MEMORANDUM

No 20/2010

Identifying Trend and Age Effects in Sickness Absence from Individual Data: Some Econometric Problems

Erik Biørn

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 19/10	Michael Hoel, Svenn Jensen *Cutting Costs of Catching Carbon* *Intertemporal effects under imperfect climate policy*
No 18/10	Hans Jarle Kind, Tore Nilssen, Lars Sørgard *Price Coordination in Two-Sided Markets: Competition in the TV Industry*
No 17/10	Vladimir Krivonozhko, Finn R. Førsund and Andrey V. Lychev *A Note on Imposing Strong Complementary Slackness Conditions in DEA*
No 16/10	Halvor Mehlum and Karl Moene *Aggressive elites and vulnerable entrepreneurs* *- trust and cooperation in the shadow of conflict*
No 15/10	Nils-Henrik M von der Fehr *Leader, Or Just Dominant? The Dominant-Firm Model Revisited*
No 14/10	Simen Gaure *OLS with Multiple High Dimensional Category Dummies*
No 13/10	Michael Hoel *Is there a green paradox?*
No 12/10	Michael Hoel *Environmental R&D*
No 11/10	Øystein Børsum *Employee Stock Options*
No 10/10	Øystein Børsum *Contagious Mortgage Default*
No 09/10	Derek J. Clark and Tore Nilssen *Learning by Doing in Contests*
No 08/10	Jo Thori Lind *The Number of Organizations in Heterogeneous Societies*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/forskning/publikasjoner/memorandum

IDENTIFYING TREND AND AGE EFFECTS IN SICKNESS ABSENCE FROM INDIVIDUAL DATA: SOME ECONOMETRIC PROBLEMS*)

ERIK BIØRN

Department of Economics, University of Oslo,
P.O. Box 1095 Blindern, 0317 Oslo, Norway

E-mail: erik.biorn@econ.uio.no

ABSTRACT: When using data from individuals who are in the labour force to disentangle the empirical relevance of cohort, age and time effects for sickness absence, the inference may be biased, affected by sorting-out mechanisms. One reason is unobserved heterogeneity potentially affecting both health status and ability to work, which can bias inference because the individuals entering the data set are conditional on being in the labour force. Can this sample selection be adequately handled by attaching unobserved heterogeneity to non-structured fixed effects? In the paper we examine this issue and discuss the econometric setup for identifying from such data time effects in sickness absence. The inference and interpretation problem is caused, on the one hand, by the occurrence of time, cohort and age effects also in the labour market participation, on the other hand by correlation between unobserved heterogeneity in health status and in ability to work. We show that running panel data regressions, ordinary or logistic, of sickness absence data on certain covariates, when neglecting this sample selection, is likely to obscure the interpretation of the results, except in certain, not particularly realistic, cases. However, the fixed individual effects approach is more robust in this respect than an approach controlling for fixed cohort effects only.

KEYWORDS: Sickness absence, health-labour interaction, cohort-age-time problem, self-selection, latent heterogeneity, bivariate censoring, truncated binormal distribution, panel data

JEL CLASSIFICATION: C23, C25, I38, J22

* This paper is part of the project "Absenteeism in Norway – Causes, Consequences, and Policy Implications", financed by the Norwegian Research Council (grant #187924). I am grateful to Knut Røed for comments.

1 Introduction

During the last two decades, the rate at which workers have been absent from work due to sickness – absenteeism – has risen in several countries. Norway, for instance, has seen a sharp increase, from around 4–5 per cent of paid hours in the early 1990s to around 6.5 per cent in 2010. This rise has occurred despite general improvements in self-reported health conditions. In a recent paper, Biørn *et al.* (2010) have, by exploiting individual data on long-term absence spells for virtually all workers in Norway over a 13-year period, addressed this problem empirically, attempting, in particular, to disentangle the empirical relevance of cohort, age and time effects by means of "fixed effect methods". It is obvious that the data available for a study of this kind are potentially affected by sorting-out mechanisms because the individuals entering the data set are conditional on being in the labour force. It may be questioned whether this sample selection can be adequately treated by handling unobserved heterogeneity through fixed effects, and whether suppressing individual heterogeneity and instead conditioning on cohort or age is likely to accentuate the bias in the estimation of time effects.

In this paper we examine this issue and discuss more thoroughly the econometric setup for identifying from such data time effects in sickness absence while, as far as possible, controlling for cohort/age effects and systematic sample selection. The inference and interpretation problems arise, on the one hand, because of the occurrence of time and cohort/age effects also in the labour market participation, on the other hand because unobserved heterogeneity which most likely affects both health status and ability to work. Specifically, the modelling and inference should account for these two latent variables being correlated. Running regressions – ordinary or logistic – of sickness absence data on certain regressors, without accounting for this sample selection, is likely to obscure the interpretation of the findings and make it difficult to explain their message to non-specialists.

The content of the paper is organized as follows. In Section 2 a simple basic model is formulated, explaining jointly degree of ability to work and degree of sickness by time, cohort and age, accounting for the exact collinearity of the latter, as well as individual heterogeneity. The modelling of unobserved heterogeneity and its implication for the interpretation of the coefficients are discussed in Section 3. Derived sickness probabilities are discussed in Section 4, where we emphasize the distinction between conditioning on individual effects and conditioning on cohort or age. Next, in Section 5, models treating degree of sickness as an observable quantitative variable are discussed, while in Section 6 models treating it as binary (sick versus non-sick) are considered. In the two latter sections, selection bias problems and ways of coming to grips with them are put in focus. Some concluding remarks follow in Section 7.

2 Notation and basic model: Heterogeneity unmodelled

Let i and t denote individual and time period (year) and let c_i and a_{it} be birth cohort and age. The three variables are collinear, since by definition

(2.1) $$a_{it} \equiv t - c_i.$$

Let $\mathbf{1}\{\mathcal{A}\}=1$ and $=0$ if event $\mathcal{A}$ is true and untrue, respectively, and define

$$\begin{aligned} w_{it} &= \mathbf{1}\{\text{Individual } i \text{ belongs to the labour force at time } t\}, \\ s_{it} &= \mathbf{1}\{\text{Individual } i \text{ is reported sick at time } t\}. \end{aligned}$$

Let also

$$\begin{aligned} w_{it}^* &= \text{Degree of ability to work, individual } i, \text{ time } t, \\ s_{it}^* &= \text{Degree of sickness, individual } t, \text{ time } t, \end{aligned}$$

both quantitative and continuous, although not frequently observable in this way.

Regardless of whether (w_{it}^*, s_{it}^*) are observable or latent, we postulate that they depend on cohort, time, age, and latent heterogeneity (μ_i^w, μ_i^s) as

(2.2) $$w_{it}^* = \beta_w c_i + \gamma_w t + \delta_w a_{it} + \mu_i^w + \varepsilon_{it}^w,$$

(2.3) $$s_{it}^* = \beta_s c_i + \gamma_s t + \delta_s a_{it} + \mu_i^s + \varepsilon_{it}^s,$$

(2.4) $$\begin{bmatrix} \varepsilon_{it}^w \\ \varepsilon_{it}^s \end{bmatrix} | [c_i, t, \mu_i^w, \mu_i^s] \sim \mathsf{IID}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_w^2 & \sigma_{ws} \\ \sigma_{ws} & \sigma_s^2 \end{bmatrix} \right) \equiv \mathsf{IID}(\mathbf{0}, \mathbf{\Sigma}).$$

where ε_{it}^w and ε_{it}^s are genuine disturbances. Whether or not the covariance matrix $\mathbf{\Sigma}$ is diagonal, *i.e.*, whether $\sigma_{ws}=0$ or $\neq 0$, will be important for the selection bias issue. Ways of modelling the latent individual effects (μ_i^w, μ_i^s) and their consequences will be discussed in Section 3.

We treat cohort, year and age as quantitative, but the terms involving these variables in (2.2)–(2.3) and formulae derived from them can be easily replaced by terms in *cohort, year, age dummies* – if desired. Specifically, we may extend t, c_i, a_{it} to (column) vectors of cohort, time, age dummies, and extend the scalar coefficients and $(\beta_w, \gamma_w, \delta_w)$ and $(\beta_s, \gamma_s, \delta_s)$ to (row) vectors of dummy coefficients, paying regard to the definitional relationships between the dummies which correspond to (2.1).

