# Strategic Behaviour in Work Absence: A Dynamic View

Yuhao Li

ENTER Seminar
Universität Mannheim

**uc3m** | Universidad **Carlos III** de Madrid

## Motivation

Work absenteeism is not uncommon and is costly for both workers and companies

- Approximately 3.3% of the U.S workforce do not report to work at any given day (Bureau of Labour Statistics data (2005)). Even higher in some developing countries (Duflo et al. (2012)).

- Absence replacement rates are usually less than 100%.

- Absenteeism costs account as high as 14.3% total payroll (Fister-Gale (2003)).

Worker's absence behaviours may be strategic:

- Dynamic: Preferences or constraints to future choices are altered by past experiences; State-dependent.

- Strategic: Decisions based on individual traits and history info; Not entirely random.

- Source of dynamic and strategic behaviour: Company's absence regulation, absence scores $\rightarrow$ benefits.

**Questions**

- Do individuals respond to the absence scores? $\longrightarrow$ Estimate a 'deep' parameter.
- Do individuals respond differently towards different absence events? Short-term, long-term; ask for leaves, return to works
- How individuals make absence decisions? $\longrightarrow$ Economic Models

**In the Literature**

- History information (e.g, absence scores) is not considered in most empirical models. Fail to address the strategic absence behaviours: Delgado and Kniesner (1997), Barmby et al.(1991), Markussen et al.(2011) and Fevang et al.(2014).
- Use conventional tools: count data regressions, duration analysis, etc that hard to model state-dependent structure.

# Motivation

**In this paper**

- Use self-exciting processes to model events (ask for leaves, return to work).

    - Self-exciting process is state-dependent
    - include past experiences

- Build econometric models that are depended on history: take absence scores into consideration.

- Main results:

    - Short-term absence events: respond to the absence scores
    - Long-term absence events: insensitive

- Inspired by the empirical results, construct a simple economic model.

- Compare new method with conventional count data regression and duration models.

# Outline

# Data

- A firm-level administrative data. The firm is UK based, produces homogeneous product.

- Has an experience rated work absence benefit scheme: lower absence scores correspond to a better financial benefit; renew every two years.

- Workers are categorised (with decreasing order) as class A, B and C based on previous two years' absence scores. The criteria scores are 21 and 41 respectively.

- All workers can receive UK statutory sick pay (SSP). To be eligible to receive SSP, workers should be absent from work for more than 3 consecutive days.

- Use data from calendar year 1987 to 1988. In total, 878 workers with 5718 absence records.

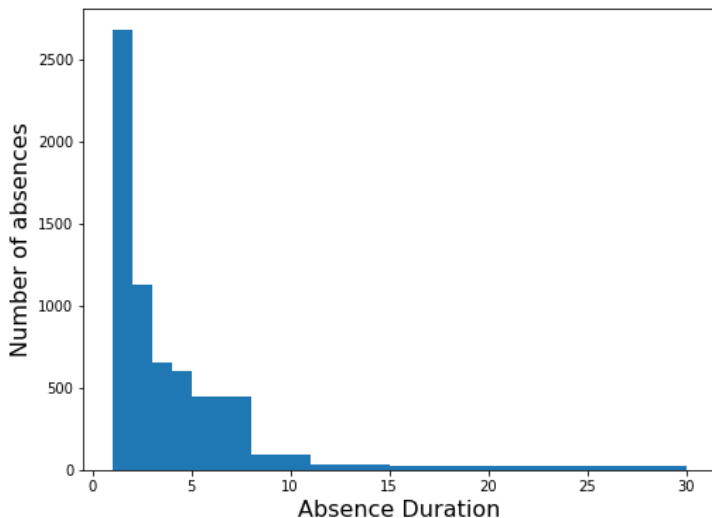- Other literatures use this data: Barmby et al. (1991,1995), etc.

FIGURE: Most frequent absence durations

# PRELIMINARY RESULTS
## CONVENTIONAL METHODS

- Count Data Regressions

    1. Poisson
    2. Negative Binomial
    3. Hurdle Model (Zero part: binomial, count part: truncated negative binomial )
    4. Zero-inflation Model (Zero part: binomial, count part:negative binomial)

- Duration Models: Study the duration before the initial absence

    1. Standard: the hazard function is constant, no individual heterogeneity is concerned.
    2. Include individual heterogeneity, use Heckman and Singer's NPMLE.

