MEMORANDUM

No 19/2013

Age-Cohort-Time Effects in Sickness Absence: Exploring a Large Data Set by Polynomial Regression

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 18/13	Halvor Mehlum and Gry Østenstad *The Political Economy of Migration Politics in Oil-rich Gulf States*
No 17/13	Halvor Mehlum, Ragnar Torvik and Simone Valente *China's Savings Multiplier*
No 16/13	Finn R. Førsund *Measuring Efficiency and Effectiveness in the Public Sector*
No 15/13	Mads Greaker and Kristoffer Midttømme *Optimal Environmental Policy with Network Effects: Is Lock-in in Dirty Technologies Possible?*
No 14/13	Diderik Lund and Ragnar Nymoen *Comparative Statistics for Real Options on Oil: What Stylized Facts to Use?*
No 13/13	Nils-Henrik M. von der Fehr *Transparency in Electricity Markets*
No 12/13	Nils Chr. Framstad *When Can Environmental Profile and Emissions Reduction Be Optimized Independently of the Pollutant Level*
No 11/13	Nils Chr. Framstad and Jon Strand *Energy Intensive Infrastructure Investments with Retrofits in Continuous Time: Effects of Uncertainty on Energy Use and Carbon Emissions*
No 10/13	Øystein Kravdal *Reflections on the Search for Fertillity Effects on Happiness*
No 09/13	Erik Biørn and Hild-Marte Bjørnsen *What Motivates Farm Couples to Seek Off-farm Labour? A Logit Analysis of Job Transitions*

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

AGE-COHORT-TIME EFFECTS IN SICKNESS ABSENCE: EXPLORING A LARGE DATA SET BY POLYNOMIAL REGRESSION

Memo 19/2013-v1
(June 2013)

ERIK BIØRN

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

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

ABSTRACT: Identification of equations explaining a continuous variable, *e.g.*, the length of sickness absence spells, by age, cohort and time (ACT), subject to their definitional identity is reconsidered. Various extensions of a linear equation to polynomials are explored. If no interactions between the ACT variables are included, only the coefficients of the linear terms create identification problems. A data set with 4.5 million individual observations for long-term sickness absence in Norway is used. The sensitivity of the estimated marginal effects of cohort and age on the length of the absence spells, at the sample mean, is illustrated. Notable differences are found between linear and quadratic equations on the one hand and cubic and fourth-order polynomials on the other. There are also notable gender differences. Representing heterogeneity by cohort effects is compared with representing heterogeneity by random and fixed individual effects. On the whole, the age coefficients in the estimated regressions are quite sensitive to how heterogeneity is modeled.

KEYWORDS: Age-cohort-time problem, identification, sickness absence, sickness and gender, panel data, polynomial regression, interaction, heterogeneity.

JEL CLASSIFICATION: C23, C24, C25, C52, H55, I18, J21.

ACKNOWLEDGEMENTS: This paper is part of the project "Absenteeism in Norway – Causes, Consequences, and Policy Implications", funded by the Norwegian Research Council (grant #187924). I am grateful to Daniel Bergsvik for preparing the data file and other valuable assistance and to Knut Røed and Terje Skjerpen for comments.

1 Introduction

When attempting to uncover relationships from individual data, the 'Age-Cohort-Time (ACT) problem', due to the identity cohort+age = time, and ways of handling it in different contexts is much discussed among social and medical researchers; see Mason *et al.* (1973), Rodgers (1982), Portrait, Alessie, and Deeg (2002), Hall, Mairesse and Turner (2007), McKenzie (2006), Winship and Harding (2008), Yang and Land (2008), and Ree and Alessie (2011). The ACT identification problem has motivated additional assumptions to reduce the parameter space. It is notorious in linear models, but also when using more flexible functional forms, *e.g.*, polynomials, problems of parameter identification arise.

An example is the problem of disentangling partial effects of birth cohort, age and time on a measure of sickness absence of individuals. Biørn *et al.* (2013), using a large set of individual data on long-term absence spells that cover virtually all workers in Norway over a 13-year period, have addressed this problem recently. They set out to separate cohort, age and time effects in the *discrete sick/non-sick response*, representing the effects by dummy variables through a fixed effects logit approach, considering the response as equivalent regardless of whether the sickness duration was, say, one week more than the minimum of 16 days, or six months.

In this paper the ACT problem for individual sickness is reconsidered in a setting which represents the degree of sickness as a continuous variable, the length of the absence spells, with age, cohort and time also measured continuously. This approach, of course, exploits a lot more of the information in the data set. It also gives challenges in quantifying marginal effects of age and cohort, related *inter alia* to the form of the relationship. Starting from a linear model, we extend it to polynomials in age, cohort and time of order up to four, with focus on *interactions* between the three variables.

A general specification of the theoretical regression for the ACT problem – with y denoting a variable to be explained, (a, c, t) the explanatory variables age, cohort, time, satisfying $a+c=t$, and x a vector of other explanatory variables – is

$$\mathsf{E}(y|a,c,t,x) = f(a,c,t,x). \tag{1}$$

Eliminating one of the ACT variables, we can write the equation as

$$\begin{aligned} \mathsf{E}(y|a,c,x) &= f(a,c,a+c,x) \equiv F_1(a,c,x), \\ \mathsf{E}(y|c,t,x) &= f(t-c,c,t,x) \equiv F_2(c,t,x), \\ \mathsf{E}(y|a,t,x) &= f(a,t-a,t,x) \equiv F_3(a,t,x). \end{aligned} \tag{2}$$

An additive subclass of (1) has the form

$$\mathsf{E}(y|a,c,t,x) = f_a(a,x) + f_c(c,x) + f_t(t,x), \tag{3}$$

and can be rewritten alternatively as

$$\begin{aligned} \mathsf{E}(y|a,c,x) &= f_a(a,x) + f_c(c,x) + f_t(a+c,x) \equiv \phi_1(a,c,x), \\ \mathsf{E}(y|c,t,x) &= f_a(t-c,x) + f_c(c,x) + f_t(t,x) \equiv \phi_2(c,t,x), \\ \mathsf{E}(y|a,t,x) &= f_a(a,x) + f_c(t-a,x) + f_t(t,x) \equiv \phi_3(a,t,x). \end{aligned} \tag{4}$$

Which of the parameters of f (or of f_a, f_c, f_t) when F_1, F_2 and F_3 (or ϕ_1, ϕ_2 or ϕ_3) are known, can be identified, depends on the functional form chosen. If f is linear, or a monotonically increasing transformation of a linear function, not all parameters can be identified. This is, loosely speaking, because the linearity of f 'interferes with' the linear definitional identity. If f, possibly after a monotonic transformation, is the sum of a linear and a non-linear part, the linear part still creates identification problems, while similar problems may not arise for the coefficients of the non-linear part.[1] If g is restricted to be non-linear, we have, for example, $g(a)+g(c) \neq g(t)$. For polynomials we can be more specific: while *e.g.* t^3 and (a^3, c^3) are not collinear, t^3 is collinear with (a^3, c^3, a^2c, ac^2), and so on. This simple example indicates that when linear functions are extended to polynomials, coefficient identification may crucially depend on whether interactions between age, cohort and time are included and on how their coefficients are restricted. This is one of the issues to be addressed.