Our primary objective is to identify γ_s, in combination with β_s or δ_s if possible, while controlling for observed and unobserved heterogeneity. Because cohort, time and age are linearly related, confer (2.1), and the equations under consideration are linear, the dimension of the equations must be reduced accordingly (2.2) or (2.3) is confronted with data. As a starting point for the empirical modelling we therefore can take either of the following versions of the equations:

(2.5) $$\begin{aligned} w_{it}^* &= (\beta_w - \delta_w) c_i + (\gamma_w + \delta_w) t + \mu_i^w + \varepsilon_{it}^w \\ &\equiv (\gamma_w + \beta_w) t + (\delta_w - \beta_w) a_{it} + \mu_i^w + \varepsilon_{it}^w \\ &\equiv (\beta_w + \gamma_w) c_i + (\delta_w + \gamma_w) a_{it} + \mu_i^w + \varepsilon_{it}^w, \end{aligned}$$

(2.6) $$\begin{aligned} s_{it}^* &= (\beta_s - \delta_s) c_i + (\gamma_s + \delta_s) t + \mu_i^s + \varepsilon_{it}^s \\ &\equiv (\gamma_s + \beta_s) t + (\delta_s - \beta_s) a_{it} + \mu_i^s + \varepsilon_{it}^s \\ &\equiv (\beta_s + \gamma_s) c_i + (\delta_s + \gamma_s) a_{it} + \mu_i^s + \varepsilon_{it}^s. \end{aligned}$$

3 Extensions: Modelling systematic heterogeneity

The latent effects are likely to be correlated with observed regressors, for instance because norms with respect to labour force participation and absenteeism are correlated with cohort. Econometrically, a 'norm' is a latent entity, to be attached to, 'proxied by', observable variables to be of relevance. A simple way of formalizing this is

$$(3.1)\qquad \mu_i^w = \alpha_w + \lambda_w c_i + \nu_i^w \equiv \alpha_w + \lambda_w(t - a_{it}) + \nu_i^w,$$

$$(3.2)\qquad \mu_i^s = \alpha_s + \lambda_s c_i + \nu_i^s \equiv \alpha_s + \lambda_s(t - a_{it}) + \nu_i^s,$$

$$(3.3)\qquad \begin{bmatrix} \nu_i^w \\ \nu_i^s \end{bmatrix} |[c_i, t] \sim \mathsf{IID}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \omega_w^2 & \omega_{ws} \\ \omega_{ws} & \omega_s^2 \end{bmatrix}\right) \equiv \mathsf{IID}(\mathbf{0}, \mathbf{\Omega}),$$

and concurrently modify (2.4) to

$$(3.4)\qquad \begin{bmatrix} \varepsilon_{it}^w \\ \varepsilon_{it}^s \end{bmatrix} |[c_i, t, \nu_i^w, \nu_i^s] \sim \mathsf{IID}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_w^2 & \sigma_{ws} \\ \sigma_{ws} & \sigma_s^2 \end{bmatrix}\right) \equiv \mathsf{IID}(\mathbf{0}, \mathbf{\Sigma}).$$

Inserting (3.1)–(3.2) in (2.5)–(2.6), we obtain

$$(3.5)\qquad \begin{aligned} w_{it}^* &= \alpha_w + (\beta_w + \lambda_w - \delta_w)c_i + (\gamma_w + \delta_w)t + \nu_i^w + \varepsilon_{it}^w \\ &\equiv \alpha_w + (\gamma_w + \beta_w + \lambda_w)t + (\delta_w - \beta_w - \lambda_w)a_{it} + \nu_i^w + \varepsilon_{it}^w \\ &\equiv \alpha_w + (\beta_w + \lambda_w + \gamma_w)c_i + (\delta_w + \gamma_w)a_{it} + \nu_i^w + \varepsilon_{it}^w, \end{aligned}$$

$$(3.6)\qquad \begin{aligned} s_{it}^* &= \alpha_s + (\beta_s + \lambda_s - \delta_s)c_i + (\gamma_s + \delta_s)t + \nu_i^s + \varepsilon_{it}^s \\ &\equiv \alpha_s + (\gamma_s + \beta_s + \lambda_s)t + (\delta_s - \beta_s - \lambda_s)a_{it} + \nu_i^s + \varepsilon_{it}^s. \\ &\equiv \alpha_s + (\beta_s + \lambda_s + \gamma_s)c_i + (\delta_s + \gamma_s)a_{it} + \nu_i^s + \varepsilon_{it}^s \end{aligned}$$

This stylized modelling of heterogeneity makes (β_w, β_s) unidentifiable, as it implies that we in (2.5)–(2.6) must extend (β_w, β_s) to $(\beta_w + \lambda_w, \beta_s + \lambda_s)$ and replace (μ_i^w, μ_i^s) by (ν_i^w, ν_i^s). In view of (3.3)–(3.4), the composite disturbances

$$(u_{it}^w, u_{it}^s) = (\nu_i^w + \varepsilon_{it}^w, \nu_i^s + \varepsilon_{it}^s)$$

have a vector error components form with components mutually orthogonal ($\varepsilon_z \perp \nu_z$, $z = w, s$) and orthogonal to both regressors, with standard deviations $(\tau_w, \tau_s) = [(\sigma_w^2 + \omega_w^2)^{\frac{1}{2}}, (\sigma_s^2 + \omega_s^2)^{\frac{1}{2}}]$, covariance $\tau_{ws} = \sigma_{ws} + \omega_{ws}$ and correlation coefficient $\kappa_{ws} = \tau_{ws}/[\tau_w \tau_s]$. We will to some extent stick to (3.1)–(3.4) as a way of modeling systematic heterogeneity on the following.

However, unobserved heterogeneity may be related also to other observable variables than cohort, some of which time-varying, reflecting (gradual) changes in 'norms' ('norm drift'); (3.1)–(3.2) may be argued to be too 'simplistic'. Consider a variant of (2.2)–(2.3) where uni-dimensional heterogeneity (μ_i^w, μ_i^s) is generalized to *two-dimensional* heterogeneity (μ_{it}^w, μ_{it}^s) and (3.1)–(3.2) are extended to

$$\begin{aligned} \mu_{it}^w &= \alpha_w + \lambda_w c_i + \gamma_w^{\ddagger} t + \delta_w^{\ddagger} a_{it} + \nu_i^w + \varepsilon_{it}^{\ddagger w}, \\ \mu_{it}^s &= \alpha_s + \lambda_s c_i + \gamma_s^{\ddagger} t + \delta_s^{\ddagger} a_{it} + \nu_i^s + \varepsilon_{it}^{\ddagger s}. \end{aligned}$$

It is easy to show that this essentially implies extending $(\gamma_w, \delta_w, \gamma_s, \delta_s)$ in (3.5)–(3.6) to include $(\gamma_w^{\ddagger}, \delta_w^{\ddagger}, \gamma_s^{\ddagger}, \delta_s^{\ddagger})$, and $(\varepsilon_{it}^{w}, \varepsilon_{it}^{s})$ to include $(\varepsilon_{it}^{\ddagger w}, \varepsilon_{it}^{\ddagger s})$, respectively.

Obvious, but important, conclusions so far are:

CONCLUSION 1: *The interpretation of 'time effect in absenteeism' depend on which mechanism determines the two kinds of unobserved heterogeneity and whether cohort or age is the other control variable.*

CONCLUSION 2: *The time effects in absenteeism obtained from (2.6), with (2.4) assumed, and with heterogeneity accounted for, i.e., $\gamma_s+\delta_s$ or $\gamma_s+\beta_s$, may be a more stable 'structure' – the equation has a higher degree of 'autonomy' – than the time effects according to (3.6), with (3.3)–(3.4) assumed, or extensions of it. The latter, unlike the former, changes when the parameters of (3.2) change.*

4 SICKNESS PROBABILITIES

4.1 *Threshold values for sickness and ability to work*

As remarked, w_{it}^* and s_{it}^*, in particular the former, may not be observable as continuous variables, while their *qualitative* counterparts – whether or not individual i is in the labour force and/or is sick at time t – are usually known. Let $\bar{w}, \bar{s}$ be unknown *critical threshold values* for the two continuous variables determining the status 'being in the labour force' and 'being reported sick':

$w_{it}^* \geq \bar{w} \implies$ Individual i is observed belonging to the labour force.

