The Strategic Behaviour in Work Absence: A Dynamic view[*]

Yuhao Li
yuli@eco.uc3m.es
Latest Version Here

Departamento de Economia, Universidad Carlos III de Madrid

October, 2018

Abstract

We use the self-exciting processes to study individuals' absence behaviours. Such behaviours are dynamic and strategic 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.

JEL. C51, C41, J22, J32, C13

Keywords. Work Absence, Self-Exciting Process, Count Data Regression, Duration Analysis

[*] I am extremely grateful to my advisor Miguel A. Delgado for his encouragement and weekly guidance throughout this project and to Winfried Stute for his inspiration. I am thankful to Tim Barmby and Michael Nolan for their generous permission to use the dataset. I would also like to thank: Carlos Velasco, Juan Carlos Escanciano, Jesus Gonzalo, Jan Stuhler, Nazarri Salish, Rui Cui, Minghai Mao, Junyi Peng Zhou, Julius Vainora and seminar participants at the Universidad Carlos III de Madrid and the UC3M-Cambridge-LSE Econometrics Workshop. All remaining errors are mine.

1 Introduction

The primary purpose of this paper is to investigate the dynamic strategic behaviours in work absences, where individuals may take past experiences into absence decision-making considerations. A typical example 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 behaviour. 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 and 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 behaviour 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 behaviour. However, we prefer to use the term 'strategic absence behaviour'. Since from a perspective of utility optimisation, what workers do is to choose strategies (whether to take an absence and how long an absence duration should be, etc.) based on current absence scores, other covariates and shocks to maximise the individual utility.

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 process and recovery process are different. Individuals may encounter shocks (e.g. illness) and seek absences, but 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 leaves 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 strategic. 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 (e.g., Delgado and Kniesner (1997)) and the duration models (Barmby et al. (1991), Markussen et al. (2011), and Fevang et al. (2014)), the self-exciting process has the advantage of modelling the dynamic decision-making process. In the count data regression models, 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 (e.g., Honoré (1993)) do exist, but they are difficult to apply.

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 strategic behaviour in the work absences. Specifically, we observe strategic behaviour in short-term absences in both incidence and recovery processes. While in the long-term absences, strategic behaviour plays an insignificant role in the decision making processes.

Second, we provide a modelling method that complements to count data regression and duration analysis models. The i.i.d assumption of absence events is not required in our self-exciting approach. In addition to the unobserved heterogeneity, because of the state dependent nature, the individual's own history could also be a source of heterogeneity.

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 behaviour 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.

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 evidences 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.

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.

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 analysis.

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, 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 1 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.

[Insert Figure 1 Here]

2.2 Preliminary Results

Conventionally, count data regression (Delgado and 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: 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 binomial distribution to describe 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 Appendix A.

[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.

Table 1 summarises 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 and Kniesner, 1997).

[Insert Table 1 Here]

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 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 A documents the details of this model. The first column in Table 2 reports the estimation results of this standard duration model.

[Insert Table 2 Here]

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 and 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 A. Column 2 of Table 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 behaviour. 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 multiple-spell models break the independence assumption and allow lagged duration dependence (e.g.,Honoré (1993)). However, this lagged duration model is in a panel setting, and its hazard rate can be very difficult to study, especially when one distinguishes short-term and long-term absences, as these two panels are shocks to 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.

2.3 State Dependence Test

The phenomenon that 'past experiences affect future ones' is called the state dependence. To ask the existence of strategic behaviour is equivalent to test the state dependence.

Heckman (1981) proposes to use the following strategy to test the state dependence in a panel data. Define an equispaced intervals of time, say a week. For each period, we observe individual's decision $d(i,t), i = 1, 2, \cdots, N; t = 1, 2, \cdots, T$ on whether to ask for a leave. $d(i,t) = 1$ if individual i choose to be absence in period t, and $d(i,t) = 0$ otherwise. For each individual, the history is made up with decisions $d(i,t), t = 1, 2, \cdots, T$, we may utilize this history data of sufficient length in the sample to estimate a regression of current absence status on previous absence status. If previous absence status has no effect on the current probability of absence, we may conclude that there is no state dependence structure.

Notice that one must permit each person to have the individual fixed effect or intercept in the regression. Otherwise, one would face the danger that individual differences in absence probabilities will be correlated with past absence status.

Specifically, for each individual, write the regression

$$d(i,t) = \nu_i + \delta \sum_{t'<t} d(i,t') + U(i,t), t + 1, \cdots, T \tag{1}$$

where $U(i,t)$ is a mean zero random variable of innovations uncorrelated with other innovations $U(i,t'), t' \neq t$, ν_i is an individual-specific effect.

After obtaining the fixed effect estimator, we perform the following test:

$$H_0 : \delta = 0 \text{ vs } H_1 : \delta \neq 0$$

At this stage, we can only check whether the incidence events have state dependence but not the recovery events. There is no economic meaning to say 'whether previous recoveries have effects on the current one.'. That is, 'the interpretation of the statistical model as an economic model is not clear' (Heckman, 1981). In the later section, we will address the issue of state dependence in the recovery process by another method.

Table 3 shows the fixed effect estimation results for both short-term and long-term incidence events. The results strongly favour the existence of state dependence.

[Insert Table 3 Here]

2.4 The Nature of the Problem

In this subsection, we try to illustrate the econometric challenges when modelling the strategic behaviour. To have a better understanding, considering Figure 2, which demonstrates 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.

[Insert Figure 2 Here]

Suppose now we are at $t \in [r_1, s_2)$ and the goal is to investigate how likely the next absence is going to occur at time $t + dt$. To account for the strategic behaviour, 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 behaviour.

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 behaviour 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 behaviour among work absences.

3 Econometric Models for Work Absenteeism

In this section, we use self-exciting processes to construct our work absenteeism models. We first present some introductions to the self-exciting process, followed by a detailed discussion about the models. We will end this section by illustrating how to estimate the models.

3.1 Introducing the Self-Exciting Process

As mentioned before, a self-exciting process is a particular counting process. A counting process (non-tied) can be regarded as a step function:

$$N(t) = \sum_{i=1}^{\infty} \mathbb{I}\{t_i \leq t\} \tag{2}$$

Its value adds one at time t if and only if an event occurs at this time. The counting process not only tells how many events have occurred before time t but also indicates exact occurrence times for each event.

The Doob-Meyer decomposition theorem states that any counting process can be decomposed into a predictable cumulative intensity part and a martingale part, i.e.

$$\begin{aligned} N(t) &= \Lambda(t) + M(t) \\ N(dt) &= \lambda(t)dt + M(dt) \end{aligned} \tag{3}$$

$\Lambda(t)$ is the predictable cumulative intensity, also known as the compensator. In brief, predictability means conditional on the information just before present (say $t-$), we should know the value of $\Lambda(t)$. Rigorously, $\Lambda(t)$ is predictable w.r.t a filtration $\mathcal{F}_{t-}$ if and only if $\Lambda(t)$ is $\mathcal{F}_{t-}$-measurable.$\lambda(t)dt = \Lambda(dt)$ is the associated intensity and $M(t)$ is the martingale satisfying $M(0) = 0$.

