ISSN 0143-4543

Centre for European Labour Market Research

Worker Absenteeism: A Study of Contagion effects

By

Tim Barmby and Makram Larguem

Discussion Paper 2004-01
February 2004

Editor: Dr W David McCausland
www.abdn.ac.uk/business/

Worker Absenteeism: A Study of Contagion effects

Tim Barmby[1] (Centre for European Labour Market Research, University of Aberdeen)

Makram Larguem (ERMES, CNRS, Université de Paris 2, Panthéon-Assas)

February 2004

1. Introduction

A number of recent studies have suggested that workers' attendance as well as their absence, could have importance for the way in which firms' design remuneration contracts, see Chatterji and Tilley (2000) and Skåtun (2002). One aspect of this is that, since worker absenteeism is in large part due to illness, if contracts impose costs on workers which induce them to attend work when ill this could result in the illness being more readily communicated to other workers with associated effects on productivity. This paper seeks to quantify such *contagion* effects by examining a personnel dataset which allows us to track daily absence decisions of a group of industrial workers employed in the same factory.

2. Model of absence

In this section we will try and incorporate some theoretical ideas from the study of sickness contagion and epidemiology into a model of worker absence, being guided by the framework used by Philipson (2000) and Skåtun (2003).

[1] Corresponding author, Centre for European Labour Market Research, Department of Economics, University of Aberdeen, Old Aberdeen, Scotland AB24 3QY. tim.barmby@abdn.ac.uk

The data we use are event histories consisting of sequences of realisations of indicator random variables

$$d_{it} = \begin{cases} 1 & absent \\ 0 & not\,absent \end{cases} \qquad (1)$$

The basic modelling framework we draw on uses a latent variable model describing the probabilistic realisation of these indicator variables, see Heckman (1981), and Barmby, Orme and Treble (1995). This will form the basis of our model of daily absence decisions of workers.

At the simplest level we consider a latent variable

$$d_{it}^{*} = x_{it}\beta + \varepsilon_{it} \qquad i = 1,......,N \quad t = 1,.....,T \qquad (2)$$

where x_{it} are a (row) vector of (possibly time-varying) variables, and β is a conformable vector of parameters, ε_{it} is a random error term with CDF F(ε_{it}). The probability of observing a worker absent can be written

$$P(d_{it} = 1) = P(d_{it}^{*} > 0) = P(x_{it}\beta + \varepsilon_{it} > 0) = F(x_{it}\beta) \qquad (3)$$

for symmetric CDF's.

The central empirical problem faced by this paper is to specify the above probability in such a way that we can interpret the estimated parameters in terms of a coherent model of sickness contagion. To this end we first consider the event that an absence spell *starts* at time t, this is the event $(d_{i,t} = 1 \mid d_{i,t-1} = 0)$. At t-1 the worker was *exposed* to $N - \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{j,t-1}$ workers who were potentially infected. The proportion of these who are infected is in epidemiological terminology the *prevalence*, which we denote, *s*. The *transmission* rate of the infection,

□ describes the ease with which the sickness can be passed from an infected to a non-infected individual. So the probability of a healthy worker being infected can therefore be written.

$$\theta \propto \mu s(N - \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{j,t-1}) \qquad (4)$$

An estimate of s, the prevalence of sickness, can, we argue, be formed by observing how many workers (apart from i) who were at work at t-1 but were absent at t, and taking this as a ratio to the total number at work at t-1.

$$\tilde{s} = \frac{\sum_{\substack{j=1 \\ j \neq i}}^{N} d_{j,t} \mid d_{j,t-1} = 0}{N - \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{j,t-1}} \qquad (5)$$

this implies that an estimate of the probability of infection can be written

$$\tilde{\theta} \propto \mu \frac{\sum_{\substack{j=1 \\ j \neq i}}^{N} d_{j,t} \mid d_{j,t-1} = 0}{N - \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{j,t-1}} (N - \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{j,t-1}) = \mu \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{j,t} \mid d_{j,t-1} = 0 \qquad (6)$$

Note that to be infected the worker has to be at work, also if the worker is absent then he/she is not exposed to infection. We assume that the probability of absence will be an increasing function of □. These, taken together, suggest the following form for the probability in equation (3)

$$P(d_{it} = 1) = F(\mu(1 - d_{i,t-1}) \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{j,t} \mid d_{j,t-1} = 0) \qquad (7)$$