# PRELIMINARY RESULTS

Count Data Regressions

|  | *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) |
| sex | −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 |
| $\theta$ |  | 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

# PRELIMINARY RESULTS

## THE DURATION MODEL

For standard duration model, the hazard rate for individual $i$:

$$h_i = \exp(X_i'\beta)$$

$$H_i(T) = \int_0^T h_i(t)dt$$

$X_i$ is a vector of covariates. The likelihood function:

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

$y_i$ is the censoring indicator.

Include individual heterogeneity, the hazard rate:

$$h_i(t) = \exp(X_i'\beta + \nu_i)$$

Use Heckman and Singer's NPMLE: Approximate the distribution of $\nu_i$ using discrete mass points. Independent Assumption violated!

# PRELIMINARY RESULTS

## DURATION MODEL 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) |
| sex | -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 |
| $\chi^2$ | 576.961*** (df = 5) | |
| Number of Mass Points |  | 2 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

# NATURE OF THE PROBLEM

Conventional methods can not model the strategic absence behaviours:

- Count Data Regressions aggregate information.
- Duration models have to maintain the i.i.d assumption among absences.
    - Multiple-spell models that allow lagged duration dependence
    - Estimate the joint duration likelihood function: must be a panel data setting.

We need to model the **State Dependent Structure** among absences.

$$
\overset{\phantom{x}}{\underset{0 \quad\quad s_1 \quad r_1 \quad\quad s_2 \quad\quad r_2 \quad S \quad\quad\quad\quad\quad\quad\quad\quad R}{\rule{10cm}{1pt}}}
$$
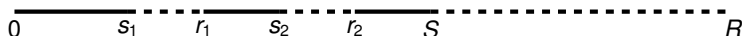
$s$, $S$ are starting dates and $r$, $R$ are ending dates for short and long-term absences.

- State Dependence means past experiences have effects on future ones.
- Larger absence scores should discourage further absence behaviours.

# NATURE OF THE PROBLEM
### STATE DEPENDENCE TEST

Following Heckman (1981), construct a weekly panel data to test the state dependent structure. For each individual $i$, write the regression:

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

$d(i, t)$ is dichotomous choice, $\mathbb{E}(d(i, t) = 1) = \Pr(\text{ask for a leave})$. $\nu_i$ is individual specific term, $U(i, t)$ satisfy usual fixed effect model assumptions.

$\hat{\delta} = -0.033316$ with a standard deviation of 0.00033: favours the existence of state-dependence in incidence data.

# Outline

# INTRO TO THE SELF-EXCITING PROCESS

The self-exciting process is a special counting process.

Counting process:

$$N(t) = \sum_i \mathbb{I}(t_i \leq t)$$

**Doob-Meyer Decomposition**:

$$N(dt) = \Lambda(dt) + M(dt)$$

$\Lambda(t)$: predictable compensator, $M(t)$: local martingale.

The compensator may be conditioned on a filtration $\mathcal{F}(t)$. If $\mathcal{F}(t)$ includes $\sigma(N(s) : s \leq t)$, we call the associated counting process a **self-exciting process**. The self-exciting process is by design **state dependent**.

Study $N(t)$ via $\Lambda(t)$ or the intensity $\lambda(t) = \Lambda(dt)/dt$

# THE ECONOMETRIC MODELS
SETUP (1/2)

- 'decompose' an absence into two events:
    - Incidence event: ask for a leave
    - Recovery event: return to work
- distinguish short and long-term absence:
    - use the duration of absences as criteria
    - the cut-off is 3 days: based on the UK's SSP regulation.
- have four counting processes.
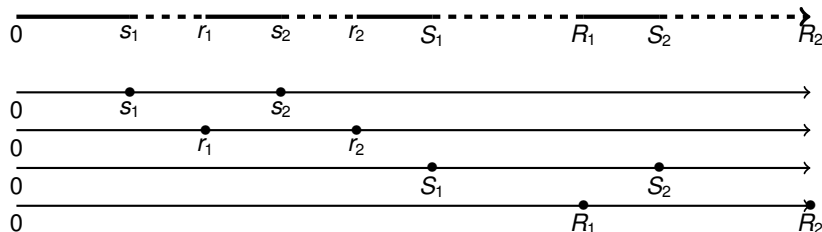
# THE ECONOMETRIC MODELS
## SETUP (2/2)

Define 3 states that one individual can occupy:

- attendance: $k = 1$

- short-term absence: $k = 2$

- long-term absence: $k = 3$

$\lambda_{1k}(t)$: Short (Long) term incidence intensity, $k = 2, 3$;

$\lambda_{k1}(t)$: Short (Long) term recovery intensity, $k = 2, 3$;

The period between previous return date and the start date of next absence is called
**attendance period**; the period between the start date and return date of same
absence is called the **absence period**.

