{X}_i'\gamma_{1k}); k = 2, 3$$

$\lambda_{1,k}(\mathbf{X}_i)$ contains all the time-invariant covariates such as age,gender and labour contract status (full time or part-time). The exponential form guarantees the intensity is positive, and it is also commonly used in duration analysis.

$$\lambda_{2,k}(t) = exp(\beta_{1k}H_i(t)); k = 2, 3$$

$\lambda_{2,k}(t)$ governs the response of one worker to her own cumulative absence time $H_i(t)$ (in terms of days). β_{1k} are our primary interested parameters. We expect they are significantly negative (at least for the short-term absences) if the strategic behavior plays some roles in the decision making processes.

$$\lambda_{3,k}(t) = 1 + |\alpha_{1k}|exp(\alpha_{1k}(t - \tau_{N_i^1(t-)})); k = 2, 3$$

Some absences might trigger further absences (e.g. minor illness might lead to a second doctor-visiting). We use $\lambda_{3,k}(t)$ to measure the time dependence since previous recovery date. The form of $|\alpha_{1k}|exp(\alpha_{1k}(t - \tau_{N_i^1(t-)}))$ is mainly for the convenient of integration.

Lastly, we need one part to measure the individual's response to Mondays and Fridays. One would expect workers tend to ask for leaving more frequently on Mondays or Fridays, as along with weekends, it would generate three consecutive off-duty days. One straightforward modeling strategy is to use indicators for Mon/Fridays, but this will create sudden jumps in the intensity function. When integrating w.r.t time to obtain the cumulative intensity, such indicators would be lost as they have zero Lebesgue measure in a continuous time framework. Thus we need a periodic continuous function with peaks on Mon/Fridays. In the end, we choose the sine function:

$$\lambda_{4,k}(t) = a_{1k}(1 + sin(b + c_{1k}t)); k = 2, 3$$

We set $c = 327.6$ such that the distance between two peaks in the sine function is equal to 7/365 years, or one week's time, we would expect $b \approx 2.5$ to match Monday/Friday's location and a to be significant if our hypothesis is correct. Figure 3.1 illustrate the idea.

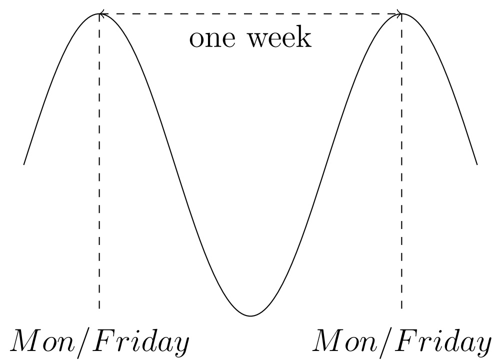


Figure 3.1: Mon/Friday Sine Function

The additive structure between $\lambda_{3,k}(t)$ and $\lambda_{4,k}(t)$ is mainly for the simplicity of integration.

A Model for Recovery Processes

The recovery intensity has the following parts:

$$\begin{aligned} &\lambda_{5,k}(X_i) = exp(X_i'\gamma_{k1}) \\ &\lambda_{6,k}(\tau) = |\beta_{k1}|exp(\beta_{k1}H_i(\tau)) \\ &\lambda_7(\tau) = 1 + |\beta_{k2}|exp(\beta_{k2}(\tau - t_{N^1_{i,13}(\tau-)})); k = 2, 3 \end{aligned} \tag{3.5}$$

these parts have similar roles as in the incidence intensities: $\lambda_{5,k}(X_i)$ contains the individual's covariates, $\lambda_{6,k}(\tau)$ measures an individual's response to the absence score and lastly $\lambda_7(\tau)$ captures the duration dependence effect.

Notice in $\lambda_{6,k}(\tau)$, the structure is different than $\lambda_{2,k}(\tau)$. This difference is for the convenient of integration. Unlike in the incidence intensity, the absence score is fixed for each attendance period (using Figure 3.3's notation, $H(t_1) = H(t_2), t_1, t_2 \in [r_{j-1}, s_j)$), in the recovery intensity, during each absence period, the absence score evolves continuously

$(H(t_2) = H(t_1) + t_2 - t_1, t_1 < t_2 \in [s_j, r_j))$. When integrate with respect to time, the structure of $\lambda_{6,k}(\tau)$ facilitates the computation.

The recovery intensities are specified as:

$$\lambda_{i,k1}(\tau) = \begin{cases} \lambda_{5,k}(X_i)(\lambda_{6,k}(\tau) + \lambda_7(\tau)), \tau \in \text{k-term absence period} \\ 0, \tau \in \text{Otherwise} \end{cases} \tag{3.6}$$

We assume that once an individual asks for a leave, she immediately knows the type of absence (short or long-term). Hence, for each absence period, only one type of the recovery intensity could be positive.

Notice that unlike the incidence processes, the recovery processes are by design conditioned on the existence of occurrence of absences. Thus they always have history information such as $H_i(\tau)$ and $\tau - t_{N^1_{i,13}(\tau-)}$.

State dependence assumption in the incidence processes is verified in the previous section. To verify the state dependent structure in the recovery intensity, one notices that if the state-dependent hypothesis is correct, coefficients of the cumulative absence time $\beta_{k1}, k = 2, 3$ should be significantly away from zero. If one (or both) of the recovery processes do not show the state dependent structure, it would be plausible to assume that the absence duration are i.i.d. In that case, a standard duration analysis would be a useful modeling alternative.

The Heterogeneity Issue and Workarounds

So far, our models have not addressed the issue of the unobserved heterogeneity. In this subsection, we briefly discuss the difficulties of including the heterogeneity in the model and some workaround methods.

Considering the following intensity,

$$\lambda_i(t|\nu_i, \mathcal{F}^i_{t-}) = \nu_i \lambda_0(t|\mathcal{F}^i_{t-})$$

where $\nu \sim G(\nu)$ is the unobserved heterogeneity with distribution G. Note that the cumulative intensity $\Lambda_i(t|\nu_i, \mathcal{F}^i_{t-}) = \nu_i \Lambda_0(t|\mathcal{F}^i_{t-})$ is not predictable w.r.t $\mathcal{F}^i_{t-}$. Hence $N_i(t) - \Lambda_i(t|\nu_i, \mathcal{F}^i_{t-})$ can not be a martingale.

One may try to integrate out with respect to ν in order to get rid of the unobserved heterogeneity as in the mixed proportional hazard (MPH) model. However, this strategy is difficult without assuming that ν is uncorrelated with the filtration, which is hard to be held in the self-exciting framework: We are conditional on past events, which, by construction, are correlated with ν.

Even one overcomes the integration problem, the difference between the observed counting process and the marginal cumulative intensity is not a martingale. By the uniqueness of the Doob-Meyer decomposition, the observed counting process is not paired with the marginal cumulative intensity.

In general, it is hard to distinguish the state dependent effect and individual heterogeneity. Heckman (1991) concluded that 'The ability to distinguish between heterogeneity and duration dependence in single spell duration models rests critically on maintaining explicit assumptions about the way unobservable and observables interact.' and '... Economically extraneous statistical assumptions drive the answer...Viewed as the prototype for identification in general nonergodic models, these results are not encouraging.' Nerlove (2014) holds a similar view: 'Unfortunately, in my view, in the more than 35 years since the Paris Conference in 1977 no solution has been found to the general problem of distinguishing between "the hidden hand of the past" and individual heterogeneity.'

Approximating the Heterogeneity

One workaround is to use the history information to approximate the unobserved heterogeneity. In the short-term incidence process, the primary unobserved heterogeneity is an individual's working attitude, while in the long-term incidence process, the primary unobserved heterogeneity is one's health status. We may approximate these two terms using the (moving) average attendance duration $\tilde{d}(t)$ along with some absence score adjustments.

Using Figure 3.3 and its notation to help define $\tilde{d}(t)$:

$$\tilde{d}(t) = \frac{\sum_{i:r_i \le t} s_i - r_{i-1}}{\#\{i : s_i \le t\}}$$

Notice here, we only consider the short-term attendance duration in the short-term process and vice versa for the long-term process.

We assume $\tilde{d}(t)$ has the following structure:

$$\log(\tilde{d}(t)) = I(t) + G(H(t)) + \epsilon$$

where $I(t)$ is the incidence index at time t. A higher index suggests a more hard-working individual in the short-term process (or better health condition in the long-term case). $G(\cdot)$ is an increasing function of the absence score $H(t)$. ϵ is a random variable with zero mean. This structure indicates that a hard-working individual on average tends to have longer work attendance period. In the meantime, a higher absence score also suppress one's further absences, generating a longer work attendance period. Notice that, unlike the conventional method, where the unobserved heterogeneity is assumed to be time-persistent, we express the working attitude to be time-variant. We argue such assumption is more realistic: as time goes by, one's attitude might be 'modified' or 'educated' by firm's absence policy such that it evolves constantly.

We approximate the index $I(t)$ by $\tilde{I}(t)$:

$$\tilde{I}(t) = \log(\tilde{d}(t)) - G(H(t)) = I(t) + \epsilon$$

In practice, we let $G(H(t)) = \log(1 + H(t))$, and to make sure $\log(\tilde{d}(t))$ has mathematical meaning in the case of $\tilde{d}(t) = 0$, we replace it as $\log(1 + \tilde{d}(t))$.

We modify the $\lambda_{1,k}$ as:

$$\lambda_{1,k} = exp(\mathbf{X}_i' \gamma_{1k}) exp(\gamma' \tilde{I}(t))$$

the structural of the overall incidence intensity remains the same.

In the long-term recovery processes, the primary unobserved heterogeneity is an individual's ability to recovery, while in the short-term, it is the willingness to return. Similarly, we could use history information to approximate them. This time, we choose the (moving) average recovery time $\tilde{c}(t)$, which is defined as (again, use Figure 3.3's notation),

$$\tilde{c}(t) = \frac{\sum_{i:r_i \leq t} r_i - s_i}{\#\{i : r_i \leq t\}}$$

We also only consider the short-term absence duration in the short-term process and vice versa for the long-term process.

Assuming $\tilde{c}(t)$ has the following structure:

$$log(\tilde{c}(t)) = R(t) - G(H(t)) + \epsilon$$

where $R(t)$ is the recovery process index with decreasing order: a higher recover index means a longer recovery time in the long-term process (or a less willing to return in the short-term case). $G(\cdot)$ is an increasing function of the absence score $H(t)$ and ϵ is a random variable with zero mean. As usual, we may approximate $R(t)$ by $\tilde{R}(t)$:

$$\tilde{R}(t) = log(\tilde{c}(t)) + G(H(t)) = R(t) + \epsilon$$

Just like before, we set $G(H(t)) = log(1 + H(t))$.

We modify $\lambda_{5,k}$ as:

$$\lambda_{5,k} = exp(X_i' \gamma_{k1}) exp(\gamma' \tilde{R}(t))$$

and the overall recovery intensities structure remain the same.

To maintain the differences between counting processes and their cumulative intensities are martingale, we need to assume that the true measurements used by individuals for working attitudes and recovery abilities are $\tilde{I}(t)$ and $\tilde{R}(t)$ respectively instead of $I(t)$ and $R(t)$.

Group Heterogeneity

Another workaround is to assume group heterogeneity instead of individual heterogeneity and to reveal the unobserved heterogeneity through an external model. Recall in the incidence processes, the primary unobserved heterogeneity is the individual's working attitude, which is correlated with the number of absences: a group of hard-working individuals have in general fewer absences, while less hard-working individuals tend to have more absences. We may then build and estimate a finite mixture count data model, and use the Bayesian rule to 'reveal' individual's group affiliation.

Now assume individuals belong to k different groups. Each individual's group affiliation is of course unobserved by the researchers. Assume the numbers of absences $\mathbf{y} = (y_1, \cdots, y_N)'$ over a period are governed by a finite mixture Poisson:

$$p(\mathbf{y}|\Theta) = w_1 f_1(\mathbf{y}|\Theta_1) + w_2 f_2(\mathbf{y}|\Theta_2) + \cdots + w_k f_k(\mathbf{y}|\Theta_k)$$

where $\Theta = (\Theta_1, \Theta_2, \cdots, \Theta_k, \mathbf{w})'$ denotes the vector of all parameters, $\mathbf{w} = (w_1, w_2, \cdots, w_k)'$ is a vector of weight whose elements are restricted to be positive and sum to unity. $f_k(\cdot|\Theta_k)$ is a Poisson density with the vector of parameters Θ_k.