$s_{it}^* \geq \bar{s} \implies$ Individual i is observed being declared sick by a doctor.

The work ability threshold $\bar{w}$ may be time invariant or time dependent, in the latter case capturing, *inter alia*, (worker) 'norm drift', the sickness threshold $\bar{s}$ may, likewise, be time invariant or time dependent, in the latter case also capturing (worker) 'norm drift' as well as drift in doctors' norms or attitudes with respect to issuing sickness certificates. We want to derive expressions for the corresponding sickness probabilities. Let, as a start, $\psi(u,v)$ be the joint density of the standardized disturbances in (2.5)–(2.6), or in (3.5)–(3.6), *i.e.*, $(u,v)=(\varepsilon_{it}^{w}/\sigma_w, \varepsilon_{it}^{s}/\sigma_s)$, or $(u,v)=(u_{it}^{w}/\tau_w, u_{it}^{s}/\tau_s)$, and define, for arbitrary a,b,

$$f(a,b) = P(u>a,\, v>b) = \int_a^\infty \int_b^\infty \psi(u,v)\,du\,dv, \tag{4.1}$$

$$g(a,b) = P(v>b|u>a) = \frac{P(u>a,\, v>b)}{P(u>a)} = \frac{f(a,b)}{f(a,-\infty)}. \tag{4.2}$$

In (2.5)–(2.6), while utilizing (3.1)–(3.2), it is convenient to define

$$\begin{aligned} \mu_i^{w*} &= (\beta_w-\delta_w)c_i + \mu_i^w = \alpha_w + (\beta_w+\lambda_w-\delta_w)c_i + \nu_i^w, \\ \mu_i^{w\dagger} &= (\beta_w+\gamma_w)c_i + \mu_i^w = \alpha_w + (\beta_w+\lambda_w+\gamma_w)c_i + \nu_i^w, \end{aligned} \tag{4.3}$$

$$\begin{aligned} \mu_i^{s*} &= (\beta_s-\delta_s)c_i + \mu_i^s = \alpha_s + (\beta_s+\lambda_s-\delta_s)c_i + \nu_i^s, \\ \mu_i^{s\dagger} &= (\beta_s+\gamma_s)c_i + \mu_i^s = \alpha_s + (\beta_s+\lambda_s+\gamma_s)c_i + \nu_i^s. \end{aligned} \tag{4.4}$$

They can be interpreted as representing 'gross individual heterogeneity' inclusive of cohort effects. Then (3.5)–(3.6) can be rewritten more simply as

$$w_{it}^{*} = (\gamma_w + \delta_w)t + \mu_i^{w*} + \varepsilon_{it}^{w} \equiv (\delta_w + \gamma_w)a_{it} + \mu_i^{w\dagger} + \varepsilon_{it}^{w}, \tag{4.5}$$

$$s_{it}^{*} = (\gamma_s + \delta_s)t + \mu_i^{s*} + \varepsilon_{it}^{s} \equiv (\delta_s + \gamma_s)a_{it} + \mu_i^{s\dagger} + \varepsilon_{it}^{s}. \tag{4.6}$$

Combining these expressions with (2.5)–(2.6), using the definition of the binary variables in Section 2, we obtain

$$\begin{aligned} w_{it} = 1 \iff & w_{it}^{*} \geq \bar{w} \iff \varepsilon_{it}^{w} \geq \bar{w} - (\gamma_w + \delta_w)t - \mu_i^{w*} = \bar{w} - (\gamma_w + \delta_w)a_{it} - \mu_i^{w\dagger}, \\ s_{it} = 1 \iff & s_{it}^{*} \geq \bar{s} \iff \varepsilon_{it}^{s} \geq \bar{s} - (\gamma_s + \delta_s)t - \mu_i^{s*} = \bar{s} - (\gamma_s + \delta_s)a_{it} - \mu_i^{s\dagger}. \end{aligned}$$

We introduce, in order to simplify notation, putting the kind of parameters identifiable from binary response data (confer Section 6) in focus, two sets of *rescaled parameters*, obtained by normalizing coefficients and thresholds against the relevant disturbance standard deviations. The first is related to (2.5)–(2.6), the second to (3.5)–(3.6), giving, respectively, *'σ-normalized' parameters:*

$$\gamma_{w\sigma} = \frac{\gamma_w + \delta_w}{\sigma_w}, \qquad \bar{w}_\sigma = \frac{\bar{w}}{\sigma_w}, \qquad \mu_{iw\sigma} = \frac{\mu_i^{w*}}{\sigma_w}, \qquad \mu_{iw\sigma}^{\dagger} = \frac{\mu_i^{w\dagger}}{\sigma_w}, \tag{4.7}$$

$$\gamma_{s\sigma} = \frac{\gamma_s + \delta_s}{\sigma_s}, \qquad \bar{s}_\sigma = \frac{\bar{s}}{\sigma_s}, \qquad \mu_{is\sigma} = \frac{\mu_i^{s*}}{\sigma_s}, \qquad \mu_{is\sigma}^{\dagger} = \frac{\mu_i^{s\dagger}}{\sigma_s}, \tag{4.8}$$

and *'τ-normalized' parameters:*

$$\gamma_{w\tau} = \frac{\gamma_w + \delta_w}{\tau_w}, \qquad \beta_{w\tau} = \frac{\beta_w + \lambda_w - \delta_w}{\tau_w}, \qquad \bar{w}_\tau = \frac{\bar{w}}{\tau_w}, \tag{4.9}$$

$$\gamma_{s\tau} = \frac{\gamma_s + \delta_s}{\tau_s}, \qquad \beta_{s\tau} = \frac{\beta_s + \lambda_s - \delta_s}{\tau_s}, \qquad \bar{s}_\tau = \frac{\bar{s}}{\tau_s}, \tag{4.10}$$

where, obviously, $(\bar{w}_\tau, \bar{s}_\tau, \gamma_{w\tau}, \gamma_{s\tau})$ are smaller (in absolute value) than $(\bar{w}_\sigma, \bar{s}_\sigma, \gamma_{w\sigma}, \gamma_{s\sigma})$.[1] We then obtain from (2.5)–(2.6)

$$w_{it} = 1 \iff \tfrac{\varepsilon_{it}^{w}}{\sigma_w} \geq \bar{w}_\sigma - \gamma_{w\sigma}t - \mu_{iw\sigma} = \bar{w}_\sigma - \gamma_{w\sigma}a_{it} - \mu_{iw\sigma}^{\dagger}, \tag{4.11}$$

$$s_{it} = 1 \iff \tfrac{\varepsilon_{it}^{s}}{\sigma_s} \geq \bar{s}_\sigma - \gamma_{s\sigma}t - \mu_{is\sigma} = \bar{s}_\sigma - \gamma_{s\sigma}a_{it} - \mu_{is\sigma}^{\dagger}, \tag{4.12}$$

and, likewise, from (3.5)–(3.6)

$$w_{it} = 1 \iff \tfrac{u_{it}^{w}}{\tau_w} \geq \bar{w}_\tau - \gamma_{w\tau}t - \beta_{w\tau}c_i = \bar{w}_\tau - (\gamma_{w\tau} + \beta_{w\tau})t + \beta_{w\tau}a_{it}, \tag{4.13}$$

$$s_{it} = 1 \iff \tfrac{u_{it}^{s}}{\tau_s} \geq \bar{s}_\tau - \gamma_{s\tau}t - \beta_{s\tau}c_i = \bar{s}_\tau - (\gamma_{s\tau} + \beta_{s\tau})t + \beta_{s\tau}a_{it}. \tag{4.14}$$

[1]Possible smooth 'norm-drift' in $\bar{w}$ and $\bar{s}$ could be absorbed into $(\gamma_{w\tau}, \gamma_{s\tau})$ or $(\gamma_{w\sigma}, \gamma_{s\sigma})$.