Let $t$ denotes the starting dates and $\tau$ denotes the return dates. The duration of $i^{th}$
absence is then $d_i = \tau_i - t_i$.

# THE ECONOMETRIC MODELS
INCIDENCE INTENSITY (1/2)

If individual *i* has no previous absence records, instead of intensity, we will study her hazard rate for the duration until first absence:

$$h_i = \exp(\nu_i) exp(\mathbf{X}_i' \gamma_{1k})$$

where $\nu_i$ is individual's random effect, $\mathbf{X}_i$ is a vector of covariates.

Two kinds of individuals fit into this situation:

- Never have absence before and not have absence in the study period (1987-1988): Censoring
- Never have absence before but have absence records in the study period.

Procedures:

- Construct a sub-dataset, using only initial absences
- Study the duration before the initial absence
- The hazard rate has the same interpretation as the intensity for the first event.
- Use Heckman and Singer's NPMLE to approximate the distribution of random effect term.

# THE ECONOMETRIC MODELS
INCIDENCE INTENSITY (2/2)

The overall incidence intensity for individual $i$, who has previous absence records would be:

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

where

- $\lambda_{1,k}(\mathbf{X}_i) = exp(\mathbf{X}_i'\gamma_{1k})$; $k = 2, 3$ contains all the time-invariant covariates

- $\lambda_{2,k}(t) = exp(\beta_{1k}H_i(t))$; $k = 2, 3$ governs the response of one worker to her own cumulative absence time $H_i(t)$.

- $\lambda_{3,k}(t) = 1 + |\alpha_{1k}|exp(\alpha_{1k}(t - \tau_{N_i^1(t-)}))$; $k = 2, 3$ measures the time dependence since previous recovery date.

- $\lambda_{4,k}(t) = a_{1k}(1 + sin(b + c_{1k}t))$; $k = 2, 3$ measures the individual's response to Mondays and Fridays. 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.

# THE ECONOMETRIC MODELS

RECOVERY INTENSITY

The intensity for recovery process (return to work) has three parts:

$$\lambda_{5,k}(\boldsymbol{X}_i) = exp(\boldsymbol{X}_i'\gamma_{k1})$$

$$\lambda_{6,k}(\tau) = |\beta_{k1}|exp(\beta_{k1}H_i(\tau))$$

$$\lambda_7(\tau) = 1 + |\beta_{k2}|exp(\beta_{k2}(\tau - t_{N_{i,13}^1(\tau-)})); k = 2, 3$$

The overall intensity for recovery process :

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

# THE ECONOMETRIC MODELS
HETEROGENEITY (1/4)

We try to use history information to approximate the unobserved heterogeneity:

- In incidence process, the primary unobserved heterogeneity is the individual's working attitude.

- It can be approximated by an individual's average attendance duration.

- In recovery process, the primary unobserved heterogeneity is the individual's recovery ability.

- It can be approximated by an individual's average recovery time.

# THE ECONOMETRIC MODELS
HETEROGENEITY (2/4)

The (moving) average attendance duration $\tilde{d}(t)$ is defined as:

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

We assume it has the following structure:

$$\log(\tilde{d}(t)) = I(t) + log(1 + H(t)) + \epsilon$$

with $\mathbb{E}(\epsilon) = 0$.

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

$$\tilde{I}(t) = \log(\tilde{d}(t)) - log(1 + H(t)) = I(t) + \epsilon$$

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

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

the structural of the overall incidence intensity remains the same.

# THE ECONOMETRIC MODELS
HETEROGENEITY (3/4)

Similarly, the (moving) average recovery time $\tilde{c}(t)$ is defined as:

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

and it has the following structure:

$$\log(\tilde{d}(t)) = R(t) - log(1 + H(t)) + \epsilon$$

with $\mathbb{E}(\epsilon) = 0$.