Additionally, we may equivalently model the finite mixture model in a hierarchical manner using a latent variable l_i, which represents the allocation of each observation y_i to one of the components:

$$
\begin{aligned}
&p(y_i|\Theta_k, l_i = k) = f_k(y_i|\Theta_k) \\
&p(l_i = k) = w_k
\end{aligned}
$$

One may have the group affiliation posterior by the Bayesian rule:

$$
\begin{aligned}
p(l_i = k|y_i, \Theta_k) &= \frac{p(y_i|l_i = k, \Theta_k) * p(l_i = k)}{p(y_i|\Theta_k)} \\
&= \frac{p(y_i|l_i = k, \Theta_k) * w_k}{\sum_{k=1}^{K} p(y_i|l_i = k, \Theta_k) * w_k}
\end{aligned}
$$

we may then assign the group affiliation according to the posteriors.

Admittedly, this workaround is not perfect. The choice of k is somehow arbitrary. The classification of groups in the finite mixture model is 'fuzziness': a certain observation y_i has probability w_k to belong to component k. However, we 'force' each observation to fit into one group by posterior, and modify the fuzzy classification into a sharp classification. In this case, some information loss is inevitable. In addition, this method only works for incidence processes. Since for the recovery processes, the primary unobserved heterogeneity is one's recovery ability, which can not be represented as counts.

Thus, we will use the group heterogeneity as a robustness tool for the incidence intensities against different heterogeneity assumptions.

3.4 Main Results

In this section, we present the estimation results along with the discussions. We first report the results for incidence processes where the decisions of asking for leave is modelled.

3.4.1 Incidence Processes Estimation Results

We will first present the results for absences that have no previous absence records (including both short term and long term records). The subject under study is the attendance duration, that is the time intervals individuals took to ask for their initial absences. Two groups of individuals may have such absences. Individuals that have no absence records in the past but have absences during the investigation period naturally fit this situation. Although we do not have exact information, we suspect these kind of individuals are most likely to be newly hired workers. The second group consists of individuals who have no absence records in our investigation period as well as in the past. And we shall treat them as censored.

Table 3.1 reports the duration analysis results using the likelihood function mentioned in the previous section. What surprises us is the lack of heterogeneity in the data: the log-likelihood values of one mass point and two mass points are incredibly close in both short term and long term cases. Another evidence that supports no heterogeneity is that when we include two mass points, the standard errors are relatively large, a sign of too many mass points (Greene & Hensher 2010).

Table 3.1:: Duration Analysis for Attendance before Initial Absence

	short term,k=2	*short term,k=2*	*long term,k=3*	*long term,k=3*
	(1)	(2)	(3)	(4)
age	-0.0392***	-0.0696*	-0.0654***	-0.1000*
	(0.0138)	(0.0402)	(0.0195)	(0.0529)
age2	0.0464**	0.0817*	0.0841 ***	0.1240**
	(0.0203)	(0.0484)	(0.0280)	(0.0633)
male	0.0834	0.0356	0.3175	0.1184
	(0.2253)	(0.2302)	(0.3461)	(0.3416)
full time	0.0703	-0.0039	0.4345	0.3861
	(0.2402)	(0.2561)	(0.3563)	(0.3602)
married	0.0357	0.0873	0.0894	0.1391
	(0.2015)	(0.2128)	(0.2570)	(0.2679)
Log Likelihood	-257.0000	-256.6635	-174.5000	-174.2110
Number of Mass Points	1	2	1	2

Note: The subject under study is the attendance duration before the initial absences. Here the initial absences are defined as the ones that when ask for leaves, the absence scores are zeros. Heckman and Singer's NPMLE is employed to approximate the distribution of unobserved heterogeneity. Column (1) and (2) report the results when we have one and two mass points for the attendance duration before short-term absences. The log-likelihood values are similar, indicating little heterogeneity; Column (3) and (4) report the results when we have one and two mass points for the attendance duration before long-term absences. The log-likelihood values are again similar, indicating litter heterogeneity. Absence duration less or equal to 3 days are categorized as short term, others are long term. $age2 = age^2/100$. *p<0.1; **p<0.05; ***p<0.01

Next, we present our main results for incidence intensities in Table 3.2. The first two columns are the results using heterogeneity approximation for short and long term absences, and the last two columns are the results using group heterogeneity. As mentioned before, our primary focus would be the ones employing the heterogeneity approximation. Group heterogeneity results are presented for robustness check purpose. To streamline the presentation, we postpone the group heterogeneity analysis in the next subsection.

The most important parameters, of course, are the $\beta_{1k}, k = 2, 3$, which are the coefficients of the absence scores for short-term and long-term incidences respectively. β_{11} is significantly less than zero while β_{12} is not. Such results suggest that the strategic behavior only exist in the short-term absences: as the absence time cumulates, workers are discouraged to take short-term absences. While in the long-term absences, these absence scores seem to be out of the decision-making processes.

Other estimators also suggest that the short-term and long-term incidence decision-making processes are entirely different. For example, in the short-term, workers are more likely to ask for leave on Monday or Fridays. This phenomenon is understandable, since along with weekends, individuals may have three consecutive off-working days. Such strategic behavior strongly indicate that short-term absences are more likely to be 'voluntary', where there is a trade-off between working time and leisure time to maximise the utility. As it could be expected, Monday/Fridays are not significant in the long-term absences, which is consistent with the 'involuntary' leave hypothesis.

In both short and long-term leaves, age is an essential element. In the short term, the general trend is to increase the intensity first and then decrease it. The peak is around 13.5 working age or 29.5 years old. The trend for long-term leaves is quite the

Table 3.2:: Incidence Intensities

	Approx. Heterogeneity		*Group Heterogeneity*	
	short term	*long term*	*short term*	*long term*
	(1)	*(2)*	*(3)*	*(4)*
β_{1k}	-0.05734195***	0.002551	-0.03207249***	-0.02183574
	(0.007313)	(0.0108902)	(0.0072219)	(0.0139920)
α_{1k}	-35.32423495**	-5.02947189	-36.90188525*	-4.92911781
	(17.09965)	(6.4427670)	(21.600208)	(7.2625815)
age	0.31746598***	-0.42761261***	0.24439097**	-0.37079208**
	(0.0639588)	(0.1598638)	(0.1135570)	(0.1448501)
age2	-1.17953689***	0.96998791***	-1.53408411**	0.86544108***
	(0.3328688)	(0.3161102)	(0.6521555)	(0.2636461)
male	-2.02277622	-0.34993766	-4.62557647	-0.39203022
	(1.512623)	(1.0142152)	(12.506801)	(1.0577588)
full time	1.2947023***	1.22844682	0.7890032***	1.33575203
	(0.438272)	(1.2074113)	(0.1905531)	(1.2328672)
married	-1.03290089***	1.33187185*	-1.50756787**	1.32788344*
	(0.350615)	(0.7269884)	(0.6975158)	(0.6999908)
Mon/Fri	2.01429447*	0.15542535	5.00469984*	0.1711705
	(1.142056)	(2.6100053)	(2.6209058)	(2.9969653)
b	2.57708555***	2.69547261	2.58967502***	2.71009241
	(0.547747)	(7.8254215)	(0.2458489)	(8.4098623)
Group 2	–	–	1.01837705**	–
	–	–	(0.4682334)	–
$I(t)$	-0.42075577***	0.03062177	–	–
	(0.095787)	(0.1093162)	–	–
Distance	0.128602	0.028854	0.146190	0.0295902

Note:

This table presents our main results for the incidence processes. Column (1) and (2) are using heterogeneity approximation, while column (3) and (4) use group heterogeneity assumption. Group affiliations are calculated using the posteriors from the finite mixture Poisson-2 models. $I(t)$ is the working attitude index. $age2 = age^2/100$. Absence duration less or equal to 3 days are categorized as short term, others are long term. β_{1k} are the coefficients of the absence score. α_{1k} are the coefficients of time dependent structure. *p<0.1; **p<0.05; ***p<0.01

opposite. It decreases first and then increases. The turning age is around 38 years. These results are reasonable and expected. Since youngsters value the leisure time much more than the elderly and are more likely to be involved in the voluntary short-term absences. They are also less likely to have major illness compare to senior workers.

Gender difference is insignificant in both short and long term cases. Full-time workers are more likely to have short-term absences compare to their part-time counterparts. Marriage plays an interesting role here. On the one hand, it serves as a stabiliser and reduces the short-term absences. On the other hand, when individuals need to ask for long-term absences, marriage seems to provide some protection against income loss during the absence period and increases the likelihood for asking leave. This case is particularly true if both spouses have jobs. Unfortunately we do not have information on this covariate.

Lastly, we have clear evidence for the time dependence in the short-term incidence process: more recent the last short absence contributes a higher propensity to ask for a short leave again. However, such time-dependent structure is not significant in the long-term incidence.

Robustness Check

For the group heterogeneity, we need first to pin down the number of groups k. Recall the results in Table 3.1, where the proper number of mass points in NPMLE is one. Therefore, we believe that $k = 2$ is reasonable. We then estimate the finite mixture Poisson model with two components and obtain the group affiliation posteriors. In the end, for the short-term absences, there are 450 individuals belong to Group 1 with average absence counts of 5.12 in the year 1988; 303 individuals are with Group 2, whose average absence counts in the same year is 7.19.

In the long-term absences, the group affiliation posteriors suggest only one group, indicating little heterogeneity among individuals. This result is somehow expected, as in the heterogeneity approximation case, the coefficient for the hard-working index is also not significant, a sign for homogeneity. We document the finite mixture Poisson model results in the Appendix.

Comparing column (1) and (3) and (2) and (4), we conclude that our results are quite robust against different heterogeneity assumptions. If one estimate is significant in the heterogeneity approximation case, it is also significant in the group heterogeneity case, the same pattern holds true for insignificant estimates. Second, as mentioned before, despite of heterogeneity settings, we have heterogeneity in the short-term absences, while in the long-term cases we end up with homogeneity.

We also estimate the incidence intensities without heterogeneity. Column (1) and (2) in Table 3.3 report the results. For the short-term incidence, the estimate for the absence score is smaller in absolute, but it is still negatively and significantly away from zero. This result indicates that the effect of the absence score dominates the effect of the unobserved heterogeneity. As for the long-term incidence, the absence score estimate is insignificant. Such result is expected since, in the heterogeneity-included results, the estimates for both absence score and heterogeneity term are insignificant.

Table 3.3:: Results When no Heterogeneity Term

	Incidence		*Recovery*	
	short term	*long term*	*short term*	*long term*
	(1)	*(2)*	*(3)*	*(4)*
β_{1k}	-0.02356803***	0.00487259	–	–
	(0.0070859)	(0.00980143)	(–)	(–)
α_{1k}	-35.32423437*	-5.02874541	–	–
	(19.987709)	(5.87600166)	(–)	(–)
β_{k1}	–	–	0.0010193***	-0.00538263***
	(–)	(–)	(0.0002110)	(0.0010231)
β_{k2}	–	–	-5.433983***	-0.60029543**
	(–)	(–)	(0.5936243)	(0.3020778)
age	0.231427***	-0.43991373***	0.171277	0.15071982***
	(0.0964381)	(0.1691635)	(0.1026832)	(0.0418187)
age2	-1.19406119**	0.9958262***	-0.396060**	-0.30651844***
	(0.5235543)	(0.3246395)	(0.1999274)	(0.0792001)
male	-2.01865347	-0.35511035	-1.271478**	2.79079758***
	(2.4986280)	(1.0628132)	(5570655)	(5163636)
full time	1.44430259***	1.23437316	-1.077283***	-1.69580522
	(0.5192177)	(1.1959571)	(0.3020042)	(2.2987771)
married	-1.00444757*	1.31971977*	4.0554886***	-1.65688059***
	(0.5257589)	(0.7042317)	(1.2183082)	(0.4934599)
Mon/Fri	2.03772383*	0.16219689	–	–
	(1.1983623)	(3.0323551)	(–)	(–)
b	2.57821261***	2.69541326	–	–
	(0.2861491)	(8.86651532)	(–)	(–)
Distance	0.142452	0.099502	0.028582	0.026130

Note:

This table presents the results when heterogeneity is not included in the models (both incidence and recovery). *p<0.1; **p<0.05; ***p<0.01

3.4.2 Recovery Processes Estimation Results

After a worker has asked for a leave, she has to decide the length of such absence. The recovery intensities portrait the counting processes that consist of all the ending days of absences. As in the incidence processes, we are also interested in discovering the differences between short-term and long-term recovery decision makings.

Caution is required for a series of scheduled absences as they will lead to a biased estimation if researchers ignore them. Unlike most absences in our study, where the decisions to be absent are due to some accidents, scheduled absences are triggered by some pre-existing events such as holiday arrangements. These specific events are known to everyone. Thus, full information is available when workers make these plans. The decisions to be absent and the duration of such absence will be made simultaneously. Thus the scheduled absence duration and the normal absence duration should come from different processes. For example, a worker might be based on her absence records and utility to decide whether to plan a leave just before the Christmas holiday. Moreover, she will at the same time pin down the duration of such absence (0 days for no absence).