The paper proceeds as follows. In Section 2 the ACT problem for a model with f (and f_a, f_c, f_t) linear and x omitted is reconsidered as a benchmark. In Section 3 we extend f (or in the additive subcase (3), f_a, f_c, f_t) to polynomials, and show that an ACT problem for the coefficients of the linear terms still exists, but that the coefficients of second- and higher order terms of f_a, f_c, f_t can be identified. The extent to which coefficients of higher-order terms in the more general polynomial version of (1) can be identified, depends on which interactions between the ACT variables are included and on their parameter restrictions. Alternative definitions of marginal effects for such models are then elaborated in Section 4.[2] Next, in Section 5, this framework is used, for polynomial orders up to four, to explore age, cohort and time effects in sickness absence from absence records from more than 1.7 million individuals in the Norwegian labour force during a 14 year period. Gender differences are examined. We conclude that long-term sickness, in absence days, is clearly non-linear in cohort and age and that the model's fit is significantly improved when polynomial additivity is relaxed by including interactions between cohort and age, at least for polynomials of order up to four. There are clear gender differences in the coefficient pattern. The overall fit, measured by R^2, is still poor, however. Modifications of the polynomial models where heterogeneity as random and fixed individual effects occurs are in Section 6 compared with the versions where heterogeneity is accounted for by (polynomial) cohort effects. This improves overall fit somewhat, but not much. Section 7 concludes.

2 Revisiting the Age-Cohort-Time problem in a linear model

Observations from n individuals on a response variable y_i, for example the length of a sickness absence spell and three covariates, birth cohort, time and age of individual i, (c_i, t_i, a_i), are assumed to be available and in the initial specification assumed to be related by the equation

$$\mathsf{E}(y_i|c_i, t_i, a_i) = \alpha + \gamma c_i + \delta t_i + \beta a_i, \qquad i = 1, \ldots, n. \tag{5}$$

[1] Fisher (1961, p. 575) indeed refers to the "the frequent claim that non-linearities aid identification or even (the claim) that the identification problem does not arise in many non-linear systems".

[2] An example of a non-linear relationship recently given attention is the possible convexity of life satisfaction, on an ordinal scale, as a function of age when estimated from panel data; see Ree and Alessie (2011) and Baetschmann (2012).

Other explanatory variables, corresponding to x in (1), are suppressed, but could easily have been included by extending the intercept α. Since in any realistic data set

$$a_i + c_i = t_i, \qquad i = 1, \dots, n, \tag{6}$$

neither of γ, δ, β represents partial effects. If, however, we believe that $\delta = 0$ and impose this as an *a priori* restriction, then γ and β can be identified as pure cohort and age effects. We have, as an example of (2),

$$\begin{aligned} \Delta \mathsf{E}(y_i|\Delta c_i, \Delta t_i, \Delta a_i) &= (\gamma+\delta)\Delta c_i + (\beta+\delta)\Delta a_i \\ &= (\gamma-\beta)\Delta c_i + (\delta+\beta)\Delta t_i \\ &= (\beta-\gamma)\Delta a_i + (\delta+\gamma)\Delta t_i. \end{aligned} \tag{7}$$

The first-order conditions for the OLS problem for (5), subject to (6), exemplifies solving a system of linear equations subject to linear variable restrictions. The problem $\min_{\alpha,\delta,\beta,\gamma} \sum_{i=1}^{n} u_i^2$, where $u_i = y_i - \mathsf{E}(y_i|c_i, t_i, a_i)$ subject to $a_i + c_i = t_i$ gives three independent conditions. Therefore only two linear combinations of the slope coefficients can be identified: either $(\gamma+\delta), (\beta+\delta)$ or $(\delta+\gamma), (\beta-\gamma)$ or $(\gamma-\beta), (\delta+\beta)$.[3] Boundary cases are:

Data from *one cohort*: Only $\beta+\delta$ can be identified, letting *either* a_i *or* t_i *be regressor.*
Data from *one period*: Only $\beta-\gamma$ can be identified, letting *either* a_i *or* c_i *be regressor.*
Data from *one age*: Only $\gamma+\delta$ can be identified, letting *either* c_i *or* t_i *be regressor.*

3 Extension to polynomial models

We consider two extensions of (5), the first has the additive form (3), the second has the more general form (1).

Additive polynomial in age, cohort and time: The first extension is a sum of Pth order polynomials in a_i, c_i, t_i, exemplifying (3), which has $3P$ coefficients. Eliminating, by using (6), alternatively, t_i, a_i and c_i, we can write the polynomial equation, now exemplifying (4), as respectively:

$$\mathsf{E}(y_i|a_i, c_i) = \alpha + \textstyle\sum_{p=1}^{P} \beta_p^* a_i^p + \sum_{p=1}^{P} \gamma_p^* c_i^p + \sum_{p=1}^{P} \delta_p^* (a_i+c_i)^p, \tag{8}$$

$$\mathsf{E}(y_i|c_i, t_i) = \alpha + \textstyle\sum_{p=1}^{P} \beta_p^* (t_i-c_i)^p + \sum_{p=1}^{P} \gamma_p^* c_i^p + \sum_{p=1}^{P} \delta_p^* t_i^p, \tag{9}$$

$$\mathsf{E}(y_i|a_i, t_i) = \alpha + \textstyle\sum_{p=1}^{P} \beta_p^* a_i^p + \sum_{p=1}^{P} \gamma_p^* (t_i-a_i)_i^p + \sum_{p=1}^{P} \delta_p^* t_i^p. \tag{10}$$

We call this an additive Pth order polynomial. Since, from the binomial formula,

$$\begin{aligned} t_i^p &= (a_i+c_i)^p = \textstyle\sum_{r=0}^{p} \binom{p}{r} a_i^r c_i^{p-r} \equiv c_i^p + \sum_{r=1}^{p-1} \binom{p}{r} a_i^r c_i^{p-r} + a_i^p, \\ a_i^p &= (t_i-c_i)^p = \textstyle\sum_{r=0}^{p} \binom{p}{r} t_i^r (-c_i)^{p-r} \equiv (-c_i)^p + \sum_{r=1}^{p-1} \binom{p}{r} t_i^r (-c_i)^{p-r} + t_i^p, \\ c_i^p &= (t_i-a_i)^p = \textstyle\sum_{r=0}^{p} \binom{p}{r} t_i^r (-a_i)^{p-r} \equiv (-a_i)^p + \sum_{r=1}^{p-1} \binom{p}{r} t_i^r (-a_i)^{p-r} + t_i^p, \end{aligned}$$

(8)–(10) can be reparametrized to give

$$\mathsf{E}(y_i|a_i, c_i) = \alpha + \beta_1 a_i + \gamma_1 c_i + \textstyle\sum_{p=2}^{P} \beta_p a_i^p + \sum_{p=2}^{P} \gamma_p c_i^p + \sum_{p=2}^{P} \delta_p \sum_{r=1}^{p-1} \binom{p}{r} a_i^r c_i^{p-r}, \tag{11}$$

$$\mathsf{E}(y_i|c_i, t_i) = \alpha + \bar{\delta}_1 t_i + \bar{\gamma}_1 c_i + \textstyle\sum_{p=2}^{P} \bar{\delta}_p t_i^p + \sum_{p=2}^{P} \bar{\gamma}_p c_i^p + \sum_{p=2}^{P} \bar{\beta}_p \sum_{r=1}^{p-1} \binom{p}{r} t_i^r (-c_i)^{p-r}, \tag{12}$$

$$\mathsf{E}(y_i|a_i, t_i) = \alpha + \tilde{\beta}_1 a_i + \tilde{\delta}_1 t_i + \textstyle\sum_{p=2}^{P} \tilde{\beta}_p a_i^p + \sum_{p=2}^{P} \tilde{\delta}_p t_i^p + \sum_{p=2}^{P} \tilde{\gamma}_p \sum_{r=1}^{p-1} \binom{p}{r} t_i^r (-a_i)^{p-r}, \tag{13}$$