We approximate $R(t)$ by $\tilde{R}(t)$:

$$\tilde{R}(t) = \log(\tilde{c}(t)) - log(1 + H(t)) = R(t) + \epsilon$$

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

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

# THE ECONOMETRIC MODELS
HETEROGENEITY (4/4)

Alternatively, instead of assuming individual heterogeneity, we may assume group heterogeneity and use finite mixture Poisson model to 'reveal' the group affiliation.

Intuition: hard-working individuals tend to (on average) have fewer absences.

Assume $k$ groups, the numbers of absences $\mathbf{y} = (y_1, \cdots, y_N)'$ follows:

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

$\Theta_k$: vector of parameters for group $k$, $w_k$: weight and $f_k(\cdot|\Theta_k)$: Poisson density.

Calculate the posteriors and assign group affiliation:

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

# How to Estimate the Models

One result of the Doob-Meyer Decomposition:

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

Kopperschmidt and Stute (2013) propose a minimum distance estimation method to overcome the mentioned difficulty.

$$\nu_n = arg \inf_{\nu \in \Theta} ||\bar{N}_K - \bar{\Lambda}_{\nu,K}||_{\bar{N}_K}$$

where $||f||_\mu = \left[ \int f^2 d\mu \right]^{1/2}$, with $\mu$ being a measure.

$\bar{N}_K = \frac{1}{K} \sum\limits_{i=1}^{K} N_i$, $\bar{\Lambda}_{\nu,K} = \frac{1}{K} \sum\limits_{i=1}^{K} \Lambda_{\nu,i}$.are the averaged counting process and compensator respectively.

The estimator is found to be consistent and asymptotically normal as long as all the observations are i.i.d.

# How to Estimate the Models

The usual way to estimate a self-exciting intensity is the maximum likelihood based technique:

- require the intensity be predictable with respect to its filtration.
- often fails in complicated economic models as the existence of external shocks.
- the intensity should be predictable with respect to the self-exciting filtration as well as shocks.

In our application:

- The four counting processes are external shocks to each other
- The switches between attendance period and absence period are external shocks.

# Outline

# THE RESULTS

INCIDENCE PROCESS (1/3)

No previous absence records: study the duration before initial absence. Those who have no absence records throughout the whole study period are treated as censoring. Usual censoring assumptions applied.

|  | short term,k=2 (1) | short term,k=2 (2) | long term,k=3 (3) | long term,k=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 | 0.3174 | 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: Note: Absence duration less or equal to 3 days are categorized as short term, others are long term. *p<0.1; **p<0.05; ***p<0.01

# THE RESULTS
## INCIDENCE PROCESS (2/3)

Two heterogeneity settings for incidence processes: approximation & group

- Finite Mixture Poisson-2 Model

- 2: limited heterogeneity & simplicity

- Posteriors suggest only one component in Long-term incidences.

Incidence intensity estimation results, both short and long-term absences:

| | Approx. Heterogeneity | | Group Heterogeneity | |
| | short term | long term | short term | long term |
| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $\beta_{1k}$ | $-0.05734195^{***}$ | $0.002551$ | $-0.03207249^{***}$ | $-0.02183574$ |
| | $(0.007313)$ | $(0.0108902)$ | $(0.0072219)$ | $(0.0139920)$ |
| $\alpha_{1k}$ | $-35.32423495^{**}$ | $-5.02947189$ | $-36.90188525^{*}$ | $-4.92911781$ |
| | $(17.09965)$ | $(6.4427670)$ | $(21.600208)$ | $(7.2625815)$ |
| 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 | $-1.03290089^{***}$ | $1.33187185^{*}$ | $-1.50756787^{**}$ | $1.32788344^{*}$ |
| | $(0.350615)$ | $(0.7269884)$ | $(0.6975168)$ | $(0.6999908)$ |
| Mon/Fri | $2.01429447^{*}$ | $0.15542535$ | $5.00469984^{*}$ | $0.1711720$ |
| | $(1.142056)$ | $(2.6100053)$ | $(2.6209058)$ | $(2.9969653)$ |
| b | $2.57708555^{***}$ | $2.69547261$ | $2.58967502^{***}$ | $2.71009241$ |
| | $(0.547747)$ | $(7.8254215)$ | $(0.2458489)$ | $(8.4098623)$ |
| Group 2 | -- | -- | $1.01837705^{**}$ | -- |
| | -- | -- | $(0.4682334)$ | -- |
| $l(t)$ | $-0.42075577^{***}$ | $0.03062177$ | -- | -- |
| | $(0.095787)$ | $(0.1093162)$ | -- | -- |
| Distance | $0.128602$ | $0.028854$ | $0.146190$ | $0.0295902$ |