The decomposition is unique: if there exists another decomposition:

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

we then have $\Lambda(t) = \tilde{\Lambda}(t)$ and $M(t) = \tilde{M}(t)$. In addition, we have

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

The proofs of the uniqueness and this equation can be found in Appendix B.

The (cumulative) intensity is usually conditioned on a filtration $\mathcal{F}_{t-}$:

$$\lambda(t|\mathcal{F}_{t-}) = \lim_{\Delta t \to 0} \frac{\mathbb{E}(N([t, t + \Delta t])|\mathcal{F}_{t-})}{\Delta t} \tag{4}$$

If the filtration is generated by the underlying counting process itself:

$$\mathcal{F}(t-) = \sigma(N(s) : s < t) \tag{5}$$

we call the counting process $N(t)$ a self-exciting process. One may study the process $N(t)$ through modelling and estimating $\lambda(t|\mathcal{F}_{t-})$ $(\Lambda(t|\mathcal{F}_{t-}))$. We postpone the estimation method in later subsection.

We may generalize the filtration by including other relevant information. Let $\mathcal{H}_{t-} = \mathcal{H}_0 \vee \mathcal{F}_{t-}$ be the conditional filtration, where $\mathcal{H}_0$ is the $\sigma - algebra$ generated by some external covariates such as age, sex, etc. We interpret this filtration as the 'whole history'. Notice that $\mathcal{H}_0$ can also be time-dependent, i.e., $\mathcal{H}_0 = \mathcal{H}_0(t-)$.

3.2 The Work Absenteeism Models

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

$$N_i^1(t) = \sum_{j=1}^{\infty} I\{t_j \leq t\} \tag{6}$$

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

$$N_i^2(\tau) = \sum_{j=1}^{\infty} I\{\tau_j \leq \tau\} \tag{7}$$

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 2 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.

3.2.1 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:

$$h(X_i, \nu_i) = \exp(\nu_i) exp(\mathbf{X}_i' \boldsymbol{\gamma}_{1k}) \tag{8}$$

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. The likelihood function is constructed in the same manner as in the section 2.2.2.

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

$$\lambda_{i,1k}(t) = \begin{cases} \lambda_{1,k}(\mathbf{X}_i)\lambda_{2,k}(t)\Big(\lambda_{3,k}(t) + \lambda_{4,k}(t)\Big), t \in \text{attendance period} \\ 0, t \in \text{absence period} \end{cases} \tag{9}$$

where:

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

$\lambda_{1,k}(\mathbf{X}_i)$ 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 behaviour 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 illustrate the idea.

[Insert Figure 3 Here]

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

3.2.2 A Model for Recovery Processes

The recovery intensity has the following parts:

$$\begin{aligned}
&\lambda_{5,k}(\boldsymbol{X}_i) = exp(\boldsymbol{X}_i'\boldsymbol{\gamma}_{k1}) \\
&\lambda_{6,k}(\tau) = |\beta_{k1}|exp(\beta_{k1}H_i(\tau)) \\
&\lambda_7(\tau) = 1 + |\beta_{k2}|exp(\beta_{k2}(\tau - t_{N_{i,13}^1(\tau-)})); k = 2, 3
\end{aligned} \tag{10}$$

these parts have similar roles as in the incidence intensities: $\lambda_{5,k}(\boldsymbol{X}_i)$ 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 2'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:

$$\lambda_{i,k1}(\tau) = \begin{cases} \lambda_{5,k}(\boldsymbol{X}_i)(\lambda_{6,k}(\tau) + \lambda_7(\tau)), \tau \in \text{absence period} \\ 0, \tau \in \text{attendance period} \end{cases} \tag{11}$$

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 durations are i.i.d. In that case, a standard duration analysis would be a useful modelling alternative.

3.2.3 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. Since 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 incidence process, especially the short-term process, the primary unobserved heterogeneity is an individual's working attitude. We may approximate it using the (moving) average attendance duration $\tilde{d}(t)$ along with some absence score adjustments.

Using Figure 2 and its notation to help define $\tilde{d}(t)$:

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

Notice here, we only consider the attendance durations among short-term absences 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 working attitude index at time t. A higher index suggests a more hard-working individual. $G(\cdot)$ 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:

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

the structural of the overall incidence intensity remains the same.

In the recovery processes, the primary unobserved heterogeneity is an individual's ability to recovery. Similarly, we could use history information to approximate it.

This time, we choose the (moving) average recovery time $\tilde{c}(t)$, which is defined as (again, use Figure 2's notation),

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

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

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

where $R(t)$ is a recovery index with decreasing order: a higher recover index means a longer recovery time and vice versa. $G(\cdot)$ 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:

$$\lambda_{5,k} = exp(\boldsymbol{X}_i' \boldsymbol{\gamma}_{k1}) exp(\gamma' \tilde{R}(t))$$

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.

3.3 How to estimate the models

Recall the Doob-Meyer decomposition, we have $\mathbb{E}N(t) = \mathbb{E}\Lambda(t)$. One may obtain the estimator by minimising the distance between the counting process and its cumulative intensity. Inspired by this idea, Kopperschmidt and Stute (2013) developed a minimum distance estimation method. This method only requires the observations (individuals) to be i.i.d. It does not assume the differentiability of the cumulative intensity and allow unexpected jumps in the intensity function. Here, we provide a summary. Technical details can be found in their paper.

Formally, let $N_1, ..., N_n$ be i.i.d copies of n observed counting process that are conditional on the increasing filtrations $\mathcal{H}_i(t), 1 \leq i \leq n$, which are comprised by the counting process N_i as well as some other external information. Let $\Lambda_{v,i}(t|\mathcal{H}_i(t-))$ with $v \in \Theta \subset \mathbb{R}^d$ be a given class of parametric cumulative intensities.