The term in the regression component here is, of course, only capturing the *contagion* component of the process, there will be other factors which we must take into account. To fully specify the probability function we would also include

1) x_{it} : these are covariates, either time variant or, invariant characteristics of workers

2) $d_{i,t-1}$: this captures the effect of own health state. If the worker was absent last period how does this affect the probability of being absent this period. Heckman (1981) terms this structural dependence

3) $\sum_{s=1}^{t-1}\prod_{k=1}^{s} d_{i,t-1}$: this captures the effect of duration in the state, this can be entered in a as a polynomial to capture potential non-linearities.

Adding these terms gives us and extended latent variable model of the form

$$d_{it}^{*} = \underbrace{\beta' x_{it} + \gamma d_{i,t-1} + \lambda \sum_{s=1}^{t-1}\prod_{k=1}^{s} d_{i,t-1} + \mu(1-d_{i,t-1})\sum_{\substack{j=1 \\ j\neq i}}^{N} d_{j,t} \mid d_{j,t-1} = 0}_{\theta Z} + \sigma u_{i} + \varepsilon_{it}$$

$$i = 1,....,N \quad t = 1,......,T_{i} \quad \theta' = (\beta\ \gamma\ \lambda\ \mu) \tag{8}$$

$$Z' = (x_{it}, d_{i,t-1}, \sum_{s=1}^{t-1}\prod_{k=1}^{s} d_{i,t-1}, (1-d_{i,t-1})\sum_{\substack{j=1 \\ j\neq i}}^{N} d_{j,t} \mid d_{j,t-1} = 0)$$

We include an unobserved term u_i in (8) since the data we are working with is gathered from a firm's personnel records. As such it will primarily include data which it needs for the accurate operation of its payroll, this means that certain data which could affect the probability of absence (whether the household has young children for instance) is not observed. So we will have to take the existence of unobserved heterogeneity seriously. We assume this unobserved term has a N(0,1) density $\phi(u)$, which we can integrate out to form a marginal likelihood,

$$LnL(\theta,\sigma) = \sum_{i=1}^{N} Ln \int_{-\infty}^{\infty} \prod_{t=2}^{T_i} F(\theta Z_{it} + u)^{d_{it}} (1 - F(\theta Z_{it} + u))^{(1-d_{it})} \phi(u) du \tag{9}$$

see Hsiao (2003)

3. Data

Our data is a sample of 957 workers in 1988 drawn from a UK manufacturing firm which produces a homogeneous product using production lines. We have data from the personnel records giving us their absence for days they were contracted to work on. The main aspects of the remuneration contract which we have information on are wage rates, contracted hours and the sick pay for which each worker is eligible. The wage information is in the form of weekly payments recorded in the payroll. A worker's remuneration is made up of a basic pay component, overtime payments and shift premia. Basic pay represents the bulk of overall earnings.

The firm also operates a sick pay scheme which entitles workers to sick pay in excess of a minimum Statutory Sick Pay (SSP) if their past attendance is sufficiently good. It is the operation of this sick pay scheme, which is the main way in which the cost of a day's absence varies across individuals. A worker's attendance record is measured by a points scheme where each day of unacceptable absence attracts at least one point. Workers are eligible for sick pay at one of three levels of generosity, A, B and C, depending on their points total over a rolling two year period.

The results are supportive of the notion that contract effects are important in explaining absence patterns. Normal daily wage has a negative effect when interacted with whether the individual is in sick-pay grade B or C where there is a higher cost of a days absence to the worker, contracted hours have a positive and significant effect when the worker is at work, both of these results are similar to Barmby (2002)[2]. Other results are in line with previous work, women have a higher propensity to be absent. There is positive duration dependence peaking around 3 days

---

[2] Note the method of interacting a regressor with $d_{i,t-1}$ allows the coefficient to vary between states (or work and absence) as discussed in Barmby (1998)

In the framework we have set up there are two main sources of variation in sickness (and therefore possible absence). The first is through own health state, which we represent in terms of the structural dependence on own absence $d_{i,t-1}$ and secondly through our the influence of the estimated prevalence of sickness of other workers, which is measured by our contagion variable $(1-d_{i,t-1})\sum_{\substack{j=1 \\ j\neq i}}^{N} d_{j,t} \mid d_{j,t-1} = 0$, (note this effect can only operate when the worker is at work which is why we premultiply $(1-d_{i,t-1})$. We get significant positive effects from both of these, as intuition would suggest.