Note: Note: Absence duration less or equal to 3 days are categorized as short term, others are long term. $\beta_{1k}$ are the coefficients of the absence score, $\alpha_{1k}$ are the coefficients of time dependent structure. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Special attention to

- Scheduled absences: e.g., absence related to holidays, information set is complete when making the absence decision; when asking for leave, the duration of that absence has already been known.

- The estimation results are biased; delete absences during the Christmas season to reduce the bias.

Recovery intensities estimation results.

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

Note: Absence duration less or equal to 3 days are categorized as short term, others are long term. $age2 = age^2/100$. $\beta_{k1}$ are the coefficients of cumulative absence time, $\beta_{k2}$ are the coefficients of time dependent structure. The first and third columns are results for all the absence records, the second and fourth columns are the results after we delete the absences during the Christmas seasons. *p<0.1; **p<0.05; ***p<0.01

Long-term recoveries are i.i.d: Standard duration model, study the duration of being absence.

For each individual $i$, the hazard rate is:

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

$$H_i(T) = \int_0^T h_i(t)dt$$

the likelihood contribution is:

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

$S_i$: the set of observed long term durations for individual $i$.

Use Heckman and Singer's NPMLE to estimate

Duration analysis for long term absences

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

Note: $age2 = age^2/100$. $b$ is the coefficient of duration dependence. *p<0.1; **p<0.05; ***p<0.01

Overall our models fit the data well.
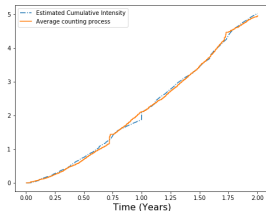


(a) Short term incidence
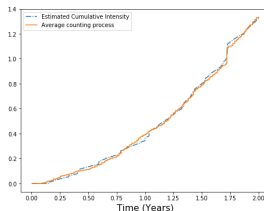


(b) Long term incidence
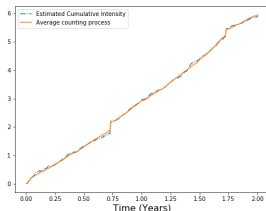


(c) Short term recovery (Original)
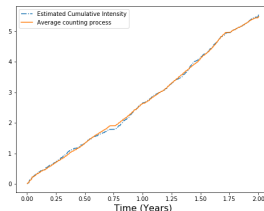


(d) Short term recovery (Bias Reduction)

**Question:** does individual's attitude towards the cumulative absence time changes as seniority grows?

Modify:

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

where $\theta(age) = \beta_0 + \beta_1 age + \beta_2 age^2/100$

Other components and the structural of both incidence and recovery intensities remain unchanged.

# The Results

Deeper into the Strategic Behaviour Effect (2/4)

|  | long term,k=3 |
|---|---|
|  | holiday |
| $\beta_0$ | -0.34215516* |
|  | (0.1904758) |
| $\beta_1$ | 0.03407675 |
|  | (0.0280976) |
| $\beta_2$ | -0.08642652 |
|  | ( 0.0960550) |
| $\alpha_{1k}$ | -39.37455248** |
|  | ( 18.741119) |
| $b$ | 2.61810899*** |
|  | ( 0.3692254) |
| Mon/Fri | 2.35010473** |
|  | (1.1141944) |
| $I(t)$ | 0.28212665*** |
|  | (0.0913577) |
| Demographic Variables | YES |
| Distance | 0.107879 |

Note: Absence duration less or equal to 3 days are categorized as short term, others are long term. $age2 = age^2/100$. We delete absences during the Christmas seasons. *p<0.1; **p<0.05; ***p<0.01

Wald Test:

- Test 1: do individuals respond to the cumulative absence time

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

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

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

Statistics for the Wald tests are 22334.637 and 355.045 respectively.