Without further information, it is almost impossible to separate scheduled absences from normal ones. Our estimation results would be inevitably biased. One obvious way to reduce (but impossible to eliminate) the bias is to delete all the absences during the Christmas seasons.

Table 3.4 reports the results for recovery intensity for both short-term and long-term. In the short-term recovery intensity, the estimator of β_{21} is significantly higher than zero in both original and bias reduction estimations. It measures how an individual responds to absence scores in the recovery decisions: the longer cumulative absence time one has, the sooner this person will choose to return to work. This, however, is not the case in the long-term, where the response to the cumulative absence time is insignificant. These facts further confirm that short-term absences are more likely to be strategic while long-term absences are mostly associated with 'involuntary' causes. In the short-term, heterogeneity is insignificant, while in the long-term, the recovery ability heterogeneity do play a role.

Age is significant in both situations of short-term recovery. The general trend is first to increase the intensity of recovering and then decrease it. Compare to their female counterparts, males stay longer in short-term absences. Full-time workers tend to stay longer in the short-term absences. Lastly, married workers would return to work much quicker from short-term absences.

Column (3) and (4) in Table 3.3 report the result when heterogeneity is not considered. The short-term absence score estimate is still significant, which is expected as the heterogeneity is shown to be insignificant in the previous result. In the long-term recoveries, however, we observe a significant and negative absence score estimate. The reason is that larger absence score is correlated with the average recovery time (the larger the score is, usually the longer the average recovery time is), which is an approximation to the unobserved heterogeneity. Recall the long-term recovery results in Table 3.4, the absence score estimate is insignificant, but the unobserved heterogeneity estimate is significant. Thus, if heterogeneity is not included, the absence score serves as an approximation to the unobserved heterogeneity.

So far, we barely mention the covariates effects in the long-term recovery process. The reason is the insignificant of β_{31} may suggest that the long-term durations are

Table 3.4:: Recovery Intensities

	short term,k=2 *original* (1)	*short term,k=2* *holiday* (2)	*long term,k=3* *original* (3)	*long term,k=3* *holiday* (4)
β_{k1}	0.0001905***	0.0008113***	1.4540654*10^{-5}	9.0460202*10^{-6}
	(7.4389542*10^{-5})	(0.0002250)	(0.0003066)	(0.0002337)
β_{k2}	-0.2974196***	-5.4340119***	-2.6893905***	-1.2034950***
	(0.0401992)	(0.5001305)	(1.0069173)	(0.3709412)
$R(t)$	-0.0192418	-0.0247104	-0.4945683***	-0.6644247**
	(0.0321509)	(0.0283878)	(0.0943706)	(0.2661450)
age	0.3777196***	0.1760568*	0.1905671***	0.1039532***
	(0.1333167)	(0.0968915)	(0.0476012)	(0.0227340)
age2	-0.8573565***	-0.3947298**	-0.5215264***	-0.1656411***
	(0.2791424)	(0.1820732)	(0.1408734)	(0.0444712)
male	-0.5670663***	-1.2714456**	4.1771252***	4.6870273***
	(0.2009157)	(0.5571283)	(0.3325700)	(0.3821019)
full time	-2.2429315***	-1.0772366***	-0.2411186	-1.7283286*
	(0.5311679)	(0.2800427)	(0.3252895)	(0.9810031)
married	3.6464467**	4.0556782***	-1.5417979***	-2.1603153***
	(1.5006498)	(1.2085741)	(0.2369997)	(0.2075685)
Distance	0.100605	0.094431	0.061892	0.044357

Note: This table reports our main recovery processes results. The first and third columns are the results for the original dataset, where we did not delete the absences during the Christmas seasons. The results in second and fourth columns are estimated using Christmas-deleted dataset. Absence duration less or equal to 3 days are categorized as short term, others are long term. β_{k1} are the coefficients of the absence score, β_{k2} are the coefficients of time dependent structure. *p<0.1; **p<0.05; ***p<0.01

memoryless. That is, conditional on the occurrence of a long-term absence, the duration of such absences are independent. Notice that 1) the average recovery time is used as an approximation to the abilities of recovery of individuals; thus it does not necessarily mean the duration is state dependent. 2) The recovery process $N^2_{i,31}(\tau)$ is still state dependent, since $N^2_{i,31}$ does not contain full information about the long-term duration. The duration can only be constructed by using both incidence and recovery process(i.e., $d = \tau - t$).

It is then reasonable to assume that within individuals, each long-term recovery duration is i.i.d. A standard duration analysis could then be used to analyse such process.

For each individual, define the hazard rate and its cumulative hazard rate as:

$$\begin{aligned} h_i(X_i, \nu_i) &= exp(X_i\beta' + \nu_i) \\ H_i(T) &= h_i(X_i, \nu_i)T \end{aligned} \tag{3.7}$$

where t is the duration (not the time stamps used in the self-exciting processes), X_i is a vector of covariates of individual i, ν_i is the individual random effect.

The likelihood contribution for each individual is

$$L_i(\nu_i) = \prod_{j \in S_i} exp(-H_i(t_j))h_i \tag{3.8}$$

where S_i the set of observed long term durations for individual i. The fact that an absence has already occurred implies that we do not have the censoring problem here.

We will, again, use Heckman & Singer (1984)'s NPMLE. The likelihood function is

$$L = \prod_{i=1}^{N} \mathbb{E}[L_i(\nu_i)] = \prod_{i=1}^{N} \sum_{l=1}^{Q} p_l L_i(\nu_l), \text{ WITH } \sum_{l=1}^{Q} p_l = 1 \tag{3.9}$$

Table 4.1 indicates the results of this duration analysis. Most of the covariates are significant. Age has a similar pattern to the short-term recovery. As age increases, the propensity to go back to work first increases but then decreases. The peak is around a working age of 31. The reason, we believe, is the individual physiological conditions. Younger workers have better physiological conditions, which lead to a faster recovery process. As ageing occurs, one's physiological conditions decline. It makes harder and longer for a person to fully recover. Male workers on average return to work faster than their female counterparts. Married workers tend to recovery slower.

Overall, our estimated self-exciting intensities fit the data quite well. We plot the estimated averaged cumulative intensities against observed averaged counting processes to demonstrate the goodness of fit. Since we believe the long-term recovery process is unfit for self-exciting, we do not report its goodness of fit.

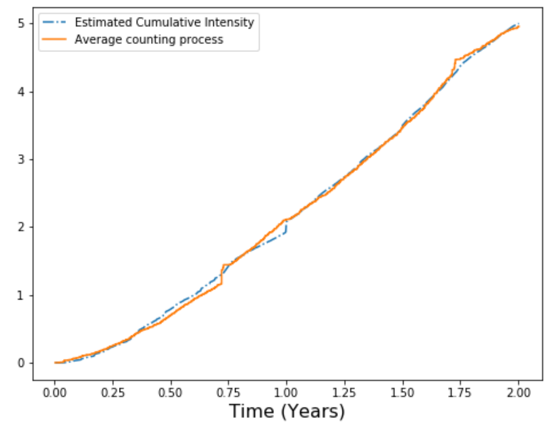
3.4.3 A Closer Look at the Strategy behavior Effect

In this subsection, we ask the question do individuals' attitudes towards the cumulative absence time changes as her seniority grows. To do so, we change the coefficient of cumulative absence time to a function of working age. Specifically, we modify $\lambda_{2,2}$ in the

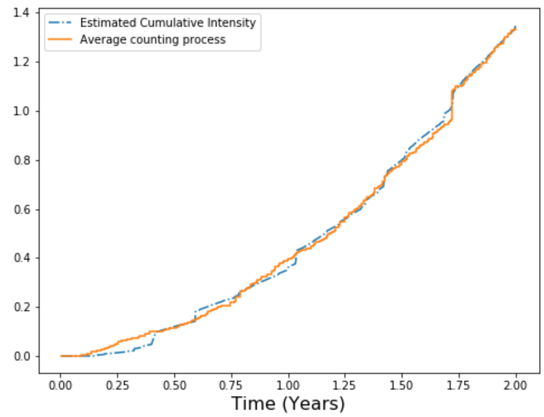
Table 3.5:: Duration Analysis for Long-term Recovery

	long term,k=3
	holiday
age	0.1062849***
	(0.0192083)
age2	-0.1849780***
	(0.0335172)
male	0.5421261***
	(0.0917358)
full time	0.1217907
	(0.1142700)
married	-0.2743786***
	(0.0952768)
log-likelihood	3361.009
Number of Mass Points	2

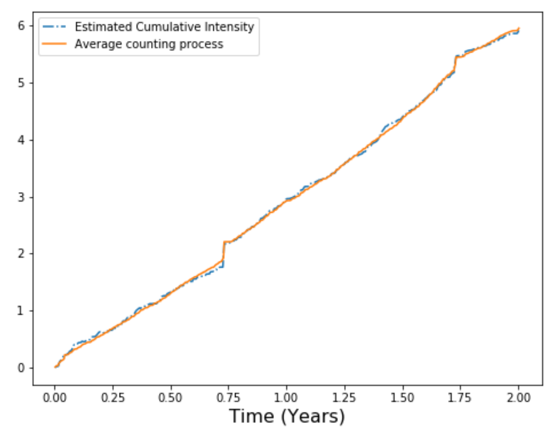
Note: Given the fact that individuals do not respond to the absence scores in the long-term recovery, we instead use the conventional duration model. The subject under study is the duration of long-term absences. Heckman and Singer's NPMLE is used. We found two mass points is good enough. $age2 = age^2/100$. b is the coefficient of duration dependence. *p<0.1; **p<0.05; ***p<0.01



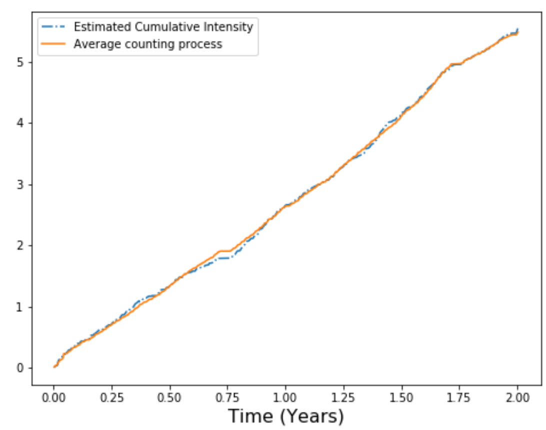
(a) Short term incidence



(b) Long term incidence



(c) Short term recovery (Original)



(d) Short term recovery (Bias Reduction)

Figure 3.1: goodness of fit

short-term incidence intensity as

$$\lambda^*_{2,2}(t) = \exp(\theta(age)H(t))$$

where $\theta(age) = \beta_0 + \beta_1 age + \beta_2 age^2/100$. Other components and the structural of incidence intensity remain unchanged. Table 3.6 reports the results.

To assess the overall significance of $\theta(age)$, we employ two Wald tests: 1) do individuals respond to the cumulative absence time:

$$H_0 : \beta_1 = \beta_1 = \beta_2 = 0$$
$$H_1 : \beta_1 \neq 0, \beta_1 \neq 0, \beta_2 \neq 0$$

and 2) do individuals' attitudes about the cumulative absence time varies as working age changes:

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_1 : \beta_1 \neq 0, \beta_2 \neq 0$$

The Wald statistics for the first and second test are 22334.637 and 355.045 respectively. The results suggest to reject both null hypothesises. We conclude that in the short-term absence, individuals are sensitive to the cumulative absence time, and individuals' attitudes vary along with age when they ask for short-term leaves. To provide a more transparent demonstration, we plot $\theta(age)$ below. What Figure 3.2 shows is that young workers tend to ignore the cumulative absence time when they are making short-term absence decisions as they grow older, but this trend stops around working age of 19.7 (or

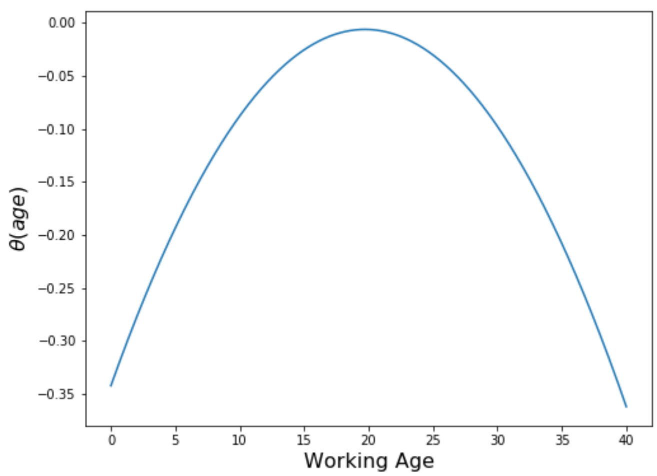
Table 3.6:: Short-term Incidence Intensity

	Incidence Intensity *short term* *k=2*
β_0	-0.34215516* (0.1904758)
β_1	0.03407675 (0.0280976)
β_2	-0.08642652 (0.0960550)
α_{1k}	-39.37455248** (18.741119)
b	2.61810899*** (0.3692254)
age	0.2953078*** (0.0991777)
age2	-1.1536771** (0.5079094)
male	-2.0585887 (1.7245835)
full time	1.31545004*** (0.5182786)
married	-1.02333993** (0.4207051)
Mon/Fri	2.35010473** (1.1141944)
$I(t)$	0.28212665*** (0.0913577)
Distance	0.107879

Note: This table reports the results that instead of estimating the coefficient of absence score, we replace it with a function working ages. Absence duration less or equal to 3 days are categorized as short term, others are long term. $age2 = age^2/100$. *p<0.1; **p<0.05; ***p<0.01

actual age of 35.7). After this age, the cumulative absence time plays more and more significant role in the decision-making process. We suspect the reason for such a pattern is because the age of 36 is the time when a typical worker gets married. So a stable family generate a more matured working attitude.