[3] For an expanded discussion, see Biørn (2013).

with coefficients (all identifiable without additional conditions being needed):

$$
\begin{array}{lllll}
(14) & \beta_1=\beta_1^*+\delta_1^*,\ \gamma_1=\gamma_1^*+\delta_1^*, & \delta_p=\delta_p^*,\ \beta_p=\beta_p^*+\delta_p^*,\ \gamma_p=\gamma_p^*+\delta_p^*, & p=2,\dots,P,\\
(15) & \bar{\delta}_1=\delta_1^*+\beta_1^*,\ \bar{\gamma}_1=\gamma_1^*-\beta_1^*, & \bar{\beta}_p=\beta_p^*,\ \bar{\delta}_p=\delta_p^*+\beta_p^*,\ \bar{\gamma}_p=\gamma_p^*+(-1)^p\beta_p^*, & p=2,\dots,P,\\
(16) & \tilde{\delta}_1=\delta_1^*+\gamma_1^*,\ \tilde{\beta}_1=\beta_1^*-\gamma_1^*, & \tilde{\gamma}_p=\gamma_p^*,\ \tilde{\delta}_p=\delta_p^*+\gamma_p^*,\ \tilde{\beta}_p=\beta_p^*+(-1)^p\gamma_p^*, & p=2,\dots,P.
\end{array}
$$

This shows that although an additive Pth order polynomial in (a_i, c_i, t_i) has seemingly no interactions, its reparametrization which creates, for example, (11) from (8), implies interactions between the (powers of the) two remaining variables and reduces the number of identifiable coefficients to $C_1=3P-1$.

Full polynomial: The above additive ACT polynomials, which exemplify (3)–(4), have an 'asymmetry'. To obtain a model which exemplifies (1)–(2) they can be extended to polynomials with a full set of interaction terms *for all powers of orders* $2,\dots,P-1$ *in, respectively,* (a_i, c_i), (t_i, c_i) *or* (t_i, a_i). The increased flexibility this creates has the potential to improve the fit to data, an issue to be addressed in Sections 5 and 6. We elaborate this extension only for (8), reparametrized as (11), and specify

$$(17)\qquad \mathsf{E}(y_i|a_i,c_i) = \alpha + \textstyle\sum_{p=1}^{P}\beta_p a_i^p + \sum_{p=1}^{P}\gamma_p c_i^p + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}a_i^r c_i^{p-r},$$

which has $C_2 = 2P+\frac{1}{2}P(P-1) = \frac{1}{2}P(P+3)$ coefficients. If $P > 2$, this is an effective increase, since C_2–C_1=$\frac{1}{2}P(P-3)+1$.[4] Model (17) specializes to (11) for

$$(18)\qquad \delta_{pr} = \binom{p}{r}\delta_p, \qquad p=2,\dots,P;\ r=1,\dots,p-1.$$

Example: Consider a full *fourth-order polynomial* ($P=4$), for which (17) gives

$$
\begin{aligned}
\mathsf{E}(y_i|a_i,c_i) = \alpha &+ \beta_1 a_i + \gamma_1 c_i + \beta_2 a_i^2 + \gamma_2 c_i^2\\
&+ \beta_3 a_i^3 + \gamma_3 c_i^3 + \beta_4 a_i^4 + \gamma_4 c_i^4\\
&+ \delta_{21}a_i c_i + \delta_{31}a_i^2 c_i + \delta_{32}a_i c_i^2\\
&+ \delta_{41}a_i^3 c_i + \delta_{42}a_i^2 c_i^2 + \delta_{43}a_i c_i^3.
\end{aligned}
$$

Imposing the $C_2-C_1=3$ restrictions $\delta_{31}=\delta_{32}(=3\delta_3)$ and $\delta_{41}=\delta_{43}=\frac{2}{3}\delta_{42}(=4\delta_4)$, implied by (18), we get, after a reparametrization which replaces $(\delta_{21},\delta_{31},\delta_{32},\delta_{41},\delta_{42},\delta_{43})$ by $(\delta_2,\delta_3,\delta_4)$, the additive polynomial model

$$
\begin{aligned}
\mathsf{E}(y_i|a_i,c_i) = \alpha + \beta_1 a_i + \gamma_1 c_i &+ \beta_2 a_i^2 + \gamma_2 c_i^2 + \delta_2 2a_i c_i\\
&+ \beta_3 a_i^3 + \gamma_3 c_i^3 + \delta_3(3a_i^2 c_i + 3a_i c_i^2)\\
&+ \beta_4 a_i^4 + \gamma_4 c_i^4 + \delta_4(4a_i^3 c_i + 6a_i^2 c_i^2 + 4a_i c_i^3).
\end{aligned}
$$

4 Marginal effects

In the empirical application to be considered, *demeaned observations* of cohort, time and age will be used. This is done not only to reduce the variables' magnitude – a notable advantage when forming powers and interactions – but also, and more importantly, to facilitate comparison of results across models of different orders.

[4] A third model with three polynomials and all interactions between (a, c), (a, t) and (c, t) included, would have had $C_3=3P+3\frac{1}{2}P(P-1)=\frac{3}{2}P(P+1)$ coefficients and hence $C_3-C_2=P^2$. It is, however, hypothetical since the inescapable restriction (6) precludes identification of all its coefficients. For examples and further discussion, see Biørn (2013).

A basis for interpreting the coefficient estimates is obtained by taking a look at expressions for 'marginal effects' of cohort and age. The following notation for *central* moments will then be needed: Let $\boldsymbol{a} = a - \mathsf{E}(a)$ and $\boldsymbol{c} = c - \mathsf{E}(c)$, and define[5]

$$\begin{aligned}
\boldsymbol{\mu}_a(p) &= \mathsf{E}[\boldsymbol{a}^p], & \boldsymbol{\mu}_c(q) &= \mathsf{E}[\boldsymbol{c}^q], \\
\boldsymbol{\mu}_{a|c}(p) &= \mathsf{E}[\boldsymbol{a}^p|\boldsymbol{c}], & \boldsymbol{\mu}_{c|a}(q) &= \mathsf{E}[\boldsymbol{c}^q|\boldsymbol{a}], \qquad & p, q = 1, 2, \dots . \\
\boldsymbol{\mu}_{ac}(p,q) &= \mathsf{E}[\boldsymbol{a}^p\boldsymbol{c}^q],
\end{aligned}$$

Corresponding to (17), after having deducted from cohort and age their expectations, *i.e.*, the theoretical counterpart to demeaning, we obtain[6]

$$\text{(19)} \qquad \mathsf{E}(y|\boldsymbol{a},\boldsymbol{c}) = \alpha + \beta_1\boldsymbol{a} + \gamma_1\boldsymbol{c} + \textstyle\sum_{p=2}^{P}\beta_p\boldsymbol{a}^p + \sum_{p=2}^{P}\gamma_p\boldsymbol{c}^p + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}\boldsymbol{a}^r\boldsymbol{c}^{p-r}.$$

The law of iterated expectations gives

$$\begin{aligned}
&\text{(20)} \qquad \mathsf{E}(y|\boldsymbol{a}) = \alpha + \beta_1\boldsymbol{a} + \textstyle\sum_{p=2}^{P}\beta_p\boldsymbol{a}^p + \sum_{p=2}^{P}\gamma_p\boldsymbol{\mu}_c(p) + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}\boldsymbol{a}^r\boldsymbol{\mu}_{c|a}(p-r), \\
&\text{(21)} \qquad \mathsf{E}(y|\boldsymbol{c}) = \alpha + \gamma_1\boldsymbol{c} + \textstyle\sum_{p=2}^{P}\beta_p\boldsymbol{\mu}_a(p) + \sum_{p=2}^{P}\gamma_p\boldsymbol{c}^p + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}\boldsymbol{\mu}_{a|c}(r)\boldsymbol{c}^{p-r}, \\
&\text{(22)} \qquad \mathsf{E}(y) = \alpha + \textstyle\sum_{p=2}^{P}\beta_p\boldsymbol{\mu}_a(p) + \sum_{p=2}^{P}\gamma_p\boldsymbol{\mu}_c(p) + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}\boldsymbol{\mu}_{ac}(r,p-r).
\end{aligned}$$

Two kinds of marginal effects 'at the mean' can now be defined.