We set,

$$< f, g >_\mu = \int_0^T fg d\mu \tag{12}$$

where T is the terminating time. If f and g are square integrable functions w.r.t. μ. The corresponding semi-norm is,

$$||f||_\mu = [< f, f >_\mu]^{1/2} \tag{13}$$

Let,

$$\bar{N}_n = \frac{1}{n}\sum_{i=1}^{n} N_i; \bar{\Lambda}_{v,n} = \frac{1}{n}\sum_{i=1}^{n} \Lambda_{v,i} \tag{14}$$

We call the former the averaged counting process and the later the averaged cumulative intensity. Naturally the associated averaged innovation martingale is,

$$d\bar{M}_n = d\bar{N}_n - d\bar{\Lambda}_{v_0,n} \tag{15}$$

If, for μ, we take $\mu = \bar{N}_n$, the quantity $||\bar{N}_n - \bar{\Lambda}_{v,n}||_{\bar{N}_n}$ is then an overall measurement of fitness of $\bar{\Lambda}_{v,n}$ to $\bar{N}_n$. The estimator v_n is computed as,

$$v_n = arg \inf_{v\in\Theta} ||\bar{N}_n - \bar{\Lambda}_{v,n}||_{\bar{N}_n} \tag{16}$$

Kopperschmidt and Stute (2013) have shown the consistency and asymptotic normality of this estimator. For technical detail, we refer readers to the Appendix B.

Thus for incidence processes, we collect starting dates of absences that have previous absent records to construct individual counting processes $N_{1k}(t), k \in \{2, 3\}$. The distance function is then:

$$||\bar{N}_{1k,n} - \bar{\Lambda}_{1k,n}||_{\bar{N}_{1k,n}}$$

where

$$\bar{\Lambda}_{1k,n}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \lambda_{i,1k}(s)ds$$

Similarly, the recovery intensities have the following distance function:

$$||\bar{N}_{k1,n} - \bar{\Lambda}_{k1,n}||_{\bar{N}_{k1,n}}$$

where $N_{k1}(t), k \in \{2, 3\}$ are made of ending dates of absences, and

$$\bar{\Lambda}_{k1,n}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \lambda_{i,k1}(s)ds$$

Conventionally, a likelihood-based method is used to obtain estimates. One may do so by exploiting the fact that $f(t) = \lambda(t)\exp(-\Lambda(t))$. For example, the Hawkes process, a special case of the self-exciting process that is widely used in the high-frequency financial analysis, uses MLE to obtain the consistent estimators (e.g., Aït-Sahalia et al. (2015),Bacry and Muzy (2014) and Bowsher (2007)). One requirement of using likelihood-based methods is the predictability of the cumulative intensity Λ w.r.t the filtration $\sigma(N(s) : s < t)$. However, as pointed by Kopperschmidt and Stute (2013), 'in many complicated economic situations, there is little reason to maintain such assumption.' Instead, the cumulative intensity should be predictable to the 'whole history' $\mathcal{H}_{t-}$, which may include external shocks or impulses. If that is the case, the model is mostly not dominated, and likelihood functions are often too difficult to write down. This dominating problem is their main motivation to develop the mentioned minimum distance estimator.

In our application, short-term absences can be regarded as the external shocks to long-term absences and vice verse. We use Figure 2 again to help to understand. The short-term absences will affect the intensities for long-term incidence and recovery processes (assuming we have strategic behaviour in both processes) mostly by the cumulative absence time. Since the locations of starting and recovery points (which determine how many short-term absences before this long-term absence) as well as the durations of the absences $r_j - s_j, j \in \{1, 2\}$ are stochastic, the short-term absences are external shocks to the long-term absence and vice versa.

Another source of external shock is the 'switches' between incidence and recovery events. As mentioned before, it is unusual for a worker to ask for a second leave without returning to work from the first absence, during an absence period, we would expect no new incidence occurs. That is, if $t \in [s_j, r_j)$, then $\lambda_{incidence}(t) = 0$. The same argument holds for the attendance period. It is meaningless to talk about a recovery event if there is no incidence occurs in the first place: if $t \in [r_{j-1}, s_j)$, then $\lambda_{recovery}(t) = 0$. Hence, there is a 'switch' such that during an absence (attendance) period, it closes the gate for an occurrence of new incidence (recovery). In the timeline, the locations of these 'switches' are stochastic, these switches then also serve as external shocks to our interested intensities.

Hence, it is, if not impossible, tough to apply the likelihood based methods in our empirical study.

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.

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 4 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 and Hensher, 2010).

[Insert Table 4 Here]

Next, we present our main results for incidence intensities in Table 5. 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.

[Insert Table 5 Here]

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 behaviour 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 behaviour strongly indicate that short-term absences are more likely to be 'voluntary' leaves, 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 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.

4.1.1 Robustness Check

For the group heterogeneity, we need first to pin down the number of groups k. Recall the results in Table 4, 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 C.

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.

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 6 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.

[Insert Table 6 Here]

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.

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 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:

$$\begin{aligned} &h_i(X_i, \nu_i) = exp(X_i\boldsymbol{\beta}' + \nu_i) \\ &H_i(T) = h_i(X_i, \nu_i)T \end{aligned} \tag{17}$$

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

$$L_i(\nu_i) = \prod_{j \in S_i} exp(-H_i(t_j))h_i \tag{18}$$

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 and Singer (1984)'s NPMLE. The likelihood function is

$$L = \prod_{i=1}^{N} \mathbb{E}[L_i(\nu_i)] = \prod_{i=1}^{N} \sum_{l=1}^{Q} p_l L_i(\nu_l), \sum_{l=1}^{Q} p_l = 1 \tag{19}$$

Table 7 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.

[Insert Table 7 Here]

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.

[Insert Figure 4 Here]

4.3 A Closer Look at the Strategy Behaviour 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 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 8 reports the results.

[Insert Table 8 Here]

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

$$\begin{aligned} H_0 &: \beta_1 = \beta_1 = \beta_2 = 0 \\ H_1 &: \beta_1 \neq 0, \beta_1 \neq 0, \beta_2 \neq 0 \end{aligned}$$

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

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = 0 \\ H_1 &: \beta_1 \neq 0, \beta_2 \neq 0 \end{aligned}$$

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 5 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 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.

[Insert Figure 5 Here]

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 $H(t)$.

[Insert Table 9 Here]

Table 9 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.

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 behaviour 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.

5.1 The incentive to the strategic absence behaviour

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.

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 shock, and a higher value of e indicates a more severe accident. 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:

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

where $a(\cdot)$ 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

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

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. 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

$$\begin{aligned} Pr(D = 1) &= Pr(\epsilon_2 - \epsilon_1 < R(A + a(\mathbb{E}(e))) + g(a(\mathbb{E}(e))) + C_1 - R(A) - C_2) \\ &= F_\epsilon(R(A + a(\mathbb{E}(e))) + g(a(\mathbb{E}(e))) + C_1 - R(A) - C_2) \end{aligned}$$

where $\epsilon = \epsilon_2 - \epsilon_1$

Since $F^{'}(\cdot) > 0, R^{'}(\cdot) < 0$ and $R^{''}(\cdot) < 0$. We have:

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

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 behaviour, 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:

$$\begin{aligned}
&\max_a R(A+a) + \omega - e + g(a) + C \\
&s.t \\
&I + w(t^c - a) + R(A+a) - C = 0
\end{aligned} \tag{20}$$

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

First order condition with respect to a yields:

$$\begin{aligned}
R^{'} + g^{'} - (w - R^{'}) = 0 \\
(w - R^{'}) = R^{'} + g^{'} > 0
\end{aligned} \tag{21}$$

By differentiating the first order condition (20) through (21), 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).