4.2 *Probabilities conditional on individual effects*

Conditioning on individual effects, we can, using (4.1)–(4.2), (4.7)–(4.8) and (4.11)–(4.12), express the probability of being sick unconditionally and conditional on being in the labour force, as, respectively,[2]

$$\begin{aligned}
(4.15)\qquad P(s_{it}=1;t,\mu_{is\sigma}) &= f(-\infty,\bar{s}_\sigma-\gamma_{s\sigma}t-\mu_{is\sigma})\\
&= f(-\infty,\bar{s}_\sigma-\gamma_{s\sigma}a_{it}-\mu^{\dagger}_{is\sigma}),\\
(4.16)\qquad P(s_{it}=1|w_{it}=1;t,\mu_{iw\sigma},\mu_{is\sigma}) &= g(\bar{w}_\sigma-\gamma_{w\sigma}t-\mu_{iw\sigma},\bar{s}_\sigma-\gamma_{s\sigma}t-\mu_{is\sigma})\\
&= g(\bar{w}_\sigma-\gamma_{w\sigma}a_{it}-\mu^{\dagger}_{iw\sigma},\bar{s}_\sigma-\gamma_{s\sigma}a_{it}-\mu^{\dagger}_{is\sigma}).
\end{aligned}$$

If ε^w_{it} and ε^s_{it} are stochastically independent, then

$$\begin{aligned}
&g(\bar{w}_\sigma-\gamma_{w\sigma}t-\mu_{iw\sigma},\bar{s}_\sigma-\gamma_{s\sigma}t-\mu_{is\sigma})\\
&\qquad\equiv g(-\infty,\bar{s}_\sigma-\gamma_{s\sigma}t-\mu_{is\sigma})\equiv f(-\infty,\bar{s}_\sigma-\gamma_{s\sigma}t-\mu_{is\sigma}).
\end{aligned}$$

4.3 *Probabilities conditional on cohort or on age*

Conditioning instead on cohort, or equivalently on age, we can, using (4.1)–(4.2), (4.9)–(4.10) and (4.13)–(4.14), express the probability of being sick unconditionally and conditional on being in the labour force, as, respectively,[3]

$$\begin{aligned}
(4.17)\qquad P(s_{it}=1;t,c_i) &= f(-\infty,\bar{s}_\tau-\gamma_{s\tau}t-\beta_{s\tau}c_i)\\
&= f(-\infty,\bar{s}_\tau-(\gamma_{s\tau}+\beta_{s\tau})t+\beta_{s\tau}a_{it}),\\
(4.18)\qquad P(s_{it}=1|w_{it}=1;t,c_i) &= g(\bar{w}_\tau-\gamma_{w\tau}t-\beta_{w\tau}c_i,\bar{s}_\tau-\gamma_{s\tau}t-\beta_{s\tau}c_i)\\
&= g(\bar{w}_\tau-(\gamma_{w\tau}+\beta_{w\tau})t+\beta_{w\tau}a_{it},\bar{s}_\sigma-(\gamma_{s\tau}+\beta_{s\tau})t+\beta_{s\tau}a_{it}).
\end{aligned}$$

If not only ε^w_{it} and ε^s_{it}, but also ν^w_i and ν^s_i are stochastically independent, then

$$\begin{aligned}
&g(\bar{w}_\tau-\gamma_{w\tau}t-\beta_{w\tau}c_i,\bar{s}_\tau-\gamma_{s\tau}t-\beta_{s\tau}c_i)\\
&\qquad\equiv g(-\infty,\bar{s}_\tau-\gamma_{s\tau}t-\beta_{s\tau}c_i)\equiv f(-\infty,\bar{s}_\tau-\gamma_{s\tau}t-\beta_{s\tau}c_i).
\end{aligned}$$

5 Models treating sickness as quantitative.

In this section, leaving the probability expressions in Section 4, we return to the setup presented in Sections 2 and 3 and consider three models with sickness assumed quantitatively observable, say measured as the number of sickness days per unit of time. All models condition on time or on age; otherwise they differ with respect to the conditioning assumed: the individual effect (Section 5.1), the birth-cohort (Section 5.2), the age (Section 5.3). Conditioning on age and on cohort give, however, models which mirror models where the conditioning is on time and cohort. We assume throughout that the observable variables are s^*_{it}, w_{it}, t, c_i.

[2]Formally, the latter probability is conditional both on being in the labour force, and on unobserved individual-specific heterogeneity in sickness and ability to work.

[3]Formally, the latter probability is conditional both on being in the labour force, and on the observed cohort to which the individual belongs.

5.1 *Conditioning on individual effect*

Assume first that (μ_i^w, μ_i^s) are treated as fixed effects and, accordingly, that the heterogeneity submodel (3.1)–(3.3) is 'suspended'. It follows from (2.4)–(2.6) and (4.6) that *only* $\gamma_s+\delta_s$ *and the composite parameters* μ_i^{s*} *defined in (4.4) can be identified.* With respect to the sample, we distinguish between two cases:

[A] *If the sample were not censored by labour force participation*, the sick-leave trend estimated by regressing s_{it}^* linearly on (t, μ_i^{s*}), would have been $\gamma_s+\delta_s$, since then

$$\mathsf{E}(s_{it}^*|t, \mu_i^{s*}) = (\gamma_s+\delta_s)t + \mu_i^{s*}. \tag{5.1}$$

[B] *If the sample is censored by labour force participation*, the sick-leave trend we actually estimate differs from $\gamma_s+\delta_s$. We have

$$\mathsf{E}(s_{it}^*|w_{it}=1; t, \mu_i^{s*}) = (\gamma_s+\delta_s)t + \mu_i^{s*} + \mathsf{E}(\varepsilon_{it}^s|w_{it}=1; t, \mu_i^{s*}). \tag{5.2}$$

This equation exemplifies a bivariate sample selection model, whose last term accounts for the sample selection; see, *e.g.*, Cameron and Trivedi (2005, Section 16.5.3). This model type is sometimes referred to as Amemiya's 'Type 2 Tobit Model'; confer Amemiya (1985, Section 10.7).

In the *binormal* case, where

$$\psi(u, v) = (2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}} e^{-\frac{1}{2}(u^2-2\rho uv+v^2)/(1-\rho^2)},$$

we can express $\mathsf{E}(\varepsilon_{it}^s|w_{it}=1; t, \mu_i^{s*})$ analytically as follows. Letting $\phi(\cdot)$ and $\Phi(\cdot)$ be the univariate normal density and c.d.f., respectively, we get, by exploiting $\phi'(u) = -u\phi(u)$, $\mathsf{E}(v|u) = \rho u$, and $\mathsf{E}(u|a \le u \le b) = [\phi(a)-\phi(b)]/[\Phi(b)-\Phi(a)]$ [see, Johnson, Kotz and Balakrishnan (1994, Section 10.1) or Biørn (2008, Appendix 8A)]:

$$\mathsf{E}(v|a \le u \le b) = \rho \frac{\phi(a)-\phi(b)}{\Phi(b)-\Phi(a)}, \tag{5.3}$$

and also that, for any a,

$$\lambda(a) \equiv \frac{\phi(a)}{\Phi(a)} \equiv \frac{\phi(-a)}{1-\Phi(-a)}, \tag{5.4}$$

$$\lambda'(a) \equiv -\xi(a) = -\lambda(a)[\lambda(a)+a]. \tag{5.5}$$

Therefore, *if* $(\varepsilon_{it}^w, \varepsilon_{it}^s)$ *are binormal*, letting $\rho_{ws} = \sigma_{ws}/(\sigma_w\sigma_s)$ and using the 'σ-normalized' parameters (4.7)–(4.8), we obtain

$$\begin{aligned} \mathsf{E}(\varepsilon_{it}^s|w_{it}=1, t, \mu_i^{s*}) &= \mathsf{E}[\varepsilon_{it}^s|\varepsilon_{it}^w \ge \bar{w} - (\gamma_w+\delta_w)t - \mu_i^{w*}; t, c_i] \\ &= \sigma_s \mathsf{E}(\tfrac{\varepsilon_{it}^s}{\sigma_s}|\tfrac{\varepsilon_{it}^w}{\sigma_w} \ge \bar{w}_\sigma - \gamma_{w\sigma}t - \mu_{iw\sigma}; t, c_i) \\ &= \rho_{ws}\sigma_s\lambda(\gamma_{w\sigma}t - \bar{w}_\sigma + \mu_{iw\sigma}). \end{aligned} \tag{5.6}$$