Figure 3.2: $\theta(age)$ in Short term incidence



3.4.4 The Cut-off Between Short and Long Terms

The criteria we used to distinguished a short and long-term absence is the length of this absence. So far this cut-off is three consecutive days of leave. The reason is due to the eligible condition of UK sick-pay regulation. Under this cut-off, we have seen that individual responds to the cumulative absence time in short-term absences but not in long-term ones. These different responses inspire us to re-define the cut-off between short and long terms.

Define the short-term absences are the ones that, when making the incidence and returning decisions, individuals will consider the cumulative absence time, while in long-term absences individuals do not take into account the cumulative absence time.

Specifically, under the proper cut-off c, the coefficients of cumulative absence time $H(t)$ in $\lambda_{2,k}$ and $\lambda_{6,k}, k = 2, 3$ satisfy 1) the short-term coefficients are significant away from zero, and 2) the long-term coefficients are insignificant.

We may use the newly introduced definition of the short and long term to find the proper cut-off c. To do so, we gradually 're-define' the short-term absence as any absence that is less or equal to c days and report the estimating results for these short and long-term intensities. The aim is to test the significance of coefficients of $H(t)$.

Table 3.7 presents the results. If the cut-off is at 2 days, both $\beta_{1k}, k = 2, 3$ are significant. This means some absences in the long-term should be categorised as short-term. When cut-off is at 4 or 5 days, both $\beta_{k1}, k = 2, 3$ are insignificant. This means some absences in the short-term should be categorised as long-term. It turns out 3 days of absence duration is the proper cut-off. The coefficient of Monday/Friday is another evidence that favours $c = 3$: only short-term absences are sensitive to the Monday and Fridays. This is exactly the case when $c = 2$ and $c = 3$, yet when $c > 3$, this coefficient is no longer significant. It suggests that for absence durations that are longer than three days, they should be classified as long-term absences.

Recall that one eligibility to claim the statutory sick pay in the UK is that individuals need to have been off work sick for beyond three days. We do not think the fact that the proper cut-off is the same as this qualification is merely coincident. Instead, it highlights the importance of this social security regulation.

Table 3.7:: Cut-off Check

	Incidence Intensities			*Recovery Intensities*	
	short term		*long term*	*short term*	*long term*
Cut-off	*Mon/Fri*	β_{12}	β_{13}	β_{21}	β_{31}
$c = 2$	1.59447952**	-0.1094399***	-0.05018605**	0.0006166***	0.00030766
	(0.5741051)	(0.0289985)	(0.0214045)	(0.0001899)	(0.0009172)
$c = 3$	2.01429447*	-0.05734195***	0.002551	0.0008113***	9.046*10^{-6}
	(1.142056)	(0.007313)	(0.0283703)	(0.0002250)	(0.0002337)
$c = 4$	1.08155452	-0.0466356***	0.00378901	0.0003203	6.802*10^{-6}
	(0.7767747)	(0.0119311)	(0.0110785)	(0.0002460)	(0.0002879)
$c = 5$	0.54790203	-0.05774703***	-0.00752412	0.0000130	9.353*10^{-6}
	(0.3641992)	(0.0121355)	(0.0484371)	(0.0004401)	(0.0004270)

Note: To be a proper cut-off, the coefficients must be the case: Short-term absence scores coefficients are significantly different than zero, while long-term absence scores coefficients are insignificant. *p<0.1; **p<0.05; ***p<0.01

3.5 An Economic Model for Work Absenteeism

In this section, inspired from the empirical results, we present a simple economic model. We first provide a narrative approach to describe the incentives to the strategic behavior in work absences. Next, we modify a standard labour-leisure model to characterise the decision-making process of asking for leave and returning to work. We also construct a structural model to describe how individuals optimise the long-term absence durations.

3.5.1 The Incentive to Absence behavior

It is known that a work search is costly. A worker may accept a job offer even though the contracted wage is not equal to the marginal rate of substitution between leisure and income. If a worker accepts such a job offer, she remains an incentive to consume more leisure, one common way to do so is, of course, to be absent from work.

Even if the marginal rate of substitution between income and leisure is equal to the contracted wage, a worker may occasionally prefer to be absent due to external accidents. A worker will choose to be absent when the (expected) size of a shock is large, and the alternative activities are more attractive.

The last element is the worker's personal absence history. Working discipline regulations in most firms specified particular reward/punishment schedules for work absences. These rules usually reward 'good reputation' workers (those who have less cumulative absence time) and punish 'bad reputation' workers (those who have more cumulative absence time). The shadow costs for workers in different positions of the cumulative absence time spectral are different. This creates the incentive to consume more (or less) absences depending on one's absence score.

3.5.2 Decision to Ask for Absence

Suppose the worker's utility is a linear function of $\omega, C, R(A)$ and some other unobserved factors. ω is the general well-being, and C is consumption. $R(A)$ is the reputation. It is a function of cumulative absence time A with $R^{'}(\cdot) < 0$ and $R^{''}(\cdot) < 0$.

Note in the incidence intensity model, we express the reputation as $\exp(-\beta A)$, whose first and second derivative are $-\beta \exp(-\beta A) < 0$ and $\beta^2 \exp(-\beta A) > 0$ respectively. We do not think this setting contradicts our economic assumptions on the reputation function. Since individual may not necessarily map utility to absence actions linearly. If current utility is quite low, one more absence may make little difference to individuals.

At the first stage, workers accept the job offers and have the same reputation. Random shocks $e \in [0, \infty)$ hit all individuals. Notice that $e = 0$ means no accident at shock, and a higher value of e indicates a more severe shock. Workers can observe the existence of the shocks but cannot observe the sizes of them without further information. We assume at this stage, after observing the shocks, workers will always ask for absence. After that, further information is given, the size of the shock is known, and workers choose the duration of the absence (The decision process for how to choose the length of an absence spell will be described later). Cumulative absence time is updated from $0 \to A$ (different values of A for different workers). The well-being $\omega^{'}$ evolves as follow:

$$\omega^{'} = \omega - e + g(A)$$

where $g(\cdot)$ is the well-being generating function with $g(0) = 0, g^{'}(\cdot) > 0$ and $g^{''}(\cdot) < 0$.

In the second stage, individuals again observe the existence of shocks. But in this stage, a worker has to decide whether to ask for the absence ($D = 1$ for absence, $D = 0$ otherwise) based on her history and the expectation on the size of the accident by:

$$\begin{aligned} D &= \mathbb{I}\{U(R(A + a(\mathbb{E}(e))), \omega^{'} - \mathbb{E}(e) + g(a(\mathbb{E}(e))), C_1) + \epsilon_1 > U(R(A), \omega^{'} - \mathbb{E}(e), C_2) + \epsilon_2\} \\ &= \mathbb{I}\{U^1 + \epsilon_1 > U^0 + \epsilon_2\} \end{aligned}$$

where $U(\cdot)$ is the utility function. Let $U_R, U_\omega, U_C > 0$ be the partial derivatives of the utility function. $a(\cdot)$ is the duration of the absence and is determined by the size of an accident with $a^{'}(\cdot) > 0$, ϵ_1, ϵ_2 represent unobserved factors that might effect the utility function.

In the case of long-term absences, individuals do not respond to the reputation, the absence decision is then governed by

$$D = \mathbb{I}\{U(\omega^{'} - \mathbb{E}(e) + g(a(\mathbb{E}(e))), C_1) + \epsilon_1 > U(\omega^{'} - \mathbb{E}(e), C_2) + \epsilon_2\}$$

This decision rule specifies that an individual will ask for a leave if and only if the expected utility for being absent is higher than the utility of attendance. The cumulative absence time is evolved as

$$A = A_{-} + a$$

Individuals' absence decisions are then depended on (a) their cumulative absence time and (b) their beliefs about the size of the accidents.

Taking the expectation, we have

$$\begin{aligned} Pr(D=1) &= Pr(\epsilon_2 - \epsilon_1 < U^1 - U^0) \\ &= F_\epsilon(U^1 - U^0) \end{aligned}$$

where $\epsilon = \epsilon_2 - \epsilon_1$

Since $F^{'}(\cdot) > 0, R^{'}(\cdot) < 0$ and $R^{''}(\cdot) < 0$. For short-term absences, we have:

$$\frac{\partial Pr(D=1)}{\partial A} < 0$$

3.5.3 Decision to Recovery

Conditional on the fact that individuals have decided to take absences, they will receive information about the size of shocks. This further information is given by, for example, doctors if workers went to hospitals. The workers then have to decide the duration of their absences. Since the empirical results suggest that only in the short-term recovery processes, workers tend to have strategic behavior, we assume that reputations will only be a part of the equation if the size of an accident is within some level. That is if $e \leq e^*$, $a(e,R) \in [0, a(e^*)]$. If $e > e^*$, $a(e)$ is then a deterministic function of accident e that can not be altered by the reputation R.

Within such size range, a worker's problem is:

$$\max_a U(R(A+a), \omega - e + g(a), C)$$

$$s.t$$

$$I + w(t^c - a) + R(A + a) - C = 0 \tag{3.10}$$

where I is non-labour income, w is wage, t^c is the contracted working time.

First order condition with respect to a yields:

$$\begin{aligned} U_R R^{'} + U_\omega g^{'} - U_C(w - R^{'}) = 0 \\ U_C(w - R^{'}) = U_R R^{'} + U_\omega g^{'} > 0 \end{aligned} \tag{3.11}$$

By differentiating the first order condition (3.10) through (3.11), one can show that

$$\frac{\partial a}{\partial A} < 0$$

That is, as long as the accident is small ($e < e^*$), the shorter the cumulative absence time, the longer absence duration one may choose.

Notice that in the case of scheduled absence, there is no stochastic in a 'shock'. The size of this 'shock' is observed all the time. And the decisions to ask for leave and to return to work should be made simultaneously: workers do not need the decision process for asking for absence, she only need to decide the duration of such absence (0,1,2 or 3 days,0 days absence means no absence).

3.5.4 A Structural Model for Long-Term Recovery

If the sizes of accidents are greater than the threshold e^*, a worker may recognise this event as a 'major' and will leave the reputation out of the equation. Statistically, this means that the duration of a long-term absence is memoryless, and there is no harm to treat each of them as independent and identical distributed.

The task of a worker under this circumstance consists of choosing an optimal duration to maximise her utility without the consideration of reputation. This task is mostly a discrete choice problem under continuous time. And the independence assumption inspires us to build a simple structural model for the long-term absence duration decision making process. This structural model is a simplified version of Honore & De Paula (2010) and de Paula & Honore (2017), in which the authors study the couple's interdependent retirement durations.

For individual i who is now in j^{th} long-term recovery period, she has a positive utility flow $K_{ij}Z_1(t)\phi_1(X_i)$, where K_{ij} is a positive random variable that could represent initial health. At any point, she may choose to 'switch' to the alternative state: returning to work, with a utility flow $Z_2(t)\phi_2(X_i)$. Assuming individuals are myopic and an exponential discount rate ρ, individual i's utility for taking part in the j^{th} long-term recovery period until time t_{ij} is:

$$\int_0^{t_{ij}} K_{ij}Z_1(s)\phi_1(X_i)e^{-\rho s}ds + \int_{t_{ij}}^{\mathbb{E}(T)} Z_2(s)\phi_2(X_i)e^{-\rho s}ds \tag{3.12}$$

where $\mathbb{E}(T)$ is the expecting beginning time of a next long-term absence.