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 Honor and De Paula (2010) and de Paula and 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:

$$\int_0^{t_{ij}} K_{ij} Z_1(s)\phi_1(X_i)e^{-\rho s}ds + \int_{t_{ij}}^{\mathbb{E}(T)} Z_2(s)\phi_2(X_i)e^{-\rho s}ds \tag{22}$$

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:

$$\begin{aligned} T_{ij} &= \inf\{t_{ij} : [K_{ij}Z_1(t_{ij})\phi_1(X_i) - Z_2(t_{ij})\phi_2(X_i)]e^{-\rho t_{ij}} < 0\} \\ &= \inf\{t_{ij} : K_{ij} - Z(t_{ij})\phi(X_i) < 0\} \end{aligned} \tag{23}$$

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

$$T_{ij} = Z^{-1}(K_{ij}/\phi(X_i)) \tag{24}$$

Notice that in this structure model, the source of randomness is K_{ij}. We can re-write equation 24 as the following:

$$\ln Z(T_{ij}) = -\ln \phi(X_i) + \epsilon \tag{25}$$

where $\epsilon = \ln K_{ij}$. Equation 25 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

$$\begin{aligned} F_{T_{ij}}(t) &= Pr[K_{ij}e^{X_i^T\beta} \leq t] \\ &= Pr[K_{ij} \leq te^{-X_i^T\beta}] \\ &= 1 - \exp(-t\exp(-X_i^T\beta)) \end{aligned} \tag{26}$$

The corresponding hazard rate is

$$\begin{aligned} h_{T_{ij}}(t) &= \frac{f_{T_{ij}}(t)}{1 - F_{T_{ij}}(t)} \\ &= \exp(-X_i^T\beta) \end{aligned} \tag{27}$$

These assumptions are mainly made to compare with our reduced form model 17 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 10 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 27 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).

[Insert Table 10 Here]

6 Discussion: 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.

6.1 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. The simulation procedure is detailed in Appendix. 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.

[Insert Figure 6 Here]

Of all 100 runs, we are able to generate 33 zeros. However, the data is still equidispersion: its sample mean is $\bar{Y} = 1.05$ and sample variance is $\hat{V}(Y) = 1.067$.

Self-Exciting DGP The DGP for the self-exciting process we picked is the ETAS (epidemic-type aftershock sequence) model, it is first introduced by Ogata and Katsura (1988) and ever since it has been widely studied in seismology (e.g. Zhuang et al. (2002)). It characterises the earthquakes occurrence times and magnitudes and belongs to a marked Hawkes process family.

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

$$\lambda(t|\mathcal{F}_{t-}) = \mu + \sum_{i:t_i<t} e^{\alpha x_i} \left(1 + \frac{t - t_i}{c}\right)^{-p} \tag{28}$$

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 and follow a exponential distribution.

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

Notice that without the exciting part, the intensity degenerates to a standard homogeneous Poisson intensity. We set the parameters as $\mu = 0.01$,$\alpha = 1.98$,$c = 0.018$,$p = 0.94$ and $\delta = log(10)$. The time interval is $[0, T^*] = [0, 100]$.

The simulation method we used is called the *thinning method*, introduced by Ogata (1981),Lewis and 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 and Shedler, 1979). To fix this problem, 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 detailed in Appendix D.

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.

[Insert Figure 7 Here]

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

6.2 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 (4)),

$$\lambda(t|\mathcal{F}(\tau)) = \frac{g(t|\mathcal{F}(\tau))}{1 - G(t|\mathcal{F}(\tau))} \tag{29}$$

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:

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{Pr\{T \in (t, t + \Delta t)\}}{Pr\{T > t\}\Delta t} \\ &= \frac{f_T(t)}{1 - F_T(t)} \end{aligned} \tag{30}$$

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 being in working at time t.

Despite the similarity between (29) and (30), the intensity and the hazard rate are conceptually different. Intensity deals with reoccurring 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.

Another feature of the self-exciting process is its ability to generate quite different individual behaviours even without the unobserved heterogeneity. The behaviours of a self-exciting process are largely shaped by its history.

Using the self-exciting DGP (the ETAS model, with its intensity as in equation 28) mentioned before as our example: The exciting part of the intensity $\sum_{i:t_i<t} e^{\alpha x_i}(1 + \frac{t-t_i}{c})^{-p}$ governs the individual heterogeneity. Figure 8 presents three quite different individual events histories simulated by our ETAS DGP using the same parameter settings stated before as an example. 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 two during the time interval $[0, 100]$. Despite the hugely different behaviour, they are governed by the same intensity function.

[Insert Figure 8 Here]

7 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 behaviour in both short-term incidence and recovery process. The strategic behaviour is generated by the cumulative absence time. However, in the long-term recovery process, we have to reject the existence of strategic behaviour 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.

References

Aït-Sahalia, Y., J. Cacho-Diaz, and R. J. Laeven (2015): “Modeling financial contagion using mutually exciting jump processes,” *Journal of Financial Economics*, 117, 585–606.

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

Bacry, E. and J.-F. Muzy (2014): “Hawkes model for price and trades high-frequency dynamics,” *Quantitative Finance*, 14, 1147–1166.

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

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

Bowsher, C. G. (2007): “Modelling security market events in continuous time: Intensity based, multivariate point process models,” *Journal of Econometrics*, 141, 876–912.

de Paula, A. and B. Honore (2017): “A New Model for Interdependent Durations,” *Quantitative Economics.*

Delgado, M. A. and T. J. Kniesner (1997): “Count data models with variance of unknown form: An application to a hedonic model of worker absenteeism,” *Review of Economics and Statistics*, 79, 41–49.

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

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

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

Gaure, S., K. Røed, and T. Zhang (2007): “Time and causality: A Monte Carlo assessment of the timing-of-events approach,” *Journal of Econometrics*, 141, 1159–1195.

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

Heckman, J. and B. Singer (1984): “A method for minimizing the impact of distributional assumptions in econometric models for duration data,” *Econometrica: Journal of the Econometric Society*, 271–320.

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

——— (1991): "Identifying the hand of past: Distinguishing state dependence from heterogeneity," *The American Economic Review*, 81, 75–79.

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

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

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

Lewis, P. A. and G. S. Shedler (1979): "Simulation of nonhomogeneous Poisson processes by thinning," *Naval Research Logistics (NRL)*, 26, 403–413.