Expected marginal effects: Definition 1 (Expectations of first-derivatives): The marginal expectations of the derivatives of sickness absence, y, with respect to age, $\boldsymbol{a}$, and cohort, $\boldsymbol{c}$ can be expressed in terms of population moments as[7]

$$\text{(23)} \qquad \begin{aligned}
\mathsf{E}[\partial y/\partial\boldsymbol{a}] &= \beta_1 + \textstyle\sum_{p=3}^{P}\beta_p p\boldsymbol{\mu}_a(p-1) + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}r\boldsymbol{\mu}_{ac}(r-1,p-r), \\
\mathsf{E}[\partial y/\partial\boldsymbol{c}] &= \gamma_1 + \textstyle\sum_{p=3}^{P}\gamma_p p\boldsymbol{\mu}_c(p-1) + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}(p-r)\boldsymbol{\mu}_{ac}(r,p-r-1).
\end{aligned}$$

Since β_2 and γ_2, *i.e.*, the coefficients of the quadratic terms in (11), do not enter these expressions, we for linear and quadratic relations simply have $\mathsf{E}[\partial y/\partial\boldsymbol{a}] = \beta_1$ and $\mathsf{E}[\partial y/\partial\boldsymbol{c}] = \gamma_1$. If $P \geq 3$, second and higher-order moments of age and cohort, interacting with the coefficients of the cubic and higher-order terms, will also be involved.

Expected marginal effects: Definition 2 (First-derivatives of conditional expectations): Two versions of these effects can be obtained from (19). Conditioning on both age and cohort and differentiating with respect to one of them, we get, respectively,

$$\text{(24)} \qquad \begin{aligned}
\partial\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c})/\partial\boldsymbol{a} &= \beta_1 + \textstyle\sum_{p=2}^{P}\beta_p p\,\boldsymbol{a}^{p-1} + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}r\boldsymbol{a}^{r-1}\boldsymbol{c}^{p-r}, \\
\partial\mathsf{E}(y|\boldsymbol{c},\boldsymbol{a})/\partial\boldsymbol{c} &= \gamma_1 + \textstyle\sum_{p=2}^{P}\gamma_p p\,\boldsymbol{c}^{p-1} + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}(p-r)\boldsymbol{a}^r\boldsymbol{c}^{p-r-1}.
\end{aligned}$$

Conditioning only on the variable on which we differentiate, (20) and (21) give

$$\text{(25)} \qquad \begin{aligned}
\partial\mathsf{E}(y|\boldsymbol{a})/\partial\boldsymbol{a} &= \beta_1 + \textstyle\sum_{p=2}^{P}\beta_p p\,\boldsymbol{a}^{p-1} + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}r\boldsymbol{a}^{r-1}\boldsymbol{\mu}_{c|a}(p-r), \\
\partial\mathsf{E}(y|\boldsymbol{c})/\partial\boldsymbol{c} &= \gamma_1 + \textstyle\sum_{p=2}^{P}\gamma_p p\,\boldsymbol{c}^{p-1} + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}\boldsymbol{\mu}_{a|c}(r)(p-r)\boldsymbol{c}^{p-r-1}.
\end{aligned}$$

There are notable differences between (24) and (25) on the one hand and (23) on the other, since in the former, the second-order coefficients β_2 and γ_2 always occur, except when the derivatives are evaluated at the expected cohort and age ($\boldsymbol{a} = \boldsymbol{c} = 0$).

---

[5]Obviously, $\boldsymbol{\mu}_a(1) = \boldsymbol{\mu}_{ac}(1,0) = \boldsymbol{\mu}_c(1) = \boldsymbol{\mu}_{ac}(0,1) = 0$, $\boldsymbol{\mu}_{ac}(p,0) = \boldsymbol{\mu}_a(p)$, and $\boldsymbol{\mu}_{ac}(0,q) = \boldsymbol{\mu}_c(q)$.

[6]For simplicity we do not change the coefficient notation here. Expressions corresponding to (11) can be obtained by substituting $\delta_{pr} = \binom{p}{r}\delta_p$ in the following expressions.

[7]These expressions are obtained by first writing (19) as $y = \mathsf{E}(y|\boldsymbol{a},\boldsymbol{c}) + u$, where $\mathsf{E}(u|\boldsymbol{a},\boldsymbol{c}) = 0$, and next using $\partial u/\partial\boldsymbol{a} = \partial u/\partial\boldsymbol{c} = 0 \Longrightarrow \partial y/\partial\boldsymbol{a} = \partial\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c})/\partial\boldsymbol{a},\ \partial y/\partial\boldsymbol{c} = \partial\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c})/\partial\boldsymbol{c}$.

5 Application: Sickness absence

In this section we explore aspects of sickness absence, measured in days, by exploiting a large panel data set for long-term sickness absence records from individuals in the Norwegian labour force. Different specifications of heterogeneity, notably with respect to gender differences, are considered. Covariates other than the ACT variables and gender, which of course also may influence observed absenteeism (and which to some extent are observable) are neglected in the application to follow. This means that, for example variables related to work-place, lifestyle, education, family situation, geographic region, working career, health performance, doctor's practice in issuing sickness certificates, etc., will, most likely, affect the coefficient estimates of the ACT and the transformations of them we consider, to the extent that they are correlated with these 'omitted variables'. Hence, the policy implications of the results are not obvious and may be an issue for discussion.

The discussion is organized in three subsections. First, data and summary statistics are presented, next follows a description of the model hierarchy, and third, OLS regression results for the linear models, the additive polynomials and the full polynomials of orders up to four are discussed.

Data and descriptive statistics:

The data set available has zero entries for sickness absences of length less than 16 days – for the following reason. Most Norwegian workers enjoy full coverage of lost earnings due to sickness absence for up to one year. For the first 16 days of absence the payment is covered by the employer; after that the Social Security Administration (SSA) provides the payment. Only the number of days of long-term sickness absence, *i.e.*, the absence spells paid for by the SSA for each worker in each year, is counted. The lowest number of absence days observed therefore is 16. Unlike the definitions used in Biørn *et al.* (2013), sickness absence are, for part-time workers, measured in *full-time equivalents*. Also the number of absence days recorded in a year refers to absence spells *starting in that year and possibly extending to the next year*.[8] The full panel data set, which also includes individuals with no SSA-paid sickness absence, is unbalanced, covers 14 years, 1994–2007, and contains 40 592 638 observations from 3 622 170 individuals. This gives an average of 11.2 observations per individual, virtually the same for males and females.