The first order condition for maximizing this with respect to t_{ij} is:

$$\Big[K_{ij}Z_1(t_{ij})\phi_1(X_i) - Z_2(t_{ij})\phi_2(X_i)\Big]e^{-\rho t_{ij}}$$

Thus the optimal T_{ij} is given by:

$$\begin{aligned} T_{ij} &= \inf\{t_{ij} : [K_{ij} Z_1(t_{ij})\phi_1(X_i) - Z_2(t_{ij})\phi_2(X_i)]e^{-\rho t_{ij}} < 0\} \\ &= \inf\{t_{ij} : K_{ij} - Z(t_{ij})\phi(X_i) < 0\} \end{aligned} \tag{3.13}$$

where $Z(\cdot)\phi(X_i) = Z_2(\cdot)\phi_2(X_i)/\big(Z_1(\cdot)\phi_1(X_i)\big)$.

Notice the above equation is in the spirit of discrete choice structure model under a latent variable framework in the sense that individual compares the instant utility between two states: $\nu^\star = K_{ij} - Z(t_{ij})\phi(X_i)$. If $\nu^\star \leq 0$, individuals will return to work, $\nu^\star > 0$ otherwise. The multiplicative structure of $Z(t)$ and $\phi(X_i)$ is explicitly designed to have the accelerated failure time model as a special case. There is no difficulty in estimation to lose this structure. To sum up, the individual will switch at

$$T_{ij} = Z^{-1}(K_{ij}/\phi(X_i)) \tag{3.14}$$

Notice that in this structure model, the source of randomness is K_{ij}. We can re-write equation 3.14 as the follows:

$$\ln Z(T_{ij}) = -\ln\phi(X_i) + \epsilon \tag{3.15}$$

where $\epsilon = \ln K_{ij}$. Equation 3.15 is a typical accelerated failure time (AFT) model. Assume $Z(t) = t$, $\phi(X_i) = e^{-X_i^T\beta}$ and $K_{ij} \sim \exp(1)$, we may end up with the exponential

AFT model. The cumulative distribution function of T_{ij} is given by

$$\begin{aligned} F_{T_{ij}}(t) &= Pr[K_{ij} e^{X_i^T \beta} \leq t] \\ &= Pr[K_{ij} \leq t e^{-X_i^T \beta}] \\ &= 1 - \exp(-t \exp(-X_i^T \beta)) \end{aligned} \tag{3.16}$$

The corresponding hazard rate is

$$\begin{aligned} h_{T_{ij}}(t) &= \frac{f_{T_{ij}}(t)}{1 - F_{T_{ij}}(t)} \\ &= \exp(-X_i^T \beta) \end{aligned} \tag{3.17}$$

These assumptions are mainly made to compare with our reduced form model 3.7 from the previous section: their hazard rates are identical except that 1) in the reduced form model, the random effect variable is included and 2) the signs of coefficients are opposite. Intuitively, a higher hazard leads to a shorter duration.

Table 4.16 presents the estimates for this exponential AFT model. Not surprisingly, the results are consistent with the reduced form hazard model. However, we have to mention that the structural model might not fit the real data well. A hazard function like 3.17 can be interpreted as no individual heterogeneity. But we have seen from the reduced form model that the long-term duration data does have heterogeneity (with mass points of two).

Table 3.1:: AFT Model Results

	Dependent variable:
	long-term duration
age	-0.26440^{***}
	(0.00685)
age2	0.44302^{***}
	(0.01569)
male	-0.58349^{***}
	(0.07593)
full time	-0.09395
	(0.08192)
married	0.14154^{*}
	(0.08066)
Observations	1,204
Log Likelihood	2,971.49500
χ^2	-362.47690 (df = 4)

Note: The structure econometric model in our setting is in fact an accelerated failure time model. This table reports the results. Comparing to the long-term duration analysis (table 4.1), 1) the coefficient sign are opposite, but the economic meaning are the same; and 2) lack of the heterogeneity term. $age2 = age^2/100$. *p<0.1; **p<0.05; ***p<0.01

3.6 Conclusion

In this paper, a series of self-exciting process models are constructed to study the work absenteeism. A minimum distance estimation method is employed. This estimation method, unlike the conventional likelihood-based method, allows including external shocks into the intensity.

In the empirical study, firm-level data is used. The firm introduced an experience rate sick pay scheme that links sick pay benefit with worker's absence history. We find the worker's decision makings are entirely different in short-term, long-term incidence and recovery processes. Specifically, we found substantial evidence supporting the existence of strategic behavior in both short-term incidence and recovery process. The strategic behavior is generated by the cumulative absence time. However, in the long-term recovery process, we have to reject the existence of strategic behavior and state-dependent structure. Instead, we adopt a conventional duration analysis and employ Heckman and Singer's NPMLE to complete the analysis.

A theoretical framework of work absence is developed. This model incorporates the strategic decision-making process and fits our empirical findings.

Appendix

3.A Count Data Regressions and Duration Models

3.A.1 Four Count Data Regressions

The dependent variable in these models is the counts of events in an interval of time. The most basic count data regression model is the Poisson, where $Pr(C_i = c \mid X_i) = \exp(-\mu(X_i))\mu(X_i)^c/c!$ $\mathbb{E}(C_i|X_i) = \mu(X_i) = Var(C_i|X_i)$, C_i and X_i are counting numbers and covariates for individual i respectively. Normally, $\mu(X_i) = \exp(X_i'\beta)$.

The equality between the mean and the variance in the Poisson model is restrictive. A popular generilisation of over-dispersion model is the negative binomial, whose density is given by

$$f_{nb}(c_i \mid X_i) = \frac{\Gamma(c_i + \psi_i)}{\Gamma(\psi_i)\Gamma(c_i + 1)} \left(\frac{\psi_i}{\lambda_i + \psi_i}\right)^{\psi_i} \left(\frac{\lambda_i}{\lambda_i + \psi_i}\right)^{c_i}$$

where $\lambda_i = \exp(X_i'\beta)$ and the precision parameter ψ_i^{-1} is specified with $\psi_i = \lambda_i/\alpha$ and a positive over-dispersion parameter α. This specification yields the mean function

$\mathbb{E}[C_i \mid X_i] = \lambda_i$ and the variance function $Var[C_i \mid X_i] = (1+\alpha)\lambda_i$.

Zero-inflation and hurdle models are good at explaining the excess of zeros. The zero-inflation model considers a mixture distribution of a degenerated distribution concentrated on zero and a negative binomial distribution. In particular,

$$Pr(C_i = 0 \mid X_i, Z_i) = \phi(Z_i) + (1 - \phi(Z_i)) f_{nb}(0 \mid X_i),$$
$$Pr(C_i = c_i \mid X_i, Z_i) = (1 - \phi(Z_i)) f_{nb}(c_i),$$

where Z_i is a vector of zero-inflated covariates,$\phi(\cdot))$ is the binomial probability. The zero-inflation model can be treated as a special case of the latent class model.

The Hurdle model, on the other hand, can be interpreted as the first part concerns the decisions to ask for leave as a binary outcome process, while the second part models the positive number of work absences conditional on the individual seeking a leave. In particular,the first part of the two-part hurdle structure is specified as

$$Pr(C_i = 0 \mid X_i) = \left(\frac{\psi_{h,i}}{\lambda_{h,i} + \psi_{h,i}}\right),$$
$$Pr(C_i > 0 \mid X_i) = 1 - \left(\frac{\psi_{h,i}}{\lambda_{h,i} + \psi_{h,i}}\right)$$

where the subscript h denotes parameters associated with the “hurdle distribution”. The likelihood function associated with this stage of the hurdle process can be maximized independently of the specification of the second stage. The second part of the model is given by the truncated negative binomial distribution:

$$f(c_i \mid X_i, C_i > 0) = \frac{\Gamma(c_i + \psi_i)}{\Gamma(\psi_i)\Gamma(c_i + 1)} \left[\left(\frac{\lambda_i + \psi_i}{\psi_i} \right)^{\psi_i} - 1 \right]^{-1} \left(\frac{\lambda_i}{\lambda_i + \psi_i} \right)^{c_i}.$$

3.A.2 Duration Models

As mentioned in the paper, the workhorse in the duration analysis is the hazard rate $h(t) = f(t)/S(t)$, where $f(t), S(t)$ are probability density function and its survival function respectively. In a basic duration model, for every individual, define the constant hazard rate and its cumulative hazard rate as:

$$\begin{aligned} h_i(X_i) &= \exp(X_i'\beta) \\ H_i(T) &= h_i(X_i)T \end{aligned}$$

Notice from the definition of the hazard rate, we have:

$$\begin{aligned} -h(t) &= \frac{dlog(S(t))}{dt} \\ -\int_0^T h(t)dt &= log(S(T)) \\ S(T) &= exp(-\int_0^T h(t)dt) \end{aligned}$$

Hence the likelihood function is

$$L = \prod_{i=1}^{N} L_i = \prod_{i=1}^{N} \exp(-H_i(t))[h_i]^{y_i}$$

where y_i is the censoring indicator: if censored, $y_i = 0$, otherwise $y_i = 1$.

One concern regarding this model is the unobserved heterogeneity among individuals. The usual way to account for this is to include a random variable $\nu \sim G$ in the hazard rate.

$$h_i(\nu_i, X_i) = \exp(X_i'\beta + \nu_i)$$

Integrate out the random variable ν, we end up with the marginal hazard rate,

$$\begin{aligned} h(t|X) &= \frac{\int_0^\infty h(X,\nu)S(t|X,\nu)dG(\nu)}{S(t|X)} \\ &= \exp(X\beta)\mathbb{E}(\exp(\nu)|T > t, X) \end{aligned} \tag{3.18}$$

where $S(t|X)$ is the associated survival function.

The second equation comes from the fact that

$$\begin{aligned} g(\nu|T > t, Z) &= \frac{Pr\{T \geq t|Z,\nu\}g(\nu)}{Pr\{T > t|Z\}} \\ &= \frac{S(t|Z,\nu)g(\nu)}{S(t|Z)} \end{aligned}$$

and

$$\mathbb{E}(\exp(\nu)|T > t, Z) = \frac{\int_0^\infty \exp(\nu)S(t|Z,\nu)g(\nu)d\nu}{S(t|Z)}$$

Assume ν is independent from X_i, one may use Heckman & Singer (1984)'s non-parametric maximum likelihood estimator (NPMLE) to avoid unjustified assumptions about the distribution G. Instead, one may approximate G in terms of a discrete distribution.

Let Q be the (prior unknown) number of support points in this discrete distribution and let $\nu_l, p_l, l = 1, 2, \cdots, Q$ be the associated location scalars and probabilities. The likelihood contribution is:

$$\mathbb{E}[L_i(\nu_i)] = \sum_{l=1}^{Q} p_l L_i(\nu_l), \sum_{l=1}^{Q} p_l = 1$$

where $L_i(\nu_l) = \exp(-H_i(t|\nu_l, X_i))[h_i(t|\nu_l, X_i)]^{y_i}$.

The likelihood function is

$$L = \prod_{i=1}^{N} \mathbb{E}[L_i(\nu)] = \prod_{i=1}^{N} \sum_{l=1}^{Q} p_l L_i(\nu_l), \sum_{l=1}^{Q} p_l = 1$$

The estimation procedure consists of maximising the likelihood function with respect to β as well as the heterogeneity parameters ν_l and their probabilities p_l for different values of Q. Starting with $Q = 1$, and then expanding the model with new support points until there is no gain in likelihood function value.

Heckman & Singer (1984) has proven that such an estimator is consistent, but its asymptotic distribution has not been discussed yet. Gaure et al. (2007) provide Monte Carlo evidence indicating the parameter estimates obtained by NPMLE are consistent and approximately normally distributed and hence can be used for standard inference purpose.

3.B Finite Mixture Poisson-2 Model

We set the number of component $k = 2$. The dependent variable is the counts of absences in the year 1988, explanatory variables include age, sex, full/part time status, marriage status and the counts of absences in previous year. The estimation results for both short-term and long-term absences are presented below.

Table 3.B.1:: Finite Mixture Poisson Model

	Dependent Variable: Counts 88		
	short term		*long term*
	Component 1	*Component 2*	*Component 1*
Intercept	2.4881479***	1.15043828	0.2400829
	(0.3704769)	(0.53720132)	(0.4198270)
age	-0.0311709	-0.01003723	-0.0015710
	(0.0219730)	(0.02758507)	(0.0225010)
age2	0.0242097	-0.00103913***	0.0094344
	(0.0285556)	(0.03301050)	(0.0268680)
sex	-1.2429011***	0.65663285***	-0.2126884*
	(0.1455399)	(0.18437828)	(0.0882432)
full	-0.8829060***	1.06507140 ***	0.0212277
	(0.1690357)	(0.18205057)	(0.0965388)
marriage	0.0240618	-0.22561099*	0.0971282
	(0.1070147)	(0.12586190)	(0.1068318)
count 87	-0.00388629*	0.00095098	-0.0022418
	(0.0020161)	(0.00157852)	(0.0015178)
Number of Individuals	450	303	562

Note: This table presents the finite mixture Poisson model with two component. For the long-term absences, the posteriors suggest homogeneity, that is, all the long-term absences come from the same data generating process. This, however, is not the case in the short-term absences. *p<0.1; **p<0.05; ***p<0.01

Chapter 4

A Dynamic Analysis of Spanish Youth Unemployment

Abstract

This chapter investigates how past (un)employment records could affect an individual's job turnover. One challenge we face is how to separate the state dependent (or the past experiences) effects from the unobserved heterogeneity. We overcome this issue by first constructing counting processes that consist of interested (un)employment duration. We specify a multiplicative accelerated failure time structure for these duration. Second, for each individual's counting process we do a first ratio transformation on duration to swipe out the unobserved heterogeneity. Comparing to existing methods, our approach includes the unobserved heterogeneity in a fixed effect way and allows researchers to build models flexibly. This new method is also an extension to the classical dynamic panel data models where weak IV problem can be solved and unit root and the non-stationary

process could also be allowed. Using Spanish social security data, we find workers mostly care about their human capital stock, duration dependence effect is different for low education and high education workers.