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

Nerlove, M. (2014): "Individual Heterogeneity and State Dependence: From George Biddell Airy to James Joseph Heckman," *Œconomia. History, Methodology, Philosophy*, 281–320.

Ogata, Y. (1981): "On Lewis' simulation method for point processes," *IEEE Transactions on Information Theory*, 27, 23–31.

Ogata, Y. and K. Katsura (1988): "Likelihood analysis of spatial inhomogeneity for marked point patterns," *Annals of the Institute of Statistical Mathematics*, 40, 29–39.

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

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

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002): "Stochastic declustering of space-time earthquake occurrences," *Journal of the American Statistical Association*, 97, 369–380.

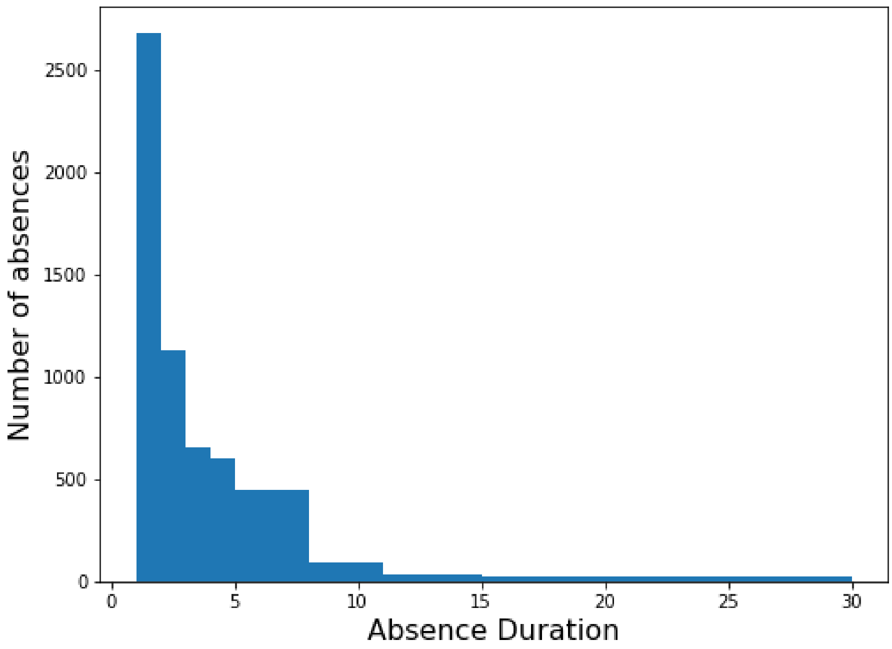


Figure 1: Most frequent absence durations

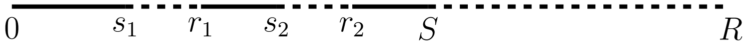


Figure 2: A possible realization of absences

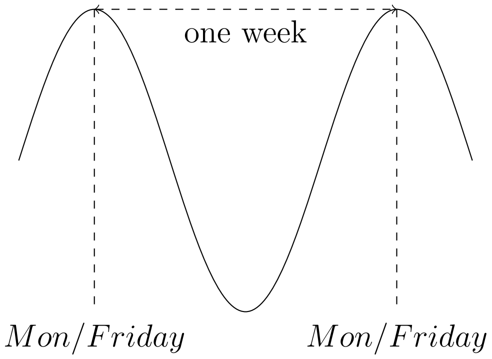
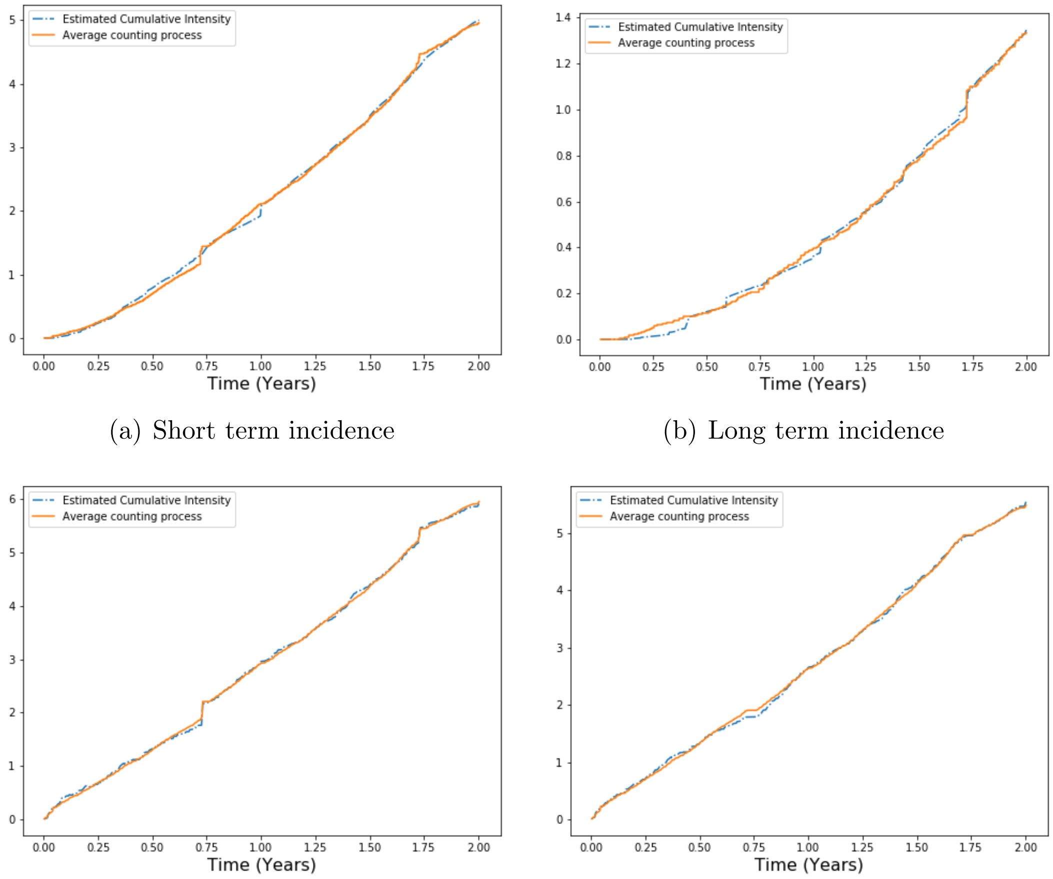