Tables 1 and 2 give summary statistics, for the full panel, for the panel truncated to contain only individuals and periods in which where a non-zero absence is recorded, and for the sub-panels containing the two genders separately. The individuals in the full panel have, on average, 12.6 absence days, 10.7 for males and 14.6 for females (Table 1, column 1). For less than half of the individuals, 1 786 105, at least one sickness absence of at least 16 days is recorded during the 14-year data period. The mean number of absence days in the truncated panel is 112.7 (Table 1, column 5). Fewer absence spells are recorded for males than for females (1.9 million against 2.6 million), while for males the spells are on average longer (113.7 days against 111.9 days). The truncated, unbalanced

[8] For more details on definitions and institutional setting otherwise, see Biørn *et al.* (2013).

data set, with 4 502 991 observations, 1 925 320 from males and 2 577 671 from females, is the one to be used in the regressions later to be presented.

Some statistics describe the unbalance: (i) 61% of the (non-truncated) individuals are observed in all the 14 years; the remaining 39 % are distributed fairly evenly by the number of observations (Table 2, columns 1 and 2). (ii) About 76% of the non-truncated observations and about 86 % of the truncated observations come from individuals observed in all the 14 years. (iii) The mean absence length (after truncation) declines from 151.3 days for those observed in one year only, to 110.5 days for those observed in all the 14 years (Table 2, column 4). The more strongly an individual, for some reason, tends to be absent from work due to sickness – which, for lack of a better term may be labeled 'latent sickness inclination' or 'weakness of health' – the larger is his/her probability to stay permanently sick, to exit from the labour force and therefore to exit from our panel. This is a *systematic selection* which may partly explain the systematic difference between the two shares in (ii) and the two means in (iii). A discussion of some related issues is given in Biørn (2010).

The year and cohort variables from which Table 1 is compiled, and used in the following regressions, are *measured from the year 1920*, giving the variables `yea` and `coh`. Their ranges extend from 74 to 87 (calendar years 1994 and 2007) and from 5 to 71 (birth years 1925 and 1991), respectively. The age variable, `age(=yea-coh)`, varies from 16 to 69. The supplementary Appendix Table A.2, contains overall, within individual and between individual standard deviations for the non-truncated and the truncated data set. While the between variation of `abs` is far smaller than the within variation in the non-truncated data set, they have more equal magnitude in the truncated data set.

Although the data set has a large number of observations, after truncation it is 'thin' along the year dimension – the individual time series have a substantial number of gaps. On average, only 2.5 observations per individual, 2.3 for males, 2.7 for females, are available. This substantial spatial/temporal 'imbalance' – the truncated data set is not far from a set of non-overlapping cross-sections – does not invite extensive application of 'panel data methods'. However, in Section 6 supplementary results with 'fixed effects' and 'random effects' modeling of individual effects will considered, to illustrate the sensitivity of the estimated time and age effects to the way unobserved heterogeneity is accounted for.

Correlation matrices for (`abs,coh,yea,age`) and the female dummy, `fdum`, are given in Table 3, for the full and the truncated data (panel A) and by gender separately (panels B and C). As expected, `abs` shows positive correlation with `age` and negative correlation with `coh`. The omission of recorded zero absence spells results in a stronger correlation across the truncated data set than across the full data set (correlation coefficients 0.0456 and −0.0376 in the latter, 0.1123 and −0.1004 in the former), which is quite reasonable. The female dummy `fdum` is positively correlated with `abs` across the full data set and weakly negatively correlated across the truncated data set, which is consistent with the gender-specific means in Table 1.

The correlation between (`coh,yea,age`) and `fdum` is weak, the latter is 'almost orthogonal to' the former, and changes sign when the data set is truncated. Considering the way the data set has been designed – `coh` spanning 66 years, `age` spanning 53 years and `yea` spanning only 13 years – strong negative correlation between `age` and `coh` is expected: -0.9630 and -0.9509 in the full and the truncated panel, respectively.[9] Turning to the gender-specific matrices (panels B and C), we find notably stronger correlation between `abs` and (`coh,age`) for males than for females.

Since polynomial regressions is a main concern, correlation coefficients for the untransformed variable are, in Table A.1, supplemented with correlation coefficients for the powers of the (demeaned) ACT variables. Sickness absence `abs` is positively correlated with all powers of (demeaned) age (panel B, column 1) and the female dummy `fdum` is negatively correlated with all powers of (demeaned) age (panel B, column 2). For (demeaned) cohort, however, sign shifts occur: Its odd-numbered powers are negatively correlation and its even-numbered powers positively correlated with `abs`. When it comes to correlation between `fdum` and powers of cohort the odd-numbered powers show positive correlation and the even-numbered powers show negative correlation (Table A.1, panel B, column 2). Table A.1, panel A, giving the full correlation matrix of the first-through fourth powers of all three ACT variables, supplements this picture: The second and fourth powers of age and cohort show all strong positive correlation. The correlations between a and a^3 and between c and c^3 are (unsurprisingly) strongly positive, while while the corresponding cross-correlations are strongly negative. On the other hand, the correlation between even and odd powers of these two variables is rather weak and sometimes negative: for example -0.0304 between a and a^2, -0.0325 between c and c^2 and -0.0583 between c^4 and a^3. This reflects, of course, that the observations are demeaned. Correlation coefficients exceeding 0.7 in absolute value, are given in boldface in this Appendix table. A clear pattern emerges.

Table 1: DESCRIPTIVE STATISTICS

Variable	*All observations* Mean	St.dev.	*Obs. with* `abs` > 16 Mean	St.dev.
`abs`	12.61670	51.14011	112.6820	110.9083
Males	10.73485	47.81207	113.7135	112.7852
Females	14.55141	54.28121	111.9116	107.4792
`coh`	39.77881	14.95959	38.71440	12.53668
`yea`	80.59854	4.03805	80.92159	3.88333
`age`	40.81973	14.58116	42.20718	12.06349
No. obs.	40 592 638		4 502 991	
Males	20 577 392		1 925 320	
Females	20 015 246		2 577 671	
No. ind.	3 622 170		1 786 105	
No. obs/ind.	11.207		2.521	
Males	11.214		2.332	
Females	11.199		2.684	

[9] If the data set had been from a cross-section, the `coh-age` correlation would have been -1; confer Case 2 in Section 2.

Table 2: Panel characteristics according to no. of observations

No. of observations before truncation	Before truncation Inds.	Before truncation Obs.	After truncation Obs.	After truncation mean `abs`
01	120 113	120 113	2 654	151.28
02	120 103	240 206	5 575	148.49
03	113 028	339 084	8 868	134.80
04	111 573	446 292	14 691	132.00
05	108 455	542 275	22 587	129.27
06	108 799	652 794	33 099	129.52
07	104 863	734 041	41 035	130.79
08	103 394	827 152	50 854	129.88
09	102 317	920 853	63 160	128.00
10	98 232	982 320	68 950	121.33
11	107 442	1 181 862	93 327	124.82
12	112 604	1 351 248	113 886	122.67
13	130 306	1 339 780	115 884	123.75
14	2 208 187	30 914 618	3 868 421	110.50