4.1 Introduction

The Spanish labor market is well-known for her high unemployment rate. In 2018, Eurostat reports that the youth unemployment rate in Spain is 32.7%, second only to Greece in EU. In this paper, we are primarily interested in studying how previous (un)employment records would affect the youth's job turnover. That is, we try to investigate the state dependent effects on the inflows and outflows to unemployment. Answering this question is challenging as 'the unemployment rate is very serially correlated and possibly nonstationary'(Ahn & Hamilton 2019) and the difficulties to separate the state dependent effect and the unobserved heterogeneity (Heckman 1981; Honoré 1993). In this paper, we aim to build a statistical model that can handle the unobserved heterogeneity and at the same time, could include different state dependent structures flexibly.

Most previous papers in this literature that tried to allow heterogeneity did so by including various observed covariates (e.g., demographics, education, occupation, geographical region and more). However, the important role of unobserved heterogeneity is increasingly recognized by the academic community: 'Any pool of unemployed individuals who share any given observed characteristics is going to become increasingly represented by those within that group who have higher ex ante continuation probabilities the longer the period of time for which the individuals have been unemployed' (Ahn &

Hamilton 2019).

Another contribution factor to the difference unemployment probabilities is what Van den Berg & Van Ours (1996) called 'genuine duration dependence': the duration of being unemployment may alter to the outflow probability. this dependence structure is essentially a particular example of the state dependence, that is the past experience could change the likelihood of occurrences of future events. Honoré (1993) developed a common practice to separate the unobserved heterogeneity and the genuine duration dependence using multiple spells of unemployment for a given individual. Assuming that the unobserved heterogeneity is persistent over time and no correlation among unemployment spells, identification could be achieved by the ratio of two spells' survival functions. The later independent assumption is used extensively in the literature as noted by Schmillen & Umkehrer (2017).

Honoré (1993) also provided the identification results on the case where multiple spells are correlated through a 'lagged duration dependent structure'. However, such structure is restrictive for several reasons: 1) It requires a balanced panel data and can be quite complicated when the number of spells is large; 2) The identification of the initial duration require the unobserved heterogeneity to be independent with other covariates (i,e., the random effect assumption). As will be shown later, our model is capable of overcoming these hurdles. Besides the duration dependence, we are also introducing the human capital stock dependence into the model. In our paper, the human capital stock is a summary of one individual's past employment and unemployment records.

In this paper, we propose to use the self-exciting process to build a statistical model that includes different state dependent structures (e.g., human capital stock, duration dependence, unemployment insurance,etc.) flexibly and at the same time allows the unobserved heterogeneity in a fixed effect way. We include both individual's employment and unemployment records in the filtration of a self-exciting process and study this process by its compensator.

We restrict our model to a multiplicative structure and then do a first ratio transformation (or equivalently, a first difference transformation after we take logarithm on both sides) to swipe out the unobserved heterogeneity. Thus, from the econometric perspective, we also generalize the classical dynamic panel data model to a non auto-regressive structure. Even in a classical dynamic panel AR model, the self-exciting process approach can easily overcome the weak-IV problem. This is because we are not focusing on the random variable per se, but instead on the whole process, as a result, we do not need IV at all. The classical first-difference transformation in a dynamic panel AR model requires the data must be stationary and can not identify the auto-regressive parameter if it is an unit root. In the self-exciting process, no stationary condition is required and one may perform a simple standard t-test to detect the unit root.

This chapter is structured as follows. In section 2 we briefly present the institution backgrounds and our database. In section 3 we discuss our econometric methods. We will demonstrate the advantages of the self-exciting process by Monte Carlo exercises. In section 4 we describe our empirical models, and in section 5 we present the empirical results. Section 6 contains our conclusions.

4.2 Institution Features and Dataset Description

4.2.1 Institution Features

Spanish labor market is known for its dual structure. The market offers one of the most stringent employment protection for permanent contracts among OECD countries, while temporary fixed-term contracts can be terminated at almost no cost.

The fixed-term contracts were first introduced in the late 1984 with intention to ease the strong employment protection. Since then, temporary contracts have increased to nearly 30% in early 2000s (Bover & Gómez 2004). In recent years, almost 60% of all employed Spanish workers in the age group 16-24 hold a temporary contract, around 40% of the less-educated and 20% of the high-educated workers still hold temporary contracts at the age of 39. The time to find a first regular job for a typical school leaver in Spain is also significantly longer compared to reference countries (Dolado et al. 2013). It is fair to say that temporary contracts fail to act as stepping-stones towards permanent jobs in Spain.

Regarding the unemployment benefit system, like most European countries, the Spanish one is consisted of two pillars. The contributory unemployment insurance provides unemployment protection to workers who contributed when employed. The unemployment allowance is an assistant benefit intended to supplement the contributory benefit. In this paper, we are primarily focus on the contributory unemployment insurance.

Workers who wish to claim the contributory unemployment entitlement must be legally unemployed and register as a jobseeker.In addition, the recipients must have made contributions for a minimum of 360 days within six years prior to becoming legally unemployed.

The benefit duration is closed related to the contribution period. The minimum period is 120 days and the maximum benefit duration is 720 days.

4.2.2 Data Description

We use the anonymized administrative data from the Muestra Continua de Vidas Laborales con Datos Fiscales or Continuous Working Life Sample(MCVL). The MCVL data is rich, containing matched social security, income tax and census records for a 4% random sample of Spanish workers, pensioners and unemployment benefit recipients. Note that this means the MCVL is only representative of the population related to the social security system.

The MCVL database is available to researchers in 2005 and is annually updated. It contains several files: personal details, details of cohabitants, data on social security contribution, data on benefits and tax data.

The data on social security contribution is the main focus in this study. Since both workers and their employers are required to contribute to the social security, we define an individual is being unemployment at time t if there is no contribution record at this time.

We define youth as someone who is below age 35. In this paper, we use the 2014 version of MCVL data and randomly select 25% of the sample size. In the end, there are $10,716$ unique youth individuals in our sample. The total amount of unemployment records is $64,549$.

4.3 Econometric Method

4.3.1 Unobserved Heterogeneity in Self-Exciting Models

Separating the state dependent effect and the unobserved heterogeneity is important in our application, both from economic and econometric point of view. An individual may have shorter unemployment duration due to better work ability rather than a shorter unemployment benefit duration, or both. The econometric challenge come from the fact that the unobserved heterogeneity is correlated with the state dependent factors, thus the conventional random effect independent assumption is not valid. In this paper, we propose a first ratio fixed effect approach by restricting the duration be a multiplicative structure that can separate the unobserved heterogeneity and other factors.

The general structure can be written as follows.

$$L(\tau_{i,k}) = G(H_{i,k};\beta)\nu_i u_{i,k} > 0 \tag{4.1}$$

where $\tau_{i,k}$ is individual i's k^{th} duration, $H_{i,k}$ is a vector of her state dependent variables prior to k^{th} duration, β is a vector of parameters that need to be estimated, $L(\cdot), G(\cdot)$ are known functions, ν_i are unobserved individual heterogeneity, and $u_{i,k}$ are i.i.d error terms across both duration and individuals.

This general form is essentially an accelerated failure time (AFT) model.

Example 1. Let $L(\cdot) = G(\cdot) = \exp(\cdot)$, $H_{i,k} = \tau_{i,k-1}$, then we have

$$\begin{aligned} \exp(\tau_{i,k}) &= \exp(\beta \tau_{i,k-1}) \nu_i u_{i,k} \\ \tau_{i,k} &= \beta \tau_{i,k-1} + \log(\nu_i) + \log(u_{i,k}) \end{aligned} \tag{4.2}$$

Equation 4.2 is the classical AR(1) dynamic panel data model with i.i.d error term assumption.

Example 2.

$$\tau_{i,k} = \exp(\beta \sum_{j=1}^{k-1} \tau_{i,j}) \nu_i u_{i,k} \tag{4.3}$$

where $\tau_{i,j}$ could be individual's employment duration and the summation $\sum_{j=1}^{k-1} \tau_{i,j}$ measures one's working experience. Notice model 4.3 can not use the widely employed IV, GMM estimation methods in the dynamic panel data models. This is because the data generating process is not necessarily stationary.

Example 3. Let $L(x) = x^2/2$, $G(\cdot) = \exp(\cdot)$, $H_{i,k} = \tau_{i,k-1}$ and $u_{i,k} \rightsquigarrow \exp(1)$ we have

$$\frac{\tau_{i,k}^2}{2} = \exp(-\beta \tau_{i,k-1}) \nu_i u_{i,k} \tag{4.4}$$

Equation 4.4 has a duration-dependent hazard rate $\lambda(\tau_{i,k})$:

$$\lambda(\tau_{i,k}) = \tau_{i,k} \exp(\beta \tau_{i,k-1}) \nu_i \tag{4.5}$$

To swipe out the the unobserved heterogeneity ν_i, we take the following first ratio transformation.

$$\tilde{\tau}_{i,k} = \frac{L(\tau_{i,k})}{L(\tau_{i,k-1})} = \frac{G(H_{i,k})}{G(H_{i,k-1})}\tilde{u}_{i,k} \tag{4.6}$$

where $\tilde{u}_{i,k} = u_{i,k}/u_{i,k-1}$. Provided that this ratio distribution $F_{\tilde{u}}(\cdot)$ is well defined, the distribution of $\tilde{\tau}_{i,k}$ is:

$$F_{\tilde{\tau}_k}(x) = F_{\tilde{u}}\Big(\frac{G(H_{i,k-1})}{G(H_{i,k})}x\Big) \tag{4.7}$$

Constructing a new counting process $\tilde{N}(t)$ whose inter-event duration are $\tilde{\tau}_{i,k}$, then by Equation 1.23, the corresponding cumulative intensity is

$$\mathbb{E}\tilde{N}(t) = \tilde{\Lambda}(t) = \tilde{\Lambda}(T_k) + \int_0^{t-\sum_{j=1}^{k+1}\tilde{\tau}_{i,j}} \frac{F_{\tilde{\tau}_k}(dx)}{1-F_{\tilde{\tau}_k}(x)}, t \in (\sum_{j=1}^{k}\tilde{\tau}_{i,j}, \sum_{j=1}^{k+1}\tilde{\tau}_{i,j}] \tag{4.8}$$

One may estimate the parameters by minimizing the distance between $\tilde{N}$ and its cumulative intensity $\tilde{\Lambda}$ as described in the previous subsection.

4.3.2 A Generalization of Dynamic Panel Models

In Example 1, we have shown that the canonical AR(1) dynamic panel model is a special case of our general structure 4.1. In this subsection, we will argue that by focusing on the whole process rather than the random variable per se, we can overcome some issues in classical IV-GMM based dynamic panel estimator.

Two advantages of the new approach are 1) avoiding weak IV problem, and 2) easy unit-root testing and compatible with non-stationary process.

For the purpose of illustration, we use the following model:

$$y_{i,k} = \nu_i + \rho y_{i,k-1} + u_{i,k} \tag{4.9}$$

where ν_i are the unobserved heterogeneity, $u_{i,k}$ are i.i.d error terms across both individuals and periods, and $k \geq 1$, $y_0 = 0$.

Weak IV Problem. When the auto-regressive parameter ρ is close to unit or the variance of the unobserved heterogeneity σ_ν is largely greater than the variance of the error term σ_u, the widely used linear GMM estimator obtained after the first difference performs poorly.

The poor performance is caused by the weak instrument problem. Blundell & Bond (1998) used the following set-up to demonstrate. Consider the case with $T = 3$. For $T = 3$, ρ is just identified and the first difference operation yields:

$$\begin{aligned} \Delta y_2 = y_2 - y_1 &= \nu_i + (\rho - 1)y_1 + u_2 \\ &= \pi y_1 + r_i \end{aligned} \tag{4.10}$$

Assuming stationarity, the plim of $\hat{\pi}$ is given by Blundell & Bond (1998)

$$plim \hat{\pi} = (\rho - 1)\frac{k}{(\sigma_\nu^2/\sigma_u^2) + k} \tag{4.11}$$

with $k = (1-\rho)^2/(1-\rho^2)$. One finds that $\hat{\pi} \to 0$ as $\rho \to 1$ or as $(\sigma_\nu^2/\sigma_u^2) \to \infty$.