Inserting (5.6) in (5.2) we obtain

$$\begin{aligned} \mathsf{E}(s_{it}^*|w_{it}=1; t, \mu_i^{s*}) &= (\gamma_s+\delta_s)t + \mu_i^{s*} + \rho_{ws}\sigma_s\lambda(\gamma_{w\sigma}t - \bar{w}_\sigma + \mu_{iw\sigma}) \\ &= \sigma_s[\gamma_{s\sigma}t + \mu_{is\sigma} + \rho_{ws}\lambda(\gamma_{w\sigma}t - \bar{w}_\sigma + \mu_{iw\sigma})]. \end{aligned} \tag{5.7}$$

Hence, utilizing (5.5), we find that the correct sickness trend, allowing for the systematic censoring, is, in general, non-linear and given by

$$\begin{aligned}(5.8)\qquad \partial \mathsf{E}(s_{it}^{*}|w_{it}=1;t,\mu_i^{s*})/\partial t &= \gamma_s+\delta_s + \rho_{ws}\sigma_s\partial\lambda(\gamma_{w\sigma}t-\bar{w}_\sigma+\mu_{iw\sigma})/\partial t \\ &= \sigma_s[\gamma_{s\sigma}-\gamma_{w\sigma}\rho_{ws}\xi(\gamma_{w\sigma}t-\bar{w}_\sigma+\mu_{iw\sigma})].\end{aligned}$$

If $\rho_{ws}\neq 0$, *i.e.*, if the genuine disturbances in (2.2) and (2.3) are correlated, the sickness trend (5.8) depends on $\sigma_s, \gamma_{w\sigma}, \bar{w}_\sigma$ and $\mu_{iw\sigma}$. *Hence, when $\rho_{ws}\neq 0$, the correct trend will be individual-specific.*

What can be said about the sign of the last component in (5.8)? First, (5.5) implies that $\xi(\gamma_{w\sigma}t-\bar{w}_\sigma+\mu_{iw\sigma})$ is likely to be positive. Second, assume that some common unspecified factors lead both to absenteeism and drop-out from the labour force and hence $\rho_{ws}<0$. Third, assume that the trend in inclusion into (exclusion from) the labour market is negative (positive), *i.e.*, $\gamma_{w\sigma}<0$. Hence, (5.8) most likely implies that $\partial\mathsf{E}(s_{it}^{*}|w_{it}=1;t,c_i,\mu_i^{s})/\partial t<\gamma_s+\delta_s$.

> Conclusion 3: *If the sample is censored by labour force participation and $\rho_{ws}\neq 0$, the (theoretical) regression $\mathsf{E}(s_{it}^{*}|w_{it}=1;t,\mu_i^{s*})$ is, in general, non-linear in (t,μ_i^{s*}). Its form depends on the coefficients of (2.5) and (2.6) as well as the distribution of $\varepsilon_{it}^{w},\varepsilon_{it}^{s}$, as expressed by (5.7) in the binormal case. A linear regression of s_{it}^{*} on (t,μ_i^{s*}) will result in biased estimation of the composite sickness trend coefficient $\sigma_s\gamma_{s\sigma}=\gamma_s+\delta_s$ and the composite individual effects μ_i^{s*}. If $\rho_{ws}=0$ the bias disappears: $\partial\mathsf{E}(s_{it}^{*}|w_{it}=1;t,\mu_i^{s*})/\partial t=\sigma_s\gamma_{s\sigma}=\gamma_s+\delta_s$. In the latter case, (2.5)–(2.6) form a recursive structure, conditional on the individual effects: first labour market participation is decided, next sickness is determined. Conditional on the individual effects, there are no latent elements bringing feedback from the latter to the former.*

5.2 *Conditioning on birth-cohort*

Assume next that heterogeneity modeled as (3.1)–(3.2) is part of the model, and let c_i be the conditioning variable in addition to t. It follows from (3.3)–(3.6) and (4.6) that *only $\gamma_s+\delta_s$ and $\beta_s+\lambda_s-\delta_s$ (or one-to-one transformations of them) can be identified.* With respect to the sample, we again distinguish between two cases.

[A] *If the sample were non-censored*, the trend coefficient we would have estimated by regressing s_{it}^{*} linearly on (t,c_i) would have been $\gamma_s+\delta_s$, since then

$$(5.9)\qquad \mathsf{E}(s_{it}^{*}|t,c_i)=\alpha_s+(\gamma_s+\delta_s)t+(\beta_s+\lambda_s-\delta_s)c_i.$$

[B] *If the sample is censored by labour force participation*, the trend coefficient actually estimated by regressing s_{it}^{*} on (t,c_i) differs from $\gamma_s+\delta_s$. We have

$$(5.10)\qquad \mathsf{E}(s_{it}^{*}|w_{it}=1;t,c_i)=\alpha_s+(\gamma_s+\delta_s)t+(\beta_s+\lambda_s-\delta_s)c_i+\mathsf{E}(u_{it}^{s}|w_{it}=1;t,c_i).$$

This equation exemplifies again a bivariate sample selection model, whose last term accounts for the effects of the sample selection on the expected response variable. Now, however, the origin of the selection is the composite disturbance $u_{it}^{s}=\nu_i^{s}+\varepsilon_{it}^{s}$.

Assume in addition that $(\varepsilon_{it}^{w}, \varepsilon_{it}^{s})$ and (ν_i^w, ν_i^s) are binormal, implying that (u_{it}^{w}, u_{it}^{s}) are binormal with standard deviations (τ_w, τ_s), covariance τ_{ws} and correlation coefficient κ_{ws}. From (5.3), (5.4) and (5.10), introducing the 'τ-normalized' parameters, (4.9)–(4.10), we then obtain

$$
\begin{aligned}
(5.11)\quad \mathsf{E}(s_{it}^*|w_{it}=1;t,c_i) &= \alpha_s+(\gamma_s+\delta_s)t+(\beta_s+\lambda_s-\delta_s)c_i+\kappa_{ws}\tau_s\lambda(\gamma_{w\tau}t+\beta_{w\tau}c_i-\bar{w}_\tau) \\
&= \alpha_s+\tau_s[\gamma_{s\tau}t+\beta_{s\tau}c_i+\kappa_{ws}\lambda(\gamma_{w\tau}t+\beta_{w\tau}c_i-\bar{w}_\tau)],
\end{aligned}
$$

where $\kappa_{ws}\neq 0$ if *at least one of* σ_{ws} and ω_{ws} is non-zero. We then find, in a similar way as (5.7), that the correct trend and the correct cohort effects are, in general, non-linear and given by, respectively,

$$
\begin{aligned}
(5.12)\quad \partial\mathsf{E}(s_{it}^*|w_{it}=1;t,c_i)/\partial t &= \gamma_s+\delta_s+\kappa_{ws}\tau_s\partial\lambda(\gamma_{w\tau}t+\beta_{w\tau}c_i-\bar{w}_\tau)/\partial t \\
&= \tau_s[\gamma_{s\tau}-\gamma_{w\tau}\kappa_{ws}\xi(\gamma_{w\tau}t+\beta_{w\tau}c_i-\bar{w}_\tau)],
\end{aligned}
$$

$$
\begin{aligned}
(5.13)\quad \partial\mathsf{E}(s_{it}^*|w_{it}=1;t,c_i)/\partial c_i &= \beta_s+\lambda_s-\delta_s+\kappa_{ws}\tau_s\partial\lambda(\gamma_{w\tau}t+\beta_{w\tau}c_i-\bar{w}_\tau)/\partial c_i \\
&= \tau_s[\beta_{s\tau}-\beta_{w\tau}\kappa_{ws}\xi(\gamma_{w\tau}t+\beta_{w\tau}c_i-\bar{w}_\tau)].
\end{aligned}
$$

If $\kappa_{ws}\neq 0$, both derivatives depend on $\tau_s, \gamma_{w\tau}, \beta_{w\tau}$ and $\bar{w}_\tau$, which implies that *the correct trend is cohort-specific, while the correct cohort effect is time-varying.*

What can be said about the sign of the last components in (5.12) and (5.13)? First, (5.5) implies that $\xi(\gamma_{w\tau}t+\beta_{w\tau}c_i-\bar{w}_\tau)$ is likely to be positive. Second, assume [1] that some common latent individual-specific factors lead to absenteeism and drop-out from labour force and hence $\omega_{ws}<0$, or [2] that some unspecified time-varying factors also lead to absenteeism and drop-out from labour force and hence $\sigma_{ws}<0$. Together, [1] or [2] suggests $\kappa_{ws}<0$. Third, assume that the trend in inclusion into (exclusion from) the labour force is negative (positive), *i.e.*, $\gamma_{w\tau}<0$. Hence, (5.12) implies that $\partial\mathsf{E}(s_{it}^*|w_{it}=1;t,c_i)/\partial t<\gamma_s+\delta_s$.