Table 3: Correlation matrices

A. Both genders

	All 40 592 638 observations					Only 4 502 991 obs. with `abs` > 16				
	`abs`	`coh`	`yea`	`age`	`fdum`	`abs`	`coh`	`yea`	`age`	`fdum`
`abs`	1.0000					1.0000				
`coh`	-0.0376	1.0000				-0.1004	1.0000			
`yea`	0.0251	0.2275	1.0000			0.0247	0.2744	1.0000		
`age`	0.0456	-0.9630	0.0435	1.0000		0.1123	-0.9509	0.0367	1.0000	
`fdum`	0.0373	-0.0123	-0.0008	0.0124	1.0000	-0.0080	0.0580	0.0220	-0.0532	1.0000

B. Males

	All 20 577 392 observations				Only 1 925 320 obs. with `abs` > 16			
	`abs`	`coh`	`yea`	`age`	`abs`	`coh`	`yea`	`age`
`abs`	1.0000				1.0000			
`coh`	-0.0535	1.0000			-0.1177	1.0000		
`yea`	0.0167	0.2259	1.0000		0.0228	0.2639	1.0000	
`age`	0.0595	-0.9626	0.0466	1.0000	0.1291	-0.9524	0.0427	1.0000

C. Females

	All 20 015 246 observations				Only 2 577 671 obs. with `abs` > 16			
	`abs`	`coh`	`yea`	`age`	`abs`	`coh`	`yea`	`age`
`abs`	1.0000				1.0000			
`coh`	-0.0228	1.0000			-0.0863	1.0000		
`yea`	0.0328	0.2291	1.0000		0.0265	0.2812	1.0000	
`age`	0.0324	-0.9633	0.0404	1.0000	0.0985	-0.9494	0.0343	1.0000

Table 4: Estimated models. Overview

Model label ($d.k$)	*Polynomial order*	*Regressors:* Linear terms	Power terms	Interaction terms	*No. of coef.* (incl. intercept)
1.1	1	c, a			3
1.2	1	c, t			3
1.3	1	t, a			3
2.0	2	c, a	c, t, a		6
2.1	2	c, a	c, a		5
2.2	2	c, t	c, t		5
2.3	2	t, a	t, a		5
3.0	3	c, a	c, t, a		9
3.1	3	c, a	c, a		7
3.2	3	c, t	c, t		7
3.3	3	t, a	t, a		7
4.0	4	c, a	c, t, a		12
4.1	4	c, a	c, a		9
4.2	4	c, t	c, t		9
4.3	4	t, a	t, a		9
2.4	2	c, a	c, a	ca	6
3.4	3	c, a	c, a	ca, ca^2, c^2a	10
4.4	4	c, a	c, a	$ca, ca^2, c^2a, c^2a^2, ca^3, c^3a$	15

Model tree

Table 4 lists 18 models of orders 1 through 4, all including only the ACT variables. For convenience, they are labeled as $d.k$, where d and k indicate, respectively, the polynomial order and the collection of power terms (when $k=1,2,3$) and interaction terms (when $k=4$). From now on c,t,a will denote *demeaned* variables. The model-tree can be described as follows. The linear models, 1.1, 1.2 and 1.3, are equivalent, which exemplifies the ACT identification problem (Section 2). Models $2.k$, $3.k$ and $4.k$ $(k = 1,2,3)$ include linear and power terms in two of the three variables and have 5, 7, and 9 coefficients (including intercept), respectively. Models 2.0, 3.0, and 4.0, with 6, 9, and 12 coefficients, respectively, include linear terms in (a,t) and powers in (a,t,c). They exemplify (11), reparametrized from (8), see (14). Models 2.4, 3.4, and 4.4 extend the additive polynomial Models $2.k$, $3.k$, and $4.k$ $(k = 1,2,3)$, by adding interaction terms to the power terms. This extension exemplifies (17) and increases the number of coefficients to 6, 10, and 15, respectively.[10]

While Model 2.4 reparametrizes Model 2.0, Model 3.0 imposes one coefficient restriction on Model 3.4, and Model 4.0 imposes three restrictions on Model 4.4; see the example with $K=4$ in Section 3. Models $2.k\,(k = 1,2,3)$ are nested within Model 2.0, Models $3.k\,(k = 1,2,3)$ are nested within Model 3.0, and Models $4.k\,(k = 1,2,3)$ are nested within Model 4.0, while Models $d.1, d.2, d.3$ $(d = 2,3,4)$ are non-nested.

Table 5: Estimated models. OLS fit statistics. Observations with abs>16 only.

	Both genders:			Males:			Females:		
Model	SSR$\times 10^{-14}$	$\sigma_u \times 10^{-4}$	R^2	SSR$\times 10^{-14}$	$\sigma_u \times 10^{-4}$	R^2	SSR$\times 10^{-14}$	$\sigma_u \times 10^{-4}$	R^2
1.1	5.4667	1.1018	0.013041	2.4075	1.1182	0.016968	3.0579	1.0892	0.010243
1.2	5.4667	1.1018	0.013041	2.4075	1.1182	0.016968	3.0579	1.0892	0.010243
1.3	5.4667	1.1018	0.013041	2.4075	1.1182	0.016968	3.0579	1.0892	0.010243
2.0	5.4385	1.0990	0.018146	2.3936	1.1150	0.022670	3.0437	1.0867	0.014818
2.1	5.4532	1.1005	0.015480	2.4010	1.1167	0.019644	3.0512	1.0880	0.012401
2.2	5.4387	1.0990	0.018099	2.3939	1.1151	0.022558	3.0438	1.0867	0.014801
2.3	5.4414	1.0993	0.017612	2.3946	1.1152	0.022265	3.0456	1.0870	0.014218
3.0	5.4306	1.0982	0.019567	2.3898	1.1141	0.024194	3.0398	1.0859	0.016096
3.1	5.4457	1.0997	0.016848	2.3975	1.1159	0.021053	3.0474	1.0873	0.013639
3.2	5.4331	1.0984	0.019111	2.3910	1.1144	0.023709	3.0411	1.0862	0.015673
3.3	5.4338	1.0985	0.018989	2.3910	1.1144	0.023706	3.0398	1.0863	0.015472
4.0	5.4279	1.0979	0.020049	2.3877	1.1136	0.025074	3.0389	1.0858	0.016371
4.1	5.4438	1.0995	0.017188	2.3957	1.1155	0.021800	3.0470	1.0872	0.013769
4.2	5.4311	1.0982	0.019474	2.3897	1.1141	0.024237	3.0403	1.0860	0.015933
4.3	5.4314	1.0983	0.019427	2.3891	1.1139	0.024511	3.0409	1.0862	0.015726
2.4	5.4385	1.0990	0.018146	2.3936	1.1150	0.022670	3.0437	1.0867	0.014818
3.4	5.4304	1.0982	0.019602	2.3898	1.1141	0.024213	3.0397	1.0859	0.016139
4.4	5.4276	1.0979	0.020104	2.3876	1.1136	0.025111	3.0387	1.0858	0.016451

[10] The equivalent models (12) and (13) are not further discussed. Restricting attention to (11) in estimation, has the advantage of involving no sign-shifts for the binomial coefficients.