FIGURE: $\theta(age)$ in Short term incidence
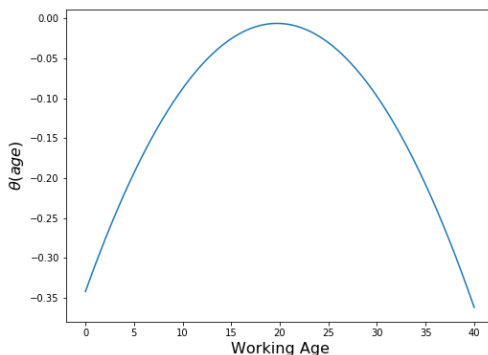
# THE RESULTS

The Cut-off between Short and Long-terms (1/2)

- So far, we select the cut-off based on the UK's SSP Regulation: three days of absence

- Redefine the short and long-term absence by individual's response to the cumulative absence time:
  - short-term: individuals respond to the cumulative absence time.
  - long-term: no responses to the cumulative absence time.

- The coefficients of cumulative absence time satisfy:
  - the short term coefficient $\beta_{12}$ and $\beta_{21}$ are significant away from zero
  - the long term coefficient $\beta_{13}$ and $\beta_{31}$ are insignificant

# THE RESULTS
## THE CUT-OFF BETWEEN SHORT AND LONG-TERMS (2/2)

| | Incidence Intensities | | | Recovery Intensities | |
|---|---|---|---|---|---|
| | short term | | long term | short term | long term |
| Cut-off | Mon/Fri | $\beta_{12}$ | $\beta_{13}$ | $\beta_{21}$ | $\beta_{31}$ |
| $c = 2$ | 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: *$p<0.1$; **$p<0.05$; ***$p<0.01$

- The proper cut-off under new definition is still three days
- Importance of Social Security Regulation!

# OUTLINE

# ECONOMIC MODELS
## ASK FOR LEAVES (1/2)

**Ask for Leaves**

- Accident, represented by a random variable $e \in [0, \infty)$
- The size of an accident is not revealed at the first place.
- Duration of an absence is determined by the accident: $a(e)$
- Individual Utility is determined by personal reputation, well-beings, and consumption

The decision to ask for a leave is governed by:

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

$R(\cdot)$: reputation, determined by the cumulative absence time $A$, in addition, $R'(\cdot) < 0, R''(\cdot) < 0$; $\omega$: a stock of well-beings, $g(\cdot)$: well-being generating function with $g'(\cdot) > 0$; $C$: consumption. $\epsilon$:unobserved factors.
Utility function satisfies: $U_R, U_\omega, U_C > 0$, partial derivatives are positive.

Notice:

$$Pr(D = 1) = Pr(\epsilon_2 - \epsilon_1 < U^1 - U^0)$$
$$= F_\epsilon(U^1 - U^0)$$

where $\epsilon = \epsilon_2 - \epsilon_1$.

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

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

# ECONOMIC MODELS
## RETURN TO WORK: SHORT-TERM (1/2)

**Return to Work**:

- Strategic behaviours only exist in the short-term recovery processes
- A threshold of the size of an accident: $e^*$ (with its corresponding absence duration $a(e^*)$)
- Within this threshold, individuals consider reputation; above this threshold, reputation is out of equation.

If $e \in [0, e^*]$ (short-term absences), the worker's problem is

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

$$s.t$$

$$I + w(t^c - a) + R(A + a) - C = 0$$

$I$: non-labour income, $w$: wage, $t^c$ contractual working time.

First Order Conditions:

$$U_R R^{'} + U_\omega g^{'} - U_C(w - R^{'}) = 0$$

$$(w - R^{'}) = \frac{U_R R^{'} + U_\omega g^{'}}{U_C} > 0$$

and

$$\frac{\partial a}{\partial A} < 0, \frac{\partial a}{\partial R} > 0$$

# ECONOMIC MODELS
RETURN TO WORK: LONG-TERM (1/3)

If $e \in [e^*, \infty)$ (long-term absences), for individual $i$ in $j^{th}$ long-term recovery period:

- Has a positive utility flow: $K_{ij}Z_1(t)\phi_1(X_i)$, $K_{ij}$ is a positive random variable, may represent initial health status; $Z_1(\cdot), \phi_1(\cdot)$: functions of time and individual covariates respectively.

- If choose to return to work: receive an utility flow $Z_2(t)\phi_2(X_i)$.

Assume an exponential discount rate $\rho$, worker's problem:

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

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

This model is essentially a simplified version of Honor and De Paula (2010).