> CONCLUSION 4: *If the sample is censored by labour force participation, the (theoretical) regression* $\mathsf{E}(s_{it}^*|w_{it}=1;t,c_i)$ *is, in general, non-linear in* (t,c_i). *Its form depends on the coefficients of both (2.2)–(2.3) and (3.1)–(3.2), as well as the distribution of* (u_{it}^w,u_{it}^s), *as expressed by (5.11) in the binormal case. A linear (empirical) regression of* s_{it}^* *on* (t,c_i) *will result in biased estimation of the adjusted trend coefficient* $\gamma_s+\delta_s$. *If both* $\sigma_{ws}=0$ *and* $\omega_{ws}=0$ *hold, implying* $\kappa_{ws}=0$, *the biases disappear:* $\partial\mathsf{E}(s_{it}^*|w_{it}=1;t,c_i)/\partial t=\tau_s\gamma_{s\tau}=\gamma_s+\delta_s$ *and* $\partial\mathsf{E}(s_{it}^*|w_{it}=1;t,c_i)/\partial c_i=\tau_s\beta_{s\tau}=\beta_s+\lambda_s-\delta_s$. *In the latter case, (3.5)–(3.6) form a recursive structure, unconditional on the individual effects: first labour market participation is decided, next sickness is determined. Conditional on cohort, but unconditional on the individual effects, there is no feedback from the latter to the former.*

5.3 *Conditioning on age*

Assume again that (3.1)–(3.2) are part of the model, and let t and a_{it} be the conditioning variables. It follows from (3.3)–(3.6) and (4.6) that *only* $\gamma_s+\beta_s+\lambda_s$ *and*

$\delta_s\!-\!\beta_s\!-\!\lambda_s$ *(or one-to-one transformations of them) can be identified.* With respect to the sample, we again distinguish between two cases.

[**A**] *If the sample were non-censored*, the trend coefficient we would have estimated by regressing s^*_{it} linearly on (t, a_{it}) would have been $\gamma_s\!+\!\beta_s\!+\!\lambda_s$, since then

$$\mathsf{E}(s^*_{it}|t, a_{it}) = \alpha_s + (\gamma_s\!+\!\beta_s\!+\!\lambda_s)t + (\delta_s\!-\!\beta_s\!-\!\lambda_s)a_{it}. \tag{5.14}$$

[**B**] *If the sample is censored by labour force participation*, the trend coefficient actually estimated by regressing s^*_{it} on (t, a_{it}) differs from $\gamma_s\!+\!\beta_s\!+\!\lambda_s$. We have[4]

$$\mathsf{E}(s^*_{it}|w_{it}\!=\!1; t, a_{it}) = \alpha_s\!+\!(\gamma_s\!+\!\beta_s\!+\!\lambda_s)t\!+\!(\delta_s\!-\!\beta_s\!-\!\lambda_s)a_{it}\!+\!\mathsf{E}(u^s_{it}|w_{it}\!=\!1; t, a_{it}). \tag{5.15}$$

From (5.3), (5.4), (5.15) and (4.9), we obtain, in the binormal case,

$$\begin{aligned}\mathsf{E}(s^*_{it}|w_{it}\!=\!1; t, a_{it}) = {} & \alpha_s + (\gamma_s\!+\!\beta_s\!+\!\lambda_s)t + (\delta_s\!-\!\beta_s\!-\!\lambda_s)a_{it} \\ & + \kappa_{ws}\tau_s\lambda[(\gamma_{w\tau}\!+\!\beta_{w\tau})t\!-\!\beta_{w\tau}a_{it}\!-\!\bar{w}_\tau].\end{aligned} \tag{5.16}$$

The correct trend and the correct age effects therefore become[5]

$$\begin{aligned}&\partial\mathsf{E}(s^*_{it}|w_{it}\!=\!1;t, a_{it})/\partial t \\ &\quad = \gamma_s\!+\!\beta_s\!+\!\lambda_s\!+\!\kappa_{ws}\tau_s\partial\lambda[(\gamma_{w\tau}\!+\!\beta_{w\tau})t\!-\!\beta_{w\tau}a_{it}\!-\!\bar{w}_\tau]/\partial t \\ &\quad = \tau_s\left[(\gamma_{s\tau}\!+\!\beta_{s\tau})\!-\!(\gamma_{w\tau}\!+\!\beta_{w\tau})\kappa_{ws}\xi[(\gamma_{w\tau}\!+\!\beta_{w\tau})t\!-\!\beta_{w\tau}a_{it}\!-\!\bar{w}_\tau]\right],\end{aligned} \tag{5.17}$$

$$\begin{aligned}&\partial\mathsf{E}(s^*_{it}|w_{it}\!=\!1;t, a_{it})/\partial a_{it} \\ &\quad = \delta_s\!-\!\beta_s\!-\!\lambda_s\!+\!\kappa_{ws}\tau_s\partial\lambda[(\gamma_{w\tau}\!+\!\beta_{w\tau})t\!-\!\beta_{w\tau}a_{it}\!-\!\bar{w}_\tau]/\partial a_{it} \\ &\quad = \tau_s\left[-\beta_{s\tau}\!+\!\beta_{w\tau}\kappa_{ws}\xi[(\gamma_{w\tau}\!+\!\beta_{w\tau})t\!-\!\beta_{w\tau}a_{it}\!-\!\bar{w}_\tau]\right].\end{aligned} \tag{5.18}$$

If $\kappa_{ws}\neq 0$, both derivatives depend on $\tau_s, \gamma_{w\tau}, \beta_{w\tau}$ and $\bar{w}_\tau$, which implies that *the correct trend is age-specific and the correct age effect is time-varying.*

> Conclusion 5: *If the sample is censored by labour force participation, the (theoretical) regression $\mathsf{E}(s^*_{it}|w_{it}\!=\!1; t, a_{it})$ is, in general, non-linear in (t, a_{it}). Its form depends on the coefficients of both (2.2)–(2.3) and (3.1)–(3.2), as well as the distribution of (u^w_{it}, u^s_{it}), as given by (5.16) in the binormal case. A linear (empirical) regression of s^*_{it} on (t, a_{it}) will result in biased estimation of the actual trend coefficient $\gamma_s+\beta_s+\lambda_s$ and of the adjusted age coefficient $\delta_s-\beta_s-\lambda_s$. If both $\sigma_{ws}=0$ and $\omega_{ws}=0$ hold, implying $\kappa_{ws}=0$, the biases disappear: $\partial\mathsf{E}(s^*_{it}|w_{it}=1; t, a_{it})/\partial t = \tau_s(\gamma_{s\tau}+\beta_{s\tau}) = \gamma_s+\beta_s+\lambda_s$ and $\partial\mathsf{E}(s^*_{it}|w_{it}=1; t, a_{it})/\partial a_{it} = -\tau_s\beta_{s\tau} = \delta_s-\beta_s-\lambda_s$. Then (3.5)–(3.6) form a recursive structure, unconditional on the individual effects. Conditional on cohort, but unconditional on the individual effects, there is no feedback from sickness to labour force participation.*

[4] This equation, of course, mirrors (5.10).
[5] These equations mirror (5.12)–(5.13).