Figure 3: Mon/Friday Sine Function



(a) Short term incidence

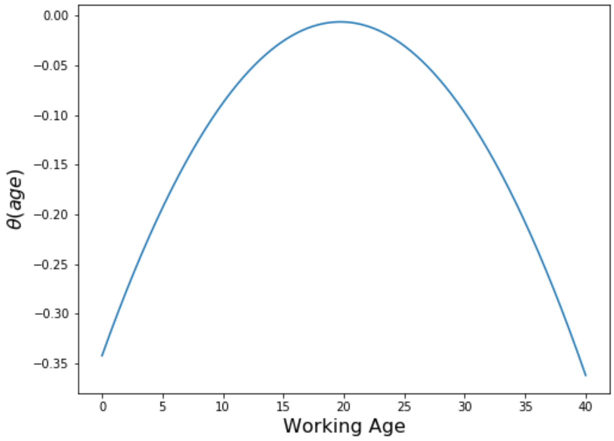
(b) Long term incidence

(c) Short term recovery (Original)

(d) Short term recovery (Bias Reduction)

Figure 4: goodness of fit

Figure 5: $\theta(age)$ in Short term incidence



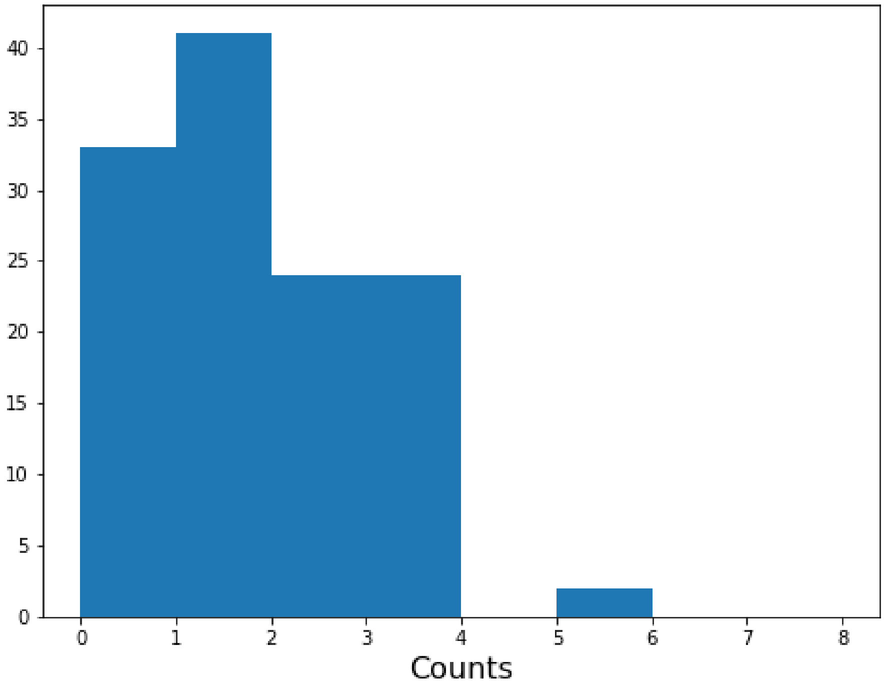


Figure 6: Result of Poisson Process Simulation

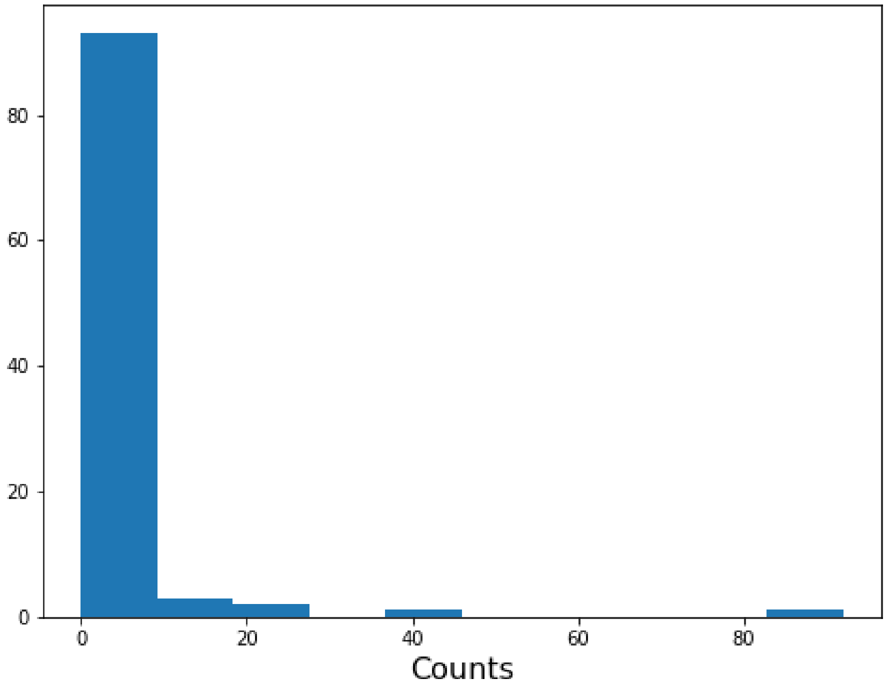
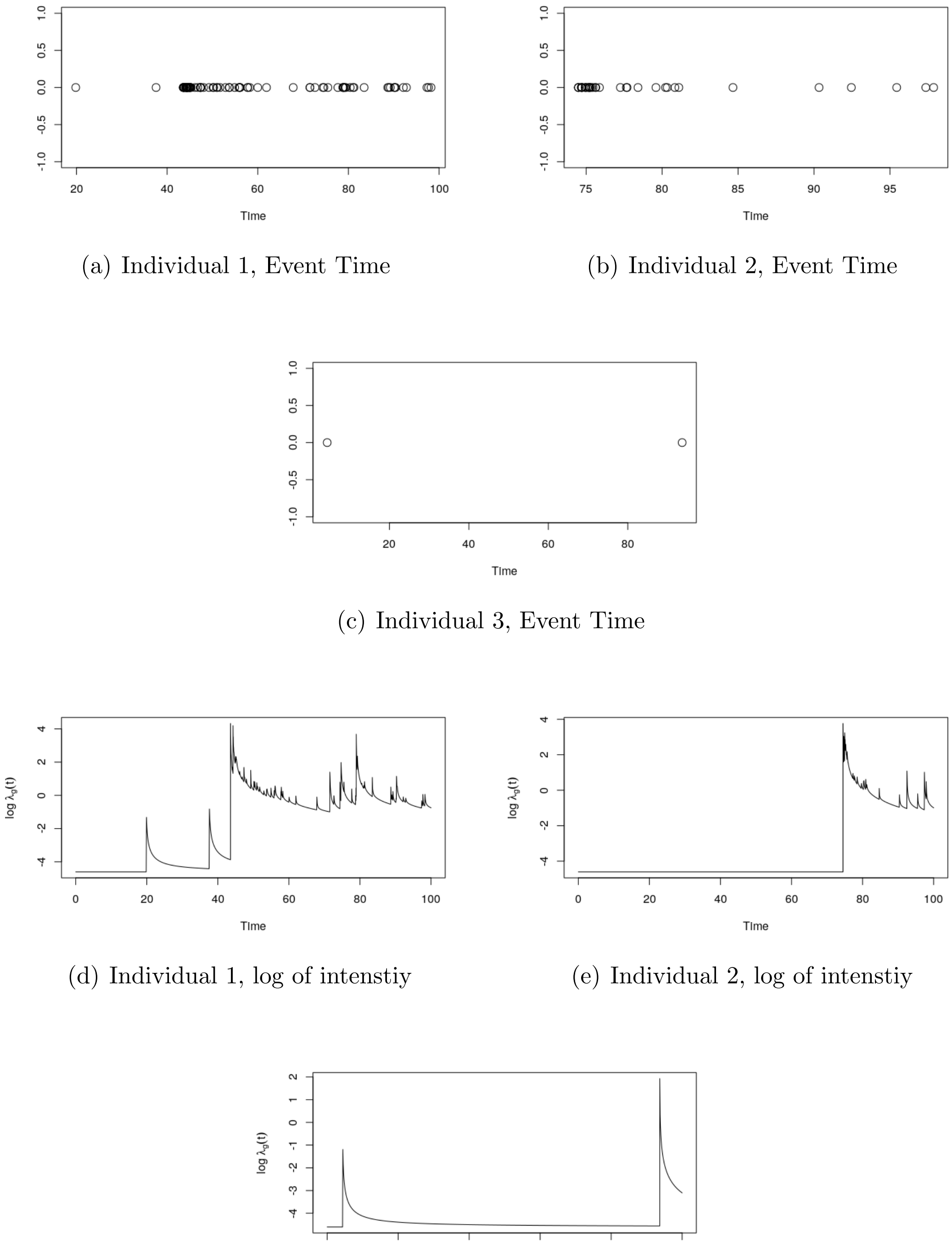