# ECONOMIC MODELS
RETURN TO WORK: LONG-TERM (2/3)

First Order Condition:

$$K_{ij}Z_1(t_{ij}^*)\phi_1(X_i)e^{-\rho t_{ij}^*} - Z_2(t_{ij}^*)\phi_2(X_i)e^{-\rho t_{ij}^*} = 0$$
$$K_{ij} - Z(t_{ij}^*)\phi(X_i) = 0$$
$$t_{ij}^* = Z^{-1}(K_{ij}/\phi(X_i))$$

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

Notice:

$$\ln Z(t_{ij}^*) = -\ln\phi(X_i) + \epsilon$$

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

- Assume $Z(t) = t$, $\phi(X_i) = e^{-X_i^T\beta}$ and $K_{ij} \sim \exp(1)$
- Is the accelerated failure time (AFT) model
- The corresponding hazard rate: $h_{t_{ij}^*}(t) = \exp(-X_i^T\beta)$

We can estimate this model:

|  | *Dependent variable:* |
| --- | --- |
|  | 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 |
| $\chi^2$ | −362.47690 (df = 4) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

# Outline

# DISCUSSION

- Poisson is the basic model of count data regression
    - require the equidispersion: the mean and variance should be equal
    - can not fit excess-zeros data
- Usual way to deal with the excess-zeros data: zeros and positives come from two data generating processes
    - Hurdle model
    - Zero-inflation model
- Simulation intensity function: $\lambda(t|\mathcal{F}_{t-}) = \mu + \sum_{i:t_i<t} e^{\alpha x_i} \left(1 + \frac{t-t_i}{c}\right)^{-p}$
- Generate data that
    - exhibits over-dispersion and
    - excess of zeros

# DISCUSSION
COMPARE TO THE COUNT DATA REGRESSION (2/2)

- mean: 3.27
- variance: 108.54
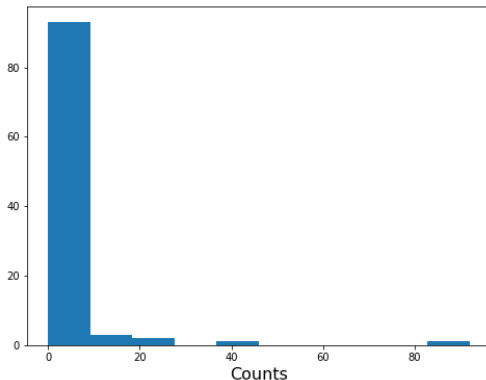


FIGURE: Result of Self-Exciting Process Simulation

# Discussion
## Compare to Duration Model (1/2)

- Duration model deals with the duration of a single event

- Self-exciting processes deals with recurrent events

- When describes a single event, the hazard rate and the intensity have the same interpretation
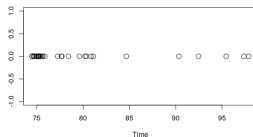
- Self-exciting may generate enough heterogeneity even without the unobserved heterogeneity.

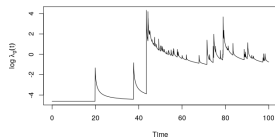# DISCUSSION

Using the same data generating process as before:
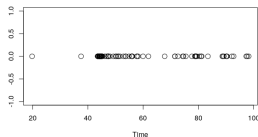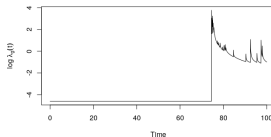


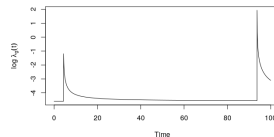(a) Individual 1, Event Time

(b) Individual 2, Event Time
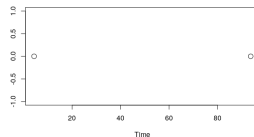
(c) Individual 3, Event Time

(d) Individual 1, log of intenstiy

(e) Individual 2, log of intenstiy

(f) Individual 3, log of intenstiy

# Outline

## CONCLUSION

In this paper, we

- provide substantial evidence on the existence of strategic absence behaviour

  - this is done by using the self-exciting process
  - study both ask for leaves and return to work decisions
  - distinguish short and long-term absences

- Only in the short-term events, individual cares about the cumulative absence time

  - can be used as one criteria to separate short and long-term absences

- build economic models to explain the empirical findings

- discuss self-exciting process as a complementary tool to conventional methods

# Thank You !