Table 6: CORRELATION COEFFICIENTS. `fdum` VERSUS POWERS OF COHORT, YEAR AND AGE

Observations with `abs` > 16 *only*

p	corr(fdum, c^p)	corr(fdum, t^p)	corr(fdum, a^p)
1	0.0580	0.0220	-0.0532
2	-0.0240	0.0035	-0.0315
3	0.0439	0.0212	-0.0386
4	-0.0305	0.0006	-0.0407

Table 7: COEFFICIENT OF `fdum` IN MODELS $d.k$ (d=2, 3, 4; k=0, 1, 2, 3, 4)

Standard errors below coefficient estimates. All coefficients multiplied by 100

Observations with `abs` > 16 *only*

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$d = 2$	-20.055299	-28.321539	-23.107976	-17.953944	-20.055299
	10.491141	10.505103	10.489306	10.493907	10.491141
$d = 3$	-35.414764	-43.605999	-32.355308	-33.752991	-36.025614
	10.485608	10.499778	10.484916	10.488588	10.485528
$d = 4$	-47.195714	-57.309748	-38.015668	-44.262887	-47.751581
	10.489016	10.503840	10.485522	10.491803	10.488848

OLS ESTIMATION RESULTS

Goodness of fit: Table 5 reports fit statistics for OLS estimation based on the truncated data set for all observations and by gender: sum of squared residuals (SSR), standard error of regression (σ_u) and squared multiple correlation (R^2). Using the two-gender panel, we obtain a fit, measured by the σ_u estimate, of about 1.1×10^{-4} in all the 18 models. When measured by R^2, the fit varies between 0.013 and 0.020. Even for Model 4.4, the most parameter-rich model, the unexplained variation accounts for a large share of the total variation. All models have lower σ_u estimates when using the female data than when using the male data. On the other hand, R^2 is *higher* for males (between 0.017 and 0.025) than for females (between 0.010 and 0.016). The latter may reflect the larger number of female long-term absence spells as well as the fact that females may tend to have more 'diverse' sickness absence patterns, less adequately captured by the ACT variables, than males.

An interesting observation if that the fit, measured by R^2, has about the same magnitude as the McFadden R-square fit measure, obtained from the discrete response (sick/non-sick) analysis of Biørn *et al.* (2013, Table 2), where a substantially larger number of parameters is, for both genders, used to capture the variation of sickness absence (541 and 57 in the model versions with cohort-specific and cohort-invariant time and age coefficients, respectively). Like the present model, the discrete response model gave a better fit to the male observations than to the female observations.

Among the models with linear and power terms in two of the three variables, those including (c, a) (Models 2.1, 3.1, and 4.1) give somewhat better fit, for both genders, than the corresponding models in (c, t) (Models 2.2, 3.2, and 4.2) or in (t, a) (Models 2.3, 3.3, and 4.3). The improvement in fit, indicated by a reduced SSR, when the regressor set includes both second, third, and fourth powers *of all the three variables* – *i.e.*, including powers of the variable which are omitted from the equation's linear part to escape the ACT problem – is clearly significant: The p-values of the F-tests for Model 2.1 against 2.0, for Model 3.1 against 3.0, and for Model 4.1 against 4.0 are all close to zero. The small increase in the respective R^2s, less than 0.003 (confer Table 5), is in the F-statistics, 'compensated' by the large number of observations, leading to a clear rejection of the restrictive model.[11]

Table 8: Linear models. OLS estimates

Standard errors below coefficient estimates. All coefficients multiplied by 100.
Observations with `abs` > 16 *only*

	Both genders. No female dummy: Model 1.1	Model 1.2	Model 1.3	*Both genders. Female dummy included:* Model 1.1	Model 1.2	Model 1.3
c	58.788114	-102.576460		58.963563	-102.448974	
	1.337988	0.430708		1.338368	0.431332	
t		161.364574	58.788114		161.412537	58.963563
		1.390470	1.337988		1.390493	1.338368
a	161.364574		102.576460	161.412537		102.448974
	1.390470		0.430708	1.390493		0.431332
`fdum`				-57.498539	-57.498539	-57.498539
				10.513268	10.513268	10.513268
	Males, 1 925 320 observations: Model 1.1	Model 1.2	Model 1.3	*Females, 2 577 671 observations:* Model 1.1	Model 1.2	Model 1.3
c	50.316711	-118.110883		65.108447	-90.029052	
	2.081852	0.658029		1.745791	0.571107	
t		168.427583	50.316711		155.137499	65.108447
		2.156397	2.081852		1.818128	1.745791
a	168.427594		118.110883	155.137499		90.029052
	2.156397		0.658029	1.818128		0.571107

Coefficient estimates: Tables 8–11 contain coefficient estimates for the 18 polynomial models.[12] Table 8 relate to linear models, Tables 9 and 10 relate to additive polynomial models, and Table 11 gives results for cubic and fourth-order models in cohort and age, with all interactions included.

Linear models: Estimates: Inclusion of the female dummy has a negligible effects on the coefficient estimates of (`coh,yea,age`) (Table 8, upper half), which reflects that `fdum` is 'almost orthogonal to' these variables (Table 3, panel A). For given `coh,yea,age`, each absence spell is about 0.57 days *shorter* for females than for males, with a p-value close to zero (Table 8, right upper part).[13] *Controlling for cohort*, we find that a one year increase in age (equivalent to a one year increase in calendar time) gives an estimated increase in (long-term) absence of 1.61 days (Table 8, upper half). *Controlling for calendar year*, while increasing birth-year by one (equivalent to being one year younger) gives an estimated reduction of absence of 1.03 days. Equivalently, *controlling for age*, while increasing birth-year by one (equivalent to increasing calendar time by one year) gives an estimated increase in absence of 0.59 days.

[11] The conclusion of rejection is also indicated from Tables 9 and 10 by the t-statistics of t^2 in Model 2.0, the t-statistics of t^2 and t^3 in Model 3.0 and the t-statistics of t^2, t^3 and t^4 in Model 4.0.

[12] The Stata software, version 12, is used in the computations.

[13] As remarked, the spells are measured in such a way that a spell starting in year t may well extend to year $t-1$.

Linear models: Gender effects: Notable gender differences emerge (Table 8, panel B). The age effect, controlling for cohort, is 1.68 days for males and 1.55 days for females (strictly, these are *age plus year effects*; see (5) and (7)). When controlling for calendar year, the age effect is 1.18 days for males and 0.90 days for females (strictly, these are *age minus cohort effects*; see (5) and (7)). The cohort effect, controlling for age, is 0.50 days for males and 0.65 days for females (strictly, these are *cohort plus year effects*; see (5) and (7)). Controlling for calendar year, the cohort effect is −1.18 days for males and −0.90 days for females (strictly, these are *cohort minus age effects*; see (5) and (7)).

Non-linear models: Estimates: For the quadratic, cubic, and fourth-order polynomial regressions, Table 9 (combined truncated panel) and Table 10 (gender-specific estimates) show that the marginal cohort and age effects at the empirical mean – the empirical counterpart to γ_1 and β_1 in (24) at the expected age and cohort ($\boldsymbol{a} = \boldsymbol{c} = 0$) – are *not invariant to* the assumed polynomial order. A certain pattern is visible though: The estimated γ_1 and β_1 from the *quadratic* Model 2.1 are close to their estimates from the *linear* Model 1.1 (in both year is omitted as a regressor): $(0.61, 1.61)$ days versus $(0.59, 1.61)$ days when using the full (truncated) panel, $(0.55, 1.69)$ versus $(0.50, 1.68)$ days for the male panel and $(0.64, 1.54)$ versus $(0.65, 1.55)$ days for the female panel. This finding may be interpreted as an empirical counterpart to (23), which implies that γ_1 and β_1 measure equally well the marginal cohort and age effects for $P = 1$ and $P = 2$. Contrasting, however, Model 2.2 with 1.2 (age omitted) and Model 2.3 with 1.3 (cohort omitted), larger discrepancies emerge. Note also that the estimates of γ_1 and β_1 from the *fourth-order* Model 4.1 are close to those from the *cubic* Model 3.1: $(0.86, 1.21)$ versus $(0.84, 1.20)$ for the full panel, $(0.89, 1.32)$ versus $(0.85, 1.34)$ for the male panel and $(0.82, 1.12)$ versus $(0.83, 1.11)$ for the female panel, respectively. On the other hand, contrasting Model 4.2 with 3.2 and Model 4.3 with 3.3, larger discrepancies again emerge. The results for the third and fourth order polynomials as well as the discrepancies between the γ_1 and β_1 coefficients in Models 3.1 and 2.1 cannot be easily explained from the expressions for the expected marginal effects, (23) or (25), however.