Arellano & Bover (1995) suggests using additional linear moment conditions that are valid when certain restrictions are imposed on the initial values,Blundell and Bond (1998)'s system GMM suggests using non-linear moment conditions and restrictions on the initial condition process to solve the weak instrument problem.

The new approach, on the other hand, is based on a functional data and focuses on process (intensity) rather than random variable (distribution). As a consequence, we do not use IV at all and such weak IV problem is irrelevant to our methodology. We demonstrate this property with a Monte Carlo study in the next subsection.

Unit-root and Non-stationarity. The IV-based estimators all require stationarity, i.e., $|\rho| < 1$ (e.g., Anderson & Hsiao (1981),Blundell & Bond (1998),Arellano & Bover (1995)). It is well known that the underlying moment conditions do not identify the auto-regressive parameter when its value is unity.

To our knowledge, so far there are only few contributions to the topic of GMM-based unit root inference. Harris & Tzavalis (1999) studied the model presented in Equation 4.9. Their test is based on a correction of the inconsistency of the least square dummy variable estimation. But it requires $\nu_i = 0$ for all i.

Bond et al. (2005) studied a model where when $\rho = 1$, there are no individual unobserved heterogeneity, so the null hypothesis is that the time series are random walks with no individual drifts.

Our approach, however, is capable of dealing with unit root as well as non-stationarity. This property is highlighted in the original Kopperschmidt & Stute (2013) paper. A simple t-test based on such estimation is able to detect the unit root. Two Monte Carlo studies are presented in the next subsection. One is with a unit root $\rho = 1$, the other is when the auto-regressive parameter is greater than the unit $\rho > 1$.

4.3.3 Monte Carlo Evidences

We perform two classes of Monte Carlo exercises. The first class has the standard AR(1) dynamic panel model structure. The second class has a non AR structure and is much similar to Example 2.

AR(1) Dynamic Panel Models. Throughout this class of Monte Carlo studies, we use the following data generating process (DGP):

$$\tau_{i,k} = exp(\nu_i + \beta X_i + \rho log(\tau_{i,k-1}))\epsilon_{i,k} \tag{4.12}$$

where $\epsilon_{i,k} \rightsquigarrow exp(1)$ are i.i.d. Taking logarithm on both side and let $y_{i,k} = log(\tau_{i,k})$, we have:

$$y_{i,k} = \nu_i + \beta X_i + \rho y_{i,k-1} + u_{i,k} \tag{4.13}$$

with $u_{i,k} = log(\epsilon_{i,k})$.

For different experiments, we have different settings on ν_i and ρ, but we always generate X_i from $N(0,5)$ and let $\beta = -1.5$.

We set $N = 100$ and $T = 6$, and we run 300 replications.

Experiment 1. In the first experiment, we set $\rho = 0.95$ and $\nu_i \rightsquigarrow N(0,5)$ such that $\rho - 1$ is close to zero and σ_ν/σ_u is large. In the conventional IV and GMM approach, this setting leads to a weak IV problem.

The average estimate is $\hat{\rho} = 0.934$, the average standard deviation is $\hat{\sigma}_\rho = 0.019$ and the empirical coverage rate is 0.95.

Experiment 2. In the second experiment, we let $\rho = 1$ and $\nu_i \rightsquigarrow N(0, 1)$. This setting generates a unit root problem. In this experiment, we generate both balanced panel data and unbalanced panel data. Existing testing methods (Harris & Tzavalis (1999), Bond et al. (2005)) require a balanced panel data.

For the balanced case, the average estimate is $\hat{\rho} = 0.982$, the average standard deviation is $\hat{\sigma}_\rho = 0.02$ and the empirical coverage rate is 0.94.

We create the the unbalanced data by restricting $\sum_k \tau_{i,k} \leq 5.5$. The average estimate is $\hat{\rho} = 0.984$, the average standard deviation is $\hat{\sigma}_\rho = 0.017$ and the empirical coverage rate is 0.93.

In both cases, a simple t-test would accept the null hypothesis that $\rho = 1$.

Experiment 3. In the third experiment, we let $\rho = 1.2$ and $\nu_i \rightsquigarrow N(0, 1)$. This setting generate a non-stationary series. Although such setting has little implication in Economic studies, we nevertheless investigate it to demonstrate the extensiveness of our approach.

The average estimate is $\hat{\rho} = 1.189$, the average standard deviation is $\hat{\sigma}_\rho = 0.014$ and the empirical coverage rate is 0.963.

Non Auto-Regressive Structure. We generate the following self-exciting process with individual heterogeneity:

$$\tau_{i,k} = \exp(-\nu_i)\exp(-\beta\sum_{j=1}^{k-1}\tau_{i,j})u_{i,k} \tag{4.14}$$

where $u_{i,k} \rightsquigarrow exp(1)$, $\nu_i \rightsquigarrow exp(1)$.

We use Ogata's thinning method to generate data, the following is the pseudo code.

Algorithm 1 Thinning algorithm, generate above self-exciting process on $[0,T]$

[1] Initialize $s = 0, n = 0, \mathcal{T} = \emptyset$ $s < T$ Set $\bar{\lambda} = \lambda(s+) = \exp(\nu_i + \beta s)$ Generate $u \rightsquigarrow U(0,1)$ Let $w = -ln(u)/\bar{\lambda}$ so that $w \rightsquigarrow exp(\bar{\lambda})$
Generate $D \rightsquigarrow U(0,1)$ $D\bar{\lambda} \leq \lambda(s)$ accepting with prob. $\lambda(s)/\bar{\lambda}$ Set $s = s + w$ s is the next point $n = n + 1$ updating the number of accepted points $t_n = s$ $\mathcal{T} = \mathcal{T} \cup \{t_n\}$ adding t_n to the ordered set $\mathcal{T}$
$t_n \leq T$ $\{t_k\}_{k=1,2,\cdots,n}$ $\{t_k\}_{k=1,2,\cdots,n-1}$

We set $\beta = -0.5$ and the terminated time $T = 7.2$. The sample size is $N = 500$ and run 300 replications.

The averaged estimate $\hat{\beta} = -0.4291$, averaged estimated stand error $\hat{se} = 0.1072$, the averaged distance $\hat{d} = 0.4892$, and the empirical coverage rate with 95% confidence level is 96%.

4.4 Modeling Strategy

Let $\tau_{i,k}$ be individual i's k^{th} unemployment duration. The length of $\tau_{i,k}$ is related to this individual's current state dependent variables.

Let $H_{i,k}$ be a vector of individual i's state dependent variables at the beginning of k^{th} unemployment spell. At this moment $H_{i,k}$ contains individual's unemployment benefit duration and her human capital stock. The latter is calculated using all previous employment and unemployment records. Our empirical model should able to constantly update $H_{i,k}$.

With these information in mind, we model the hazard rate of $\tau_{i,k}$ as

$$\lambda(\tau_{i,k}) = \tau_{i,k}^{\alpha} \exp(-\beta^{'} H_{i,k} + \nu_i) \tag{4.15}$$

where ν_i is the unobserved heterogeneity. We highlight the role of the duration dependence by separating $\tau_{i,k}$ from the state dependent variables $H_{i,k}$.

This hazard rate corresponds to the following AFT model

$$\tau_{i,k}^{\alpha+1} = (\alpha+1) \exp(\beta^{'} H_{i,k} - \nu_i) u_{i,k} \tag{4.16}$$

where $u_{i,k} \rightsquigarrow exp(1)$ are i.i.d across both individuals and duration.

To see this, recall the hazard rate is defined as $f(x)/(1-F(x))$ where $f(x), F(x)$ are random variable's p.d.f and c.d.f respectively. The c.d.f of $\tau_{i,k}$ is

$$\begin{aligned} F_{\tau_{i,k}}(x) = Pr(\tau_{i,k} \leq x) &= \Pr((\alpha+1)^{1/(\alpha+1)} \exp(\beta H_{i,k} - \nu_i)^{1/(\alpha+1)} u_{i,k}^{1/(\alpha+1)} \leq x) \\ &= Pr(u_{i,k} \leq 1/(\alpha+1) x^{\alpha+1} \exp(-\beta H_{i,k} + \nu_i)) \\ &= 1 - \exp(-1/(\alpha+1) x^{\alpha+1} \exp(-\beta H_{i,k} + \nu_i)) \end{aligned} \tag{4.17}$$

Plugging $F_{\tau_{i,k}}(x), f_{\tau_{i,k}}(x)$ into the hazard rate definition, we ended up with previous asserted.

We choose the error terms to be i.i.d exponential distributed because without exposed to the state dependent structure, it is reasonable to assume the unemployment counting process to be the standard Poisson process and the intervals (or duration) of a Poisson process is $exp(1)$ distributed.

Taking the first ratio transformation, we have

$$\tilde{\tau}_{i,k}^{\alpha+1} = \frac{\tau_{i,k}^{\alpha+1}}{\tau_{i,k-1}^{\alpha+1}} = \exp(\beta(H_{i,k} - H_{i,k-1}))\tilde{u}_{i,k} \tag{4.18}$$

where $\tilde{u}_{i,k} = u_{i,k}/u_{i,k-1}$. In general, the ratio distribution of $\tilde{u}$ is not easy to obtain, however, with $u \rightsquigarrow exp(1)$, the calculation is quite easy. One may show

$$F_{\tilde{u}}(x) = \frac{x}{1+x} \tag{4.19}$$

Thus

$$F_{\tilde{\tau}_{i,k}}(x) = \frac{\exp(\beta H_{i,k-1} - \beta H_{i,k})x^{\alpha+1}}{1 + \exp(\beta H_{i,k-1} - \beta H_{i,k})x^{\alpha+1}} \tag{4.20}$$

Constructing a counting process $\tilde{N}(t)$ whose inter-event duration is $\tilde{\tau}_{i,k}$. We may write the counting process's cumulative intensity $\tilde{\Lambda}(t)$ using Equation 4.8.

4.5 Empirical Results and Discussion

The following table reports the main results.

Table 4.1:: Main Results

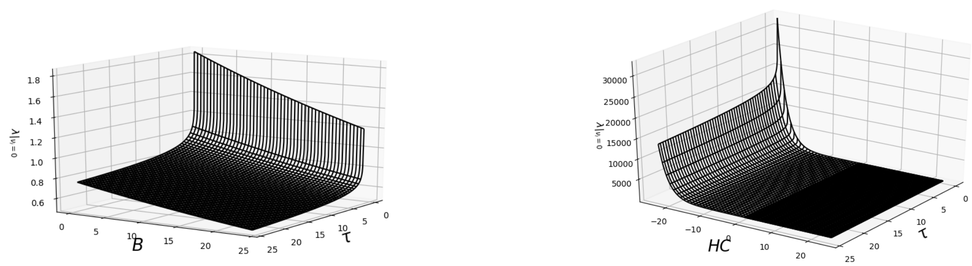
α	-0.0882302***
	(0.02044)
Benefit Duration	0.01782836*
	(0.01002)
Human Capital	0.40773478***
	(0.00858)

Note: *p<0.1; **p<0.05; ***p<0.01

A negative α suggests that the longer one is in the unemployment, the harder to exit from the current status, this result is consistent with previous literature. Both unemployment benefit duration and human capital are positive, indicating that higher volume of benefit duration (human capital) leads to a longer unemployment duration.

The economical interpretation could be as follow: different benefit duration and the human capital volumes may alter one's employment preference. When the volumes are high, it is more likely for an individual to search and wait for a better contract, hence a longer unemployment duration; When the volumes are low, individuals' preferences may change and one may become more likely to accept some low-quality job offers whenever are available.

The following figures summarize our estimates.



(a) w.r.t unemployment benefit duration (b) w.r.t human capital stock

Figure 4.1: Hazard Rate for overall Unemployment Duration

We also investigated how different sub-populations would react to these state-dependent structures. To this end, we divide the sample by education gender.

Education. We classify individuals who did not enter senior high school or senior professional education as low education, others are high education. The following table reports the results. The difference between these two groups are quite large. First, the reaction to the duration dependence is opposite. Low education individuals are more likely to exit from the unemployment as the unemployment duration grow, but for high education individual, this is not the case. A reasonable explanation is that low education people are willing to accept bad temporary contracts, while high education people tend to wait for better terms.

Low education individuals are also more sensitive to the benefit duration, while high education people seem care little about the unemployment benefit duration. But both group are respond to the human capital stock positively.