Figure 7: Result of Self-Exciting Process Simulation



(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 8: Three individual events history

Table 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		

Note: $age2 = age^2/100$. *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.

Table 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

Table 3: State Dependence Check

	Dependent variable:	
	$d(i,t)$	
	short term	*long term*
δ	-0.0053926***	-0.00576899***
	(0.000364)	(0.000611)

Note: This table presents the fixed effect panel data regression. Absence duration less or equal to 3 days are categorized as short term, others are long term. The results favor the existence of state-dependent structure among the data. *p<0.1; **p<0.05; ***p<0.01

Table 4: 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

Table 5: 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, see Appendix C for details. $I(t)$ 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

Table 6: Recovery Intensities

	short term,k=2 *original*	*short term,k=2* *holiday*	*long term,k=3* *original*	*long term,k=3* *holiday*
	(1)	(2)	(3)	(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)
c	-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

Table 7: Duration Analysis for Long-term Recovery

	long term,k=3 *holiday*
age	0.07259135*** (0.0128741)
age2	-0.12391056*** (0.0237107)
male	0.04056263 (0.0761176)
full time	-0.09775137 (0.0969011)
married	-0.26665599** (0.0963911)
log-likelihood	3390.528
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

Table 8: 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

Table 9: 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

Table 10: 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 7), 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

A Count Data Regressions and Duration Models

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, $\mu(X_i) = \exp(X_i'\boldsymbol{\beta})$.

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 $\lambda_i = \exp(X_i'\boldsymbol{\beta})$ 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,

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

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

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

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:

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

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:

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

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.

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

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

$$\begin{aligned} h(t|X) &= \frac{\int_0^\infty h(X,\nu)S(t|X,\nu)dG(\nu)}{S(t|X)} \\ &= \exp(X\boldsymbol{\beta})\mathbb{E}(\exp(\nu)|T > t, X) \end{aligned} \tag{31}$$

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 and 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 and 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.

B Properties of the Doob-Meyer Decomposition and the Inference Theorems

B.1 Properties of the Decomposition

Uniqueness of the Doob-Meyer Decomposition Suppose we have an alternative decomposition:

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

Conclude that

$$\Lambda(t) - \tilde{\Lambda}(t) = M(t) - \tilde{M}(t)$$

whence

$$\mathbb{E}[\Lambda(t) - \tilde{\Lambda}(t)|\mathcal{F}(t-)] = \mathbb{E}[M(t) - \tilde{M}(t)|\mathcal{F}(t-)]$$

Since $\Lambda(t)$ and $\tilde{\Lambda}(t)$ are measurable w.r.t $\mathcal{F}(t-)$ and $M(t)$ and $\tilde{M}(t)$ are martingales, the last equation yields

$$\Lambda(t) - \tilde{\Lambda}(t) = M(t-) - \tilde{M}(t-) = \Lambda(t-) - \tilde{\Lambda}(t-)$$

By induction, it follows that

$$\Lambda(t-) - \tilde{\Lambda}(t-) = \Lambda(0) - \tilde{\Lambda}(0) = 0 - 0 = 0$$

and, finally, $M(t) = \tilde{M}(t)$.

Proof of $\mathbb{E}N(t) = \mathbb{E}\Lambda(t)$ Before going into the proof, we first introduce the following lemma:

Lemma Let the random variable X_i records the occurrence time of i^{th} event and let $F_i(\cdot)$ be the distribution function of X_i. Denote $\mathbb{I}(\cdot)$ the indicator function. We have

$$\mathbb{E}(N(t)|\mathcal{F}(t-)) = N(t-) + \mathbb{I}(X_{N(t-)+1} > t-)\frac{F_{N(t-)+1}(dt)}{1 - F_{N(t-)+1}(t-)}$$

Proof. Note

$$N(t) = N(t-) + \mathbb{I}(t- < X_{N(t-)+1} \leq t)$$

The first component is measurable w.r.t $\mathcal{F}(t-)$. For the second component we get

$$\begin{aligned}
&\mathbb{E}[\mathbb{I}(t- < X_{N(t-)+1} \leq t)|\mathcal{F}(t-)] \\
&= \frac{\mathbb{I}(t- < X_{N(t-)+1}) \int_{\{t- < X_{N(t-)+1}\}} \mathbb{I}(t- < X_{N(t-)+1} \leq t) dF_{N(t-)+1}(x)}{1 - F_{N(t-)+1}(t-)} \\
&= \mathbb{I}(X_{N(t-)+1} > t-)\frac{F_{N(t-)+1}(dt)}{1 - F_{N(t-)+1}(t-)}
\end{aligned}$$

□

Now we prove the main claim.