6 Models treating sickness as dichotomously observable

6.1 *General remarks*

Having explored the situation where the degree of absenteeism, s^*_{it}, is assumed to be recorded quantitatively, we next consider models where absenteeism is assumed to be recorded qualitatively (dichotomously). This may sometimes be a more realistic assumption. Or even if continuous observations are available, the analyst may want to exploit it only dichotomously for 'institutional' reasons, because of measurement problems which may plague the data collection, suggesting a need for 'robustifying' the results, etc. This corresponds to the approach of Biørn *et al.* (2010). With respect to the sample, we distinguish between cases [**A**] and [**B**], as in Section 5.

[**A**] *Data for all individuals, whether in the labour force or outside, are in the sample.* Then we could want to make inference on trend effects in the sickness probability from (4.15) or (4.17). If we base inference on (4.15) when (3.1)–(3.2) are part of the data generating mechanism, using standard *binomial* logit or probit analysis – and hence conditioning on c_i or a_{it} – we would estimate τ*-normalized* coefficients. If we base inference on (4.17), using binomial logit or probit analysis – and hence conditioning on $\mu_{is\sigma}$ – we would estimate σ*-normalized* coefficients.[6] Derivatives of the (log-)probabilities, 'marginal effects', could be estimated from either.

[**B**] *The sample is only from individuals being in the labour force.* Then, to obtain valid inference on trend effects in the sickness probability, we should account for the implicit censoring. Again, we could only obtain inference on τ- or σ-normalized coefficients. Since the relevant sickness-absence probabilities underlying our binary response data are conditional on $w_{it}=1$, they are of the form (4.16) or (4.18). When conditioning on $\mu_{is\sigma}$ ($\mu^\dagger_{is\sigma}$), we obtain more robust inference on the trend in the sickness probability than when conditioning on c_i (a_{it}).

To see this we differentiate the relevant expressions for the conditional log-probability of absenteeism with respect to time and the other relevant covariates. Let $\Psi_u(u;b)=\int_b^\infty \psi(u,v)dv$ and $\Psi_v(v;a)=\int_a^\infty \psi(u,v)du$, and write (4.1) as

$$(6.1)\qquad f(a,b)=\int_a^\infty \Psi_u(u;b)du=\int_b^\infty \Psi_v(v;a)dv \Longrightarrow \begin{cases} \partial f(a,b)/\partial a=-\Psi_u(a;b), \\ \partial f(a,b)/\partial b=-\Psi_v(b;a). \end{cases}$$

Now differentiation of (4.2) gives

$$(6.2)\qquad \frac{\partial g(a,b)}{\partial a} = -g(a,b)G_a(a,b) \iff \frac{\partial \ln g(a,b)}{\partial a} = -G_a(a,b),$$

$$(6.3)\qquad \frac{\partial g(a,b)}{\partial b} = -g(a,b)G_b(a,b) \iff \frac{\partial \ln g(a,b)}{\partial b} = -G_b(a,b),$$

[6] In both cases the non-normalized coefficients are non-identifiable when only discrete information is exploited since no metric for (w^*_{it}, s^*_{it}) and $(\bar{w}, \bar{s})$ is exploited.

where

$$G_a(a,b) = \frac{\Psi_u(a;b)}{f(a,b)} - \frac{\Psi_u(a;-\infty)}{f(a,-\infty)} = \frac{\int_b^\infty \psi(a,v)dv}{\int_a^\infty\int_b^\infty \psi(u,v)du\,dv} - \frac{\int_{-\infty}^\infty \psi(a,v)dv}{\int_a^\infty\int_{-\infty}^\infty \psi(u,v)du\,dv},$$

$$G_b(a,b) = \frac{\Psi_v(b;a)}{f(a,b)} = \frac{\int_a^\infty \psi(u,b)du}{\int_a^\infty\int_b^\infty \psi(u,v)du\,dv}.$$

6.2 *Conditioning on individual effect*

It follows by combining (6.2)–(6.3) with (4.16) that the derivative of the log-probability of absenteeism with respect to time is

$$\begin{aligned}(6.4)\qquad & \partial \ln P(s_{it}=1|w_{it}=1;t,\mu_{iw\sigma},\mu_{is\sigma})/\partial t \\ &= \partial \ln g(\bar{w}_\sigma - \gamma_{w\sigma}t - \mu_{iw\sigma}, \bar{s}_\sigma - \gamma_{s\sigma}t - \mu_{is\sigma})/\partial t \\ &= \gamma_{s\sigma} G_b(\bar{w}_\sigma - \gamma_{w\sigma}t - \mu_{iw\sigma}, \bar{s}_\sigma - \gamma_{s\sigma}t - \mu_{is\sigma}) \\ &\quad + \gamma_{w\sigma} G_a(\bar{w}_\sigma - \gamma_{w\sigma}t - \mu_{iw\sigma}, \bar{s}_\sigma - \gamma_{s\sigma}t - \mu_{is\sigma}).\end{aligned}$$

The first term after the last equality sign represents the direct effect of the trend in absenteeism – mirroring the effect of the trend term in (2.6). It is positive when $\gamma_{s\sigma} > 0$ since $G_b(\bar{w}_\sigma - \gamma_{w\sigma}t - \mu_{iw\sigma}, \bar{s}_\sigma - \gamma_{s\sigma}t - \mu_{is\sigma})$ is positive. The second term represents the indirect effect, via the trend in the ability to work and dropping out of the labour market – mirroring the effect of the trend term in (2.5). It is negative if $\gamma_{w\sigma} < 0$, since $G_a(\bar{w}_\sigma - \gamma_{w\sigma}t - \mu_{iw\sigma}, \bar{s}_\sigma - \gamma_{s\sigma}t - \mu_{is\sigma})$ is, most likely, positive.

6.3 *Conditioning on cohort or age*

Combining (6.2)–(6.3) with (4.18), it follows, likewise, that

$$\begin{aligned}(6.5)\qquad & \partial \ln P(s_{it}=1|w_{it}=1;t,c_i)/\partial t \\ &= \partial \ln g(\bar{w}_\tau - \gamma_{w\tau}t - \beta_{w\tau}c_i, \bar{s}_\sigma - \gamma_{s\tau}t - \beta_{s\tau}c_i)/\partial t \\ &= \gamma_{s\tau} G_b(\bar{w}_\tau - \gamma_{w\tau}t - \beta_{w\tau}c_i, \bar{s}_\tau - \gamma_{s\tau}t - \beta_{s\tau}c_i) \\ &\quad + \gamma_{w\tau} G_a(\bar{w}_\tau - \gamma_{w\tau}t - \beta_{w\tau}c_i, \bar{s}_\tau - \gamma_{s\tau}t - \beta_{s\tau}c_i),\end{aligned}$$

$$\begin{aligned}(6.6)\qquad & \partial \ln P(s_{it}=1|w_{it}=1;t,c_i)/\partial c_i \\ &= \partial \ln g(\bar{w}_\tau - \gamma_{w\tau}t - \beta_{w\tau}c_i, \bar{s}_\sigma - \gamma_{s\tau}t - \beta_{s\tau}c_i)/\partial c_i \\ &= \beta_{s\tau} G_b(\bar{w}_\tau - \gamma_{w\tau}t - \beta_{w\tau}c_i, \bar{s}_\tau - \gamma_{s\tau}t - \beta_{s\tau}c_i) \\ &\quad + \beta_{w\tau} G_a(\bar{w}_\tau - \gamma_{w\tau}t - \beta_{w\tau}c_i, \bar{s}_\tau - \gamma_{s\tau}t - \beta_{s\tau}c_i),\end{aligned}$$

or equivalently

$$\begin{aligned}(6.7)\quad & \partial \ln P(s_{it}=1|w_{it}=1;t,a_{it})/\partial t \\ &= \partial \ln g(\bar{w}_\tau - (\gamma_{w\tau}+\beta_{w\tau})t + \beta_{w\tau}a_{it}, \bar{s}_\sigma - (\gamma_{s\tau}+\beta_{s\tau})t + \beta_{s\tau}a_{it})/\partial t \\ &= (\gamma_{s\tau}+\beta_{s\tau}) G_b(\bar{w}_\tau - (\gamma_{w\tau}+\beta_{w\tau})t + \beta_{w\tau}a_{it}, \bar{s}_\sigma - (\gamma_{s\tau}+\beta_{s\tau})t + \beta_{s\tau}a_{it}) \\ &\quad + (\gamma_{w\tau}+\beta_{w\tau}) G_a(\bar{w}_\tau - (\gamma_{w\tau}+\beta_{w\tau})t + \beta_{w\tau}a_{it}, \bar{s}_\sigma - (\gamma_{s\tau}+\beta_{s\tau})t + \beta_{s\tau}a_{it}),\end{aligned}$$