Table 4.2:: Sub-population Results, Education

	Low Education	High Education
α	0.15151556***	-0.25632063***
	(0.04274)	(0.02235)
Benefit Duration	0.05544162***	0.00012458
	(0.01176)	(0.02225)
Human Capital	0.46336057***	0.31247732***
	(0.01639)	(0.00899)

Note: *p<0.1; **p<0.05; ***p<0.01

Gender. As shown by the next table, it seems that there is no big different reactions towards state dependent variables across gender.

Table 4.3:: Sub-population Results, Gender

	Male	Female
α	-0.08627595***	-0.1125127***
	(0.02578)	(0.03329)
Benefit Duration	0.02389269**	0.02693626*
	(0.01172)	(0.01555)
Human Capital	0.41330633***	0.38903453***
	(0.01128)	(0.01302)

Note: *p<0.1; **p<0.05; ***p<0.01

4.6 Future Research

One distinctive characteristic of the Spanish labor market is the extensive use of temporary employment, as a result, the probability of receiving a permanent job offer is much lower than the one of receiving a temporary offer (Bover & Gómez 2004). The temporary offers are fixed-term labor contracts. These contracts have lower firing costs than the permanent ones. Most new hirings were under the low cost fixed-term contracts and temporary workers were typically dismissed at the end of the maximum contract length to avoid transfer to the high-cost permanent contracts (Bover et al. 2002). In the future, I am also interested in investigating how temporary contracts affects an individual's job turnover. Specifically, how large number of temporary contracts impact the duration of both employment and unemployment duration. Does large number of temporary contracts increase of decrease the probability to find a permanent job? Does large number of temporary contracts increase of decrease the risk of being lay off?

4.6.1 Unemployment Duration Model that distinguish Permanent and Temporary Contracts

Model with Exclusive Assumption

In general, we are unable identify the model if one distinguishes the duration from unemployment to permanent and temporary contracts if no additional assumption is made. This is because we can not write a accelerate failure time model with cause specific structure. We use the following figure to illustrate.

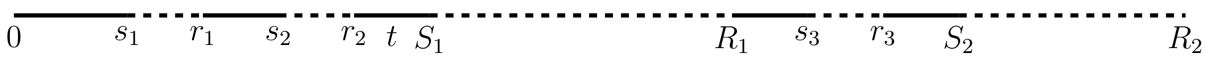


Figure 4.1: A possible realization of job turnover

Here lower cap letter s denotes the beginning dates of temporary contracts, lower cap letter r denotes the end dates of temporary contracts. Correspondingly, upper letter S denotes the beginning dates of permanent contracts and the upper letter R denotes the ending dates of permanent contracts. The dashed lines are then the employment spells, while the solid lines are unemployment spells.

Now consider we are at time t, during an unemployment spell. Without additional assumption, it is possible that the next employment could be both permanent and temporary, that is the intensity is non-negative for both finding a permanent or a temporary job. The two non-negative intensities create a cause specific structure for the unemployment duration $\tau \in [0, S_1 - r_2]$.

To better understanding the consequence of such structure, let's focus on modeling the duration from the unemployment until a permanent job. There are two interpretation for this duration. The first one would be the total duration until the individual find a permanent job, using Figure 4.1, this could be the duration of $S_1 - 0$ and $S_2 - R_1$. In these duration, an individual might have several temporary jobs. The other interpretation would be the duration from current unemployment to a permanent job. In Figure 4.1, these are $S_1 - r_2$ and $S_2 - r_3$. There are no other jobs, temporary or permanent in these duration.

If we allow the cause specific structure, we are using the first interpretation. This is precisely because the hazard rate for such permanent job duration exists and is positive even an individual existed from unemployment by a mean of temporary jobs. The temporary jobs will change the state dependent variable. Put it another way, temporary jobs act as external shocks that lead to the change of hazard rate for a permanent job duration. These external shocks make it impossible to link the hazard rate to a accelerated failure model and hence we are unable to do the first ratio transformation to swipe out the unobserved heterogeneity.

To solve this problem, we assume the **exclusive assumption**:

EA. There is no cause specific structure, individuals will subjectively decide the type of next employment.

The exclusive assumption means that if an individual decide to find a permanent (temporary) job, she will not accept a temporary (permanent) offer and remain unemployed until the desired type of offer arrives. With this assumption, we are adopting the second interpretation of the permanent job duration, the following figure illustrate:

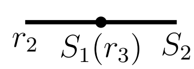


Figure 4.2: Permanent Contract Process

For the duration from unemployment to temporary contracts, we study the following counting process:

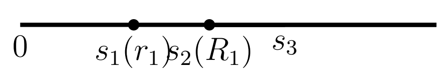


Figure 4.3: Temporary Contract Process

The corresponding AFT models are

$$\tau_{i,k,J}^{\alpha_J+1} = (\alpha_J + 1)\exp(\beta_J' H_{i,k} - \nu_i)u_{i,k} \tag{4.21}$$

where $J \in \{P, T\}$ is contract type, permanent(P) or temporary(T). $H_{i,k}$ is a vector of state dependent variables including the number of temporary contracts so far, the unemployment benefit duration and the human capital stock.

Model without Exclusive Assumption

As mentioned before, without the exclusive assumption, we are unable to write down the accelerated failure time model, hence unable to swipe out the unobserved heterogeneity term. Arguably, this assumption is strong, and if readers are willing to trade-off by ignoring the unobserved heterogeneity, we may study the duration from unemployment until find a permanent job. Using Figure 4.1 as an example, we are able to study the following permanent contract process (the first interpretation,temporary contract process can be constructed in a similar way).

Figure 4.4: Duration from Unemployment until find a Permanent Job

Notice that during temporary employment spells (dashed lines in Figure 4.4), the intensity of finding a permanent job should be zero. The hazard rate is constructed as follow:

$$\lambda(\tau_{i,k,J}) = \begin{cases} \tau_{i,k,J}^{\alpha_J} \exp(-\beta_J' H_{i,k,J} + \gamma_J' X_i), \tau_{i,k,J} \in \text{unemployment period} \\ 0, \tau_{i,k,J} \in \text{employment period} \end{cases} \tag{4.22}$$

where $\tau_{i,k,J}$ is the duration before finding a type $J \in \{P, T\}$ job. $H_{i,k,J}$ is a vector of state dependent covariates, X_i is a vector of observed time-independent individual heterogeneity such as gender, education, region, income,etc.

4.6.2 Employment Duration that distinguish Permanent and Temporary Contracts

To complete the analysis of job turnover circle, we need to study the employment duration. Here we are primarily interested in the permanent contract employment duration, as in most temporary contracts, the length of job is pre-specified.

It is clear there would be no cause specific structure in these duration, as individuals are very clear what type of they are in. Thus we might just use the model as described in Equations 4.15 or 4.16. The only differences are that here duration $\tau_{i,k}$ are employment duration and the state variables $H_{i,k}$ include human capital stock and the number of temporary jobs so far.

4.7 Conclusion

We developed a new estimator to separate the state dependent effect from the unobserved heterogeneity in duration models. The new estimator allows a fixed effect unobserved heterogeneity and is an extension to the classical dynamic panel data models.

The duration are history dependent and the dependent structure is flexible, no stationary or auto-regressive structure is needed. The key assumption is that the duration can be written as a multiplicative form of accelerated failure time model. To swipe out the unobserved heterogeneity, we do a first ratio transformation on the duration.

Using Spanish social security data, we find that individuals respond to state dependent variables. In general, during the current unemployment period, the longer the spell is, the harder to exit from unemployment (duration dependent structure). However, a sub-population analysis reveals that low education workers have the opposite duration dependent structure. Unemployment benefit duration and the human capital stock also play important roles: workers treat these state dependent variables as 'assets', small volumes of these state dependent variables lead to a shorter unemployment duration.

References

Ahn, H. J., & Hamilton, J. D. 2019, Journal of Business & Economic Statistics, 1

Aït-Sahalia, Y., Cacho-Diaz, J., & Laeven, R. J. 2015, Journal of Financial Economics, 117, 585

Allen, S. G. 1981, The Review of Economics and Statistics, 77

Anderson, T. W., & Hsiao, C. 1981, Journal of the American statistical Association, 76, 598

Arellano, M., & Bover, O. 1995, Journal of econometrics, 68, 29

Aron-Dine, A., Einav, L., & Finkelstein, A. 2012, The RAND health insurance experiment, three decades later, Tech. rep., National Bureau of Economic Research

Aron-Dine, A., Einav, L., Finkelstein, A., & Cullen, M. 2015, Review of Economics and Statistics, 97, 725

Bacry, E., & Muzy, J.-F. 2014, Quantitative Finance, 14, 1147

Barmby, T., Orme, C., & Treble, J. 1995, Labour Economics, 2, 53

Barmby, T. A., Orme, C. D., & Treble, J. G. 1991, The Economic Journal, 101, 214

Blundell, R., & Bond, S. 1998, Journal of econometrics, 87, 115

Bond, S. R., Nauges, C., & Windmeijer, F. 2005, CEMMAP Working Paper No. CWP07/05.

Bover, O., Arellano, M., & Bentolila, S. 2002, The Economic Journal, 112, 223

Bover, O., & Gómez, R. 2004, investigaciones económicas, 28

Brot-Goldberg, Z. C., Chandra, A., Handel, B. R., & Kolstad, J. T. 2017, The Quarterly Journal of Economics, 132, 1261

Cardon, J. H., & Hendel, I. 2001, RAND Journal of Economics, 408

Cutler, D. M., & Zeckhauser, R. J. 2000, Handbook of health economics, 1, 563

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

Dalton, C. M. 2014, International Journal of Industrial Organization, 37, 178

de Paula, A., & Honore, B. 2017, Quantitative Economics

Delgado, M. A., & Kniesner, T. J. 1997, Review of Economics and Statistics, 79, 41

Dolado, J. J., Felgueroso, F., & Jansen, M. 2013, Intereconomics, 4, 209

Duflo, E., Hanna, R., & Rya, S. P. 2012, The American Economic Review, 102, 1241

Einav, L., Finkelstein, A., & Schrimpf, P. 2015, The quarterly journal of economics, 130, 841

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. 1996in , 226–231

Fevang, E., Markussen, S., & Røed, K. 2014, Journal of Labor Economics, 32, 305

Fister-Gale, S. 2003, Workforce, 82, 72

Gaure, S., Røed, K., & Zhang, T. 2007, Journal of Econometrics, 141, 1159

Greene, W. H., & Hensher, D. A. 2010, Modeling ordered choices: A primer (Cambridge University Press)

Handel, B. R., Kolstad, J. T., & Spinnewijn, J. 2015, Information frictions and adverse selection: Policy interventions in health insurance markets, Tech. rep., National Bureau of Economic Research

Harris, R. D., & Tzavalis, E. 1999, Journal of econometrics, 91, 201

Heckman, J., & Singer, B. 1984, Econometrica: Journal of the Econometric Society, 271

Heckman, J. J. 1981, in Studies in labor markets (University of Chicago Press), 91–140

—. 1991, The American Economic Review, 81, 75

Honoré, B. E. 1993, The Review of Economic Studies, 60, 241

Honore, B. E., & De Paula, Á. 2010, The Review of Economic Studies, 77, 1138

Karr, A. 1991, Point processes and their statistical inference, Vol. 7 (CRC press)

Keeler, E. B., Newhouse, J. P., & Phelps, C. E. 1977, Econometrica, 641

Keeler, E. B., & Rolph, J. E. 1988, Journal of Health Economics, 7, 337

Kopperschmidt, K., & Stute, W. 2013, Stast. Sinica, 23, 1273

Kowalski, A. E. 2015, International journal of industrial organization, 43, 122

Lewis, P. A., & Shedler, G. S. 1979, Naval Research Logistics Quarterly, 26, 403

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., & Leibowitz, A. 1987, The American economic review, 251

Markussen, S., Røed, K., Røgeberg, O. J., & Gaure, S. 2011, Journal of health economics, 30, 277

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. 2012, Journal of the American Statistical Association

Nerlove, M. 2014, Œconomia. History, Methodology, Philosophy, 281

Ogata, Y. 1981, Information Theory, IEEE Transactions on, 27, 23

Ogata, Y., & Katsura, K. 1988, Annals of the Institute of Statistical Mathematics, 40, 29

Schmillen, A., & Umkehrer, M. 2017, International Labour Review, 156, 465

Steel, R. P., Rentsch, J. R., & Van Scotter, J. R. 2007, Journal of Management, 33, 180

Steers, R. M., & Rhodes, S. R. 1978, Journal of applied Psychology, 63, 391

Van den Berg, G. J. 2001, Handbook of econometrics, 5, 3381

Van den Berg, G. J., & Van Ours, J. C. 1996, Journal of Labor Economics, 14, 100

Zhuang, J., Ogata, Y., & Vere-Jones, D. 2002, Journal of the American Statistical Association, 97, 369