Since we have in mind that absence is being generated by some underlying stochastic sickness process, we have to take into account the influence of the weekend. We interact $d_{i,t-1}$ with a dummy for Monday to pick up the possible weakening of the effect of past absence over the weekend. This clearly shows up with a significant negative coefficient on this interaction. However doing this for our contagion variable presents some difficulty in interpreting the coefficient on the interaction. Essentially two things are going on, firstly there is the possible weakening of the contagion effect over the weekend, but in addition to this our estimate of the prevalence on Friday is not as good as we are using the observed numbers of absentees on Monday.

Table 1 : ML Estimation marginal likelihood for absence model (without weekends)

Variable	**Coefficient (se)**
Monday	-5.918 (0.260)
Tuesday	-6.804 (0.273)
Wednesday	-7.188 (0.273)
Thursday	-7.2340 (0.272)
Friday	-7.583 (0.271)
1) Worker Characteristics	
Gender	0.2654 (0.0788)
Married	0.0488 (0.0676)
Contracted hours	0.0130 (0.0057)
Contracted hours*Lag own absence	-0.0541 (0.0071)
Grade B	1.1703 (0.0973)
Grade C	1.5113 (0.0964)
Normal Daily Wage	-0.0003 (0.0005)
Normal Daily wage*(grades B+C)	-0.0048 (0.0007)
2) Structural dependence	
Lag own Absence $d_{i,t-1}$	6.6911 (0.3342)
Lag own Absence*Mon	-1.2471 (0.4163)
3) Duration in state	
Duration	1.9323 (0.1841)
Duration^2	-0.3663 (0.0382)
4) Contagion Effects	
Contagion $(1-d_{i,t-1})\sum_{\substack{j=1 \\ j\neq i}}^{N} d_{j,t} \mid d_{j,t-1}=0$	0.0189 (0.0033)
Contagion * Mon	0.0129 (0.0041)
Unobserved Heterogeneity and endpoint parameters	
□	0.7929 (0.0298)
Log likelihood	-16384.735
N (NT)	957 (244 035)

6. Concluding Remarks

The model developed above shows how an individual workers propensity to absent themselves from work is affected by a range of factors including their personal characteristics and the terms of the workers contract. We find that absence is positively affected by a worker's own absence, which is in line with other work. We construct a model of absence incorporating an epidemiological structure and find significant contagion effects of our measure of sickness prevalence in the (rest of the) workforce on the absence probabilities of individual workers. This provides empirical support for both common-sense intuition and theoretical models exploring further aspects of the way in which worker sickness and contagion affect the way in which firms might set their working arrangements.

References

Barmby T, Orme C and Treble J (1995) "Worker Absence Histories: a panel data study" *Labour Economics* Vol 2 pp 53-65.

Barmby T (2002) "Worker Absenteeism: A Discrete Hazard Model with Bivariate Heterogeneity" *Labour Economics* Vol 9 pp 469-476.

Barmby T (1998) "The Relationship between Event History and Discrete Time Duration Models: an Application to the analysis of Personnel Absenteeism" *Oxford Bulletin of Economics and Statistics* Vol 60(2) pp 261-265

Chatterji M and Tilley C (2002) "Sickness Absenteeism, Presenteeism, and Sickpay" *Oxford Economic Papers* Vol 54 pp 669-87

Heckman J.J. (1981) "Statistical models for discrete panel data" In Structural Analysis of Discrete Data with Econometric Applications (ed C.F.Manski and D.McFadden) Cambridge, Mass. MIT press.

Hsiao C (2003) Analysis of Panel Data 2nd Edition CUP

Philipson T (2000) "Economic Epidemiology and Infectious Diseases" Chapter 33 in Handbook of Health Economics Vol 1B by Culyer A J and Newhouse J P.

Skåtun J (2003) "Take some time off, why don't you : endogenous sickleave and pay", *Journal of Health Economics*. Vol. 22, issue 3, Pages 379-402