Proof. Let $N(0) = \Lambda(0) = M(0) = 0$. By the Doob-Meyer decomposition, the proof is completed if $\mathbb{E}M(t) = 0, \forall t > 0$. For $\forall t > 0$, set, by recursion,

$$\Lambda(t) = \Lambda(t-) - N(t-) + \mathbb{E}(N(t)|\mathcal{F}(t-))$$

$$M(t) = M(t-) + N(t) - \mathbb{E}(N(t)|\mathcal{F}(t-))$$

From the above lemma, the martingale part of the counting process then satisfies the recursion

$$M(t) = M(t-) + \mathbb{I}(t- < X_{N(t-)+1} \leq t) - \mathbb{I}(X_{N(t-)+1} > t-)\frac{F_{N(t-)+1}(dt)}{1 - F_{N(t-)+1}(t-)}$$

$$M(dt) = \mathbb{I}(t- < X_{N(t-)+1} \leq t) - \mathbb{I}(X_{N(t-)+1} > t-)\frac{F_{N(t-)+1}(dt)}{1 - F_{N(t-)+1}(t-)}$$

Taking integral on both side w.r.t a Lebesgue measure leads to

$$M(t) = \mathbb{I}(X_{N(t-)+1} \leq t) - \int_0^t \frac{\mathbb{I}(x \leq X_{N(t-)+1})}{1 - F_{N(t-)+1}(x-)} F_{N(t-)+1}(dx)$$

The proof is completed by taking expectation on both sides on the last equation. □

B.2 Inference Theorems

The following theorems come from Kopperschmidt and Stute (2013). let $v_0 \in \Theta \subset \mathbb{R}^d$ be the true parameters, and let $\Lambda_{v,i}$ with $v \in \Theta \subset \mathbb{R}^d$ be a given class of parametric cumulative intensities.

Theorem (Consistency) Let $\Theta \in \mathbb{R}^d$ be a bounded open set and for each $\epsilon > 0$, we assume,

$$\inf_{||v-v_0|| \geq \epsilon} ||\mathbb{E}\Lambda_{v_0} - \mathbb{E}\Lambda_v||_{\mathbb{E}\Lambda_{v_0}} > 0 \tag{32}$$

$$\textit{The process}(t, v) \to \Lambda_v(t) \textit{ is continuous with probability one} \tag{33}$$

Then

$$\lim_{n\to\infty} v_n = v_0 \textit{ with probability one} \tag{34}$$

Condition 32 is a weak identifiability condition, while condition 33 guarantees continuity (but not differentiability) of Λ_v in t and allows for unexpected jumps in the intensity function λ_v as well.

Theorem (Asymptotic Behaviour) Let

$$\Phi_0(v) = \frac{\partial}{\partial v}\int_E (\mathbb{E}\Lambda_{g,v}(t) - \mathbb{E}\Lambda_{g,v_0}(t))\mathbb{E}\frac{\partial}{\partial v}\Lambda_{g,v}(t)^T \mathbb{E}\Lambda_{g,v_0}(dt) \tag{35}$$

a matrix-valued function, where T denotes transposition, $E = [\underline{t}, \bar{t}]$. And suppose (32) and (33) hold, furthermore, assume that

$$\parallel \frac{\partial}{\partial v}(\mathbb{E}\Lambda_{g,v}(t) - \mathbb{E}\Lambda_{g,v_0}(t)\mathbb{E}\frac{\partial}{\partial v}\Lambda_{g,v}(t)^T) \parallel \leq C(t) \tag{36}$$

for all v in a neighborhood of v_0, function C is integrable w.r.t $\mathbb{E}\Lambda_{v_0}$, and

$$\phi(x) = \int_{[x,\bar{t}]} \mathbb{E}\frac{\partial}{\partial v}\Lambda_{g,v}(t)\mathbb{E}\Lambda_{g,v_0}(dt) \mid_{v=v_0}, \underline{t} \leq x \leq \bar{t} \tag{37}$$

is square integrable w.r.t. $\mathbb{E}\Lambda_{v_0}$. Then as $n \to \infty$

$$\sqrt{n}\Phi_0(v_0)(v_n - v_0) \to \mathscr{N}_d(0, C(v_0)) \tag{38}$$

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

$$C_{ij}(v_0) = \int_E \phi_i(x)\phi_j(x)\mathbb{E}\Lambda_{g,v_0}(dx) \tag{39}$$

Remark Let Φ_n be the empirical analogue of Φ_0,

$$\Phi_n(v) = \frac{\partial}{\partial v}\int_E (\bar{\Lambda}_{v,n}(t) - \bar{\Lambda}_{v_0,n}(t))\frac{\partial}{\partial v}\bar{\Lambda}_{v,n}(t)^T \bar{\Lambda}_{v_0,n}(dt) \tag{40}$$

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

$$\Phi_n(v) = \Phi_0(v) + op(1) \tag{41}$$

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 $\mathscr{N}_d(0, C(v_0))$, where C^n is the sample analog of C. In practice, one need to plug and replace the true ones with estimators and replace $\mathbb{E}\Lambda_{v_0}(dt)$ with its empirical counterpart $\bar{N}(dt)$.

C 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 11: 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: *p<0.1; **p<0.05; ***p<0.01

D Simulation Procedure

Here we describe the simulation procedures of the Poisson and the ETAS processes.

D.1 Simulation of Poisson Process

We use the fact that the inter-event durations $d_i = t_i - t_{i-1}, i = 2, 3, \cdots$ of a Poisson process are exponentially distributed:

$$F(d) = 1 - \exp(-\lambda d)$$

Procedure

1. Draw a uniformly distributed random variable $U \in [0, 1]$, the duration until next event is then generated by:

$$nextTime = \frac{-\log(1 - U)}{\lambda}$$

2. Update the time list by:

$$\text{TimeList} = \text{append(TimeList,TimeList+nextTime)}$$

3. if the latest time stamps is below the terminated time $TimeList \leq tMax$, repeat steps 1 and 2, otherwise, terminate the program.

D.2 Simulation of ETAS Process

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

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

   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

E Optimization Procedure

Due to the highly non-linear nature of our intensity function (hence the distance function), there is no obvious closed-form solution. A set of numerical optimization routines are employed.

It is well known that for non-linear optimization, the starting points (or guess points) are important. Different starting points may lead to different optimization results. To minimize such impacts, we first use heuristic algorithms such as simulated annealing to find a set of ‘proper’ starting points. These algorithms, however, are usually not speedy enough, we thus break the process after 24 hours.

Using these 'proper' starting points, we next perform the usual optimization routine such as BFGS ,L-BFGS-B, Nelder-Mead and other Newton-based methods. We fail to find a single optimization routine that can ‘fit’ all the distance functions. For example, in the short-term incidence, although the BFGS routine can find the smallest distance value, the estimator for the absence score is positive (and significant), i.e., the higher the absence score is, the more likely to ask for leaves. This result is obviously wrong. Instead, we perform a bounded optimization using the L-BFGS-B routine. The new distance value is larger, but difference is quite small.

We write the code in Python with several packages installed, among them, the most important ones are numpy and numba.