Non-linear models: Gender effects: While in all polynomial models the fit is slightly improved when the female dummy is added to the ACT regressors (Table 5), the coefficient estimates of the power and interaction terms are rather insensitive to whether this dummy is included or not. This reflects the weak correlation between the female dummy and powers of the demeaned ACT variables, as shown in Table 6: `fdum` is positively correlated with all powers of t and negatively correlated with all powers of a, while for c, `fdum` is positively correlated with its odd-numbered powers and negatively correlated with its even-numbered powers. The 'female effect' is somewhat sensitive to the polynomial order chosen, however. The lowest estimates are obtained for the second-order polynomials ($d = 2$), while the highest estimates occur for the fourth-order polynomials ($d = 4$) (Table 7). In, *e.g.*, Model 2.3, `fdum` is only 'marginally significant' at the 5% level (t-value around 1.8). For all polynomial orders the models which include cohort and age (Models 2.1, 3.1 and 4.1) have the largest coefficient of `fdum` in absolute value.

Table 10 shows that the sign of the coefficients of the linear and the quadratic terms in the additive polynomials are the same for males and females. The same holds for the models which include interactions; see Table 11. For the cubic and fourth-order terms, however, there are some notable gender differences: (i) In the cubic additive models where t^3 is included, this variable has negative coefficient for males and positive coefficient for females. The same is true for the fourth-order additive models. In Model 4.0 the coefficients of c^4 also come out with different signs. (ii) In the non-additive Model 4.4, the coefficients of all the fourth-order terms, $a^4, a^3c, a^2c^2, ac^3, c^4$, are positive, while in Model 3.4, the coefficients of all the cubic terms, a^3, a^2c, ac^2, c^3, are negative for males and positive for females.

Curvature: The quadratic model: The quadratic *Model 2.0* can be written in several forms. Let us take a closer look at its estimated regression Table 9. We have

$$\begin{aligned}
(26)\qquad &\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c},\boldsymbol{t})} = \text{constant} + 42.380\,\boldsymbol{c} + 144.231\,\boldsymbol{a} + 2.715\,\boldsymbol{c}^2 - 40.724\,\boldsymbol{t}^2 + 0.947\,\boldsymbol{a}^2,\\
&\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c},\boldsymbol{t})} = \text{constant} - 101.851\,\boldsymbol{c} + 144.231\,\boldsymbol{t} + 2.715\,\boldsymbol{c}^2 - 40.724\,\boldsymbol{t}^2 + 0.947\,\boldsymbol{a}^2,\\
&\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c},\boldsymbol{t})} = \text{constant} + 42.380\,\boldsymbol{t} + 101.851\,\boldsymbol{a} + 2.715\,\boldsymbol{c}^2 - 40.724\,\boldsymbol{t}^2 + 0.947\,\boldsymbol{a}^2.
\end{aligned}$$

By manipulating the second-order terms, eliminating, respectively, $\boldsymbol{t}^2$, $\boldsymbol{a}^2$ and $\boldsymbol{c}^2$, which creates interactions between the two remaining variables, we get

$$\begin{aligned}
(27)\qquad &\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c},\boldsymbol{a}+\boldsymbol{c})} = \text{constant} + 42.380\,\boldsymbol{c} + 144.231\boldsymbol{a} - 38.009\,\boldsymbol{c}^2 - 81.448\,\boldsymbol{ac} - 39.777\,\boldsymbol{a}^2,\\
&\widehat{\mathsf{E}(y|\boldsymbol{t}-\boldsymbol{c},\boldsymbol{c},\boldsymbol{t})} = \text{constant} - 101.851\,\boldsymbol{c} + 144.231\boldsymbol{t} + 3.662\,\boldsymbol{c}^2 - 1.894\,\boldsymbol{ct} - 39.777\,\boldsymbol{t}^2,\\
&\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{t}-\boldsymbol{a},\boldsymbol{t})} = \text{constant} + 42.380\,\boldsymbol{t} + 101.851\boldsymbol{a} - 38.009\,\boldsymbol{t}^2 - 5.430\,\boldsymbol{ta} + 3.662\,\boldsymbol{a}^2.
\end{aligned}$$

Since neither of the regressions in (27) has a Hessian matrix that is positive or negative definite, neither are convex or concave *in their two variables.*[14] However, *controlling for one variable*, the curvature of the other (around mean) can be described as follows:[15]

$\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c},\boldsymbol{a}+\boldsymbol{c})}$: Positively sloping and concave in $\boldsymbol{c}$, when $\boldsymbol{a}$ is controlled for:

$$m_{c|a} \equiv \partial\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c},\boldsymbol{a}+\boldsymbol{c})}/\partial\boldsymbol{c} = 42.4 - 76.0\boldsymbol{c} - 81.4\boldsymbol{a}.$$

Positively sloping and concave in $\boldsymbol{a}$, when $\boldsymbol{c}$ is controlled for:

$$m_{a|c} \equiv \partial\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c},\boldsymbol{a}+\boldsymbol{c})}/\partial\boldsymbol{a} = 144.2 - 79.5\boldsymbol{a} - 81.4\boldsymbol{c}.$$

Strictly, $m_{c|a}$ and $m_{a|c}$ are *marginal cohort+year and age+year effects*; cf. (2) and (7).

$\widehat{\mathsf{E}(y|\boldsymbol{t}-\boldsymbol{c},\boldsymbol{c},\boldsymbol{t})}$: Negatively sloping and convex in $\boldsymbol{c}$, when $\boldsymbol{t}$ is controlled for:

$$m_{c|t} \equiv \partial\widehat{\mathsf{E}(y|\boldsymbol{t}-\boldsymbol{c},\boldsymbol{c},\boldsymbol{t})}/\partial\boldsymbol{c} = -101.9 + 7.3\boldsymbol{c} - 1.9\boldsymbol{t} \equiv -101.9 + 5.4\boldsymbol{c} - 1.9\boldsymbol{a}.$$

Positively sloping and concave in $\boldsymbol{t}$, when $\boldsymbol{c}$ is controlled for:

$$m_{t|c} \equiv \partial\widehat{\mathsf{E}(y|\boldsymbol{t}-\boldsymbol{c},\boldsymbol{c},\boldsymbol{t})}/\partial\boldsymbol{t} = 144.2 - 79.5\boldsymbol{t} - 1.9\boldsymbol{c} \equiv 144.2 - 79.5\boldsymbol{a} - 81.4\boldsymbol{c}.$$

Strictly, $m_{c|t}$ and $m_{t|c}$ are *marginal cohort−age and year+age effects*; cf. (2) and (7).

$\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{t}-\boldsymbol{a},\boldsymbol{t})}$: Positively sloping and convex in $\boldsymbol{a}$, when $\boldsymbol{t}$ is controlled for:

$$m_{a|t} \equiv \partial\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{t}-\boldsymbol{a},\boldsymbol{t})}/\partial\boldsymbol{