$$\begin{aligned}(6.8)\quad & \partial \ln P(s_{it}=1|w_{it}=1;t,a_{it})/\partial a_{it} \\ &= \partial \ln g(\bar{w}_\tau - (\gamma_{w\tau}+\beta_{w\tau})t + \beta_{w\tau}a_{it}, \bar{s}_\sigma - (\gamma_{s\tau}+\beta_{s\tau})t + \beta_{s\tau}a_{it})/\partial a_{it} \\ &= -\beta_{s\tau} G_b(\bar{w}_\tau - (\gamma_{w\tau}+\beta_{w\tau})t + \beta_{w\tau}a_{it}, \bar{s}_\sigma - (\gamma_{s\tau}+\beta_{s\tau})t + \beta_{s\tau}a_{it}) \\ &\quad - \beta_{w\tau} G_a(\bar{w}_\tau - (\gamma_{w\tau}+\beta_{w\tau})t + \beta_{w\tau}a_{it}, \bar{s}_\sigma - (\gamma_{s\tau}+\beta_{s\tau})t + \beta_{s\tau}a_{it}).\end{aligned}$$

Again, the first terms after the last equality signs represents the direct effects, while the second terms represent the indirect effect, via the ability to work and dropping out of the labour market.

6.4 *The recursive case and a synthesis*

It is illuminating to compare the last five expressions with those obtained when the structure is recursive, *i.e.*, $\rho_{ws}=0$ or $\tau_{ws}=0$. We then have $g(a,b) = \int_b^\infty \phi(v)dv = 1-\Phi(b)$, which imply:[7]

$$G_a(a,b) = 0, \quad G_b(a,b) = \frac{\phi(b)}{1-\Phi(b)} = \frac{\phi(-b)}{\Phi(-b)} = \lambda(-b).$$

Then (6.4)–(6.8) are simplified to

$$\rho_{ws} = 0 \text{ (Recursivity conditional on individual effects)} \implies$$

$$\frac{\partial \ln P(s_{it}{=}1|w_{it}{=}1;t,\mu_{iw\sigma},\mu_{is\sigma})}{\partial t} = \frac{\partial \ln P(s_{it}{=}1;t,\mu_{is\sigma})}{\partial t} = \gamma_{s\sigma}\lambda(\gamma_{s\sigma}t+\mu_{is\sigma}-\bar{s}_\sigma),$$

$$\tau_{ws} = 0 \text{ (Recursivity conditional on cohort)} \implies$$

$$\frac{\partial \ln P(s_{it}{=}1|w_{it}{=}1;t,c_i)}{\partial t} = \frac{\partial \ln P(s_{it}{=}1;t,c_i)}{\partial t} = \gamma_{s\tau}\lambda(\gamma_{s\tau}t+\beta_{s\tau}c_i-\bar{s}_\tau),$$

$$\frac{\partial \ln P(s_{it}{=}1|w_{it}{=}1;t,c_i)}{\partial c_i} = \frac{\partial \ln P(s_{it}{=}1;t,c_i)}{\partial c_i} = \beta_{s\tau}\lambda(\gamma_{s\tau}t+\beta_{s\tau}c_i-\bar{s}_\tau),$$

$$\tau_{ws} = 0 \text{ (Recursivity conditional on age)} \implies$$

$$\frac{\partial \ln P(s_{it}{=}1|w_{it}{=}1;t,a_{it})}{\partial t} = \frac{\partial \ln P(s_{it}{=}1;t,a_{it})}{\partial t} = (\gamma_{s\tau}{+}\beta_{s\tau})\lambda((\gamma_{s\tau}+\beta_{s\tau})t-\beta_{s\tau}a_{it}-\bar{s}_\tau),$$

$$\frac{\partial \ln P(s_{it}{=}1|w_{it}{=}1;t,a_{it})}{\partial a_{it}} = \frac{\partial \ln P(s_{it}{=}1;t,a_{it})}{\partial a_{it}} = -\beta_{s\tau}\lambda((\gamma_{s\tau}{+}\beta_{s\tau})t-\beta_{s\tau}a_{it}-\bar{s}_\tau).$$

CONCLUSION 6: *When we condition on (a) the individual latent heterogeneity or (b) cohort (or age) only, we should account for sample truncation when formulating the appropriate response probabilities as functions of covariates and likelihood functions for estimating trends in absenteeism, except when $\sigma_{ws}=0$ (in case (a)) and $\omega_{ws}=\sigma_{ws}=0$ (in latter case (b)). The correct form of the likelihood function will, in the general case, reflect the mixture of discrete choice and sample truncation.*

7 CONCLUDING REMARKS

We have in this paper presented a simple model framework for analyzing jointly degree of sickness and degree of work ability with two kinds of latent heterogeneity

[7] If u,v are independent $\implies \psi(u,v)=\phi(u)\phi(v)$, $f(a,b)=[1{-}\Phi(a)][1{-}\Phi(b)]$, $g(a,b)=[1{-}\Phi(b)]\,\forall\,a$, where $\phi(u)$ and $\Phi(u)$ are the univariate density and the c.d.f. of u or v [confer Section 4.1], then $\Psi_u(u;b)=[1{-}\Phi(b)]\phi(u)$, $\Psi_v(v;a)=[1{-}\Phi(a)]\phi(v)$, and hence $G_a(a,b)=0$, $G_b(a,b)=\phi(b)/[1{-}\Phi(b)]$.

interacting, one related to absenteeism (sickness absence), the other related to ability to work. Obtaining valid inference on trend effects has been the main focus of the paper. Sometimes also cohort effects or age effects can be uncovered. We have shown that correlation pattern of the two kinds of latent heterogeneity is important. Treating the two decisions as recursive may not be always be the answer, and neglecting the sample selection may obscure the interpretation of the coefficients estimated.

An overall conclusion, somewhat related to and extending conclusions derived for bivariate 'Tobit models' in literature, is that when we stick to linear regression, the conditions which need to be satisfied for estimated composite trends (time effects) to be unbiased are stronger when the other covariate (conditioning variable) is cohort or age, than when we condition on individual effects (and, by implication, eliminate any relationship between individual heterogeneity and cohort). In the former case, the genuine disturbances in the underlying sickness equation and work ability equation should be uncorrelated. The latter case, a kind of 'double recursivity' should hold: both the genuine disturbances and the latent individual effects in the two equations should be uncorrelated. Inference on sickness absence trends obtained by linear regression with fixed individual effects (additive shifts in the intercept) included, may therefore be characterized as *more robust* than that obtained when including only cohort or age as regressors and throwing all heterogeneity into (gross) disturbances. Essentially, these conclusions also carry over to the case where absenteeism is only observed dichotomously.

Natural, and rather straightforward, extensions, not elaborated in the paper, could be to replace the time, cohort and age variable by corresponding time, cohort and age dummies. Genuine 'economic regressors' could also be included, formally as extensions of the models' intercepts, except that no such regressor could be individual specific, in order to avoid perfect collinearity with the individual effects. Neither could, for a similar reason, time specific regressors be included in models where time dummies replace the continuous time variable.

References

Amemiya, T. (1985): *Advanced Econometrics.* Cambridge (Ma.): Harvard University Press.

Biørn, E. (2008): *Økonometriske emner. En videreføring.* Oslo: Unipub.

Biørn, E., Gaure, S., Markussen, S., and Røed, K. (2010): The Rise in Absenteeism: Disentangling the Impacts of Cohort, Age and Time. IZA Discussion Paper No. 5091. IZA, Bonn.

Cameron, A.C. and Trivedi, P.K. (2005): *Microeconometrics. Methods and Applications.* Cambridge: Cambridge University Press.

Johnson, N.L., Kotz, S., and Balakrishnan, N. (1994): *Continuous Univariate Distributions, Volume 1, Second Edition.* New York: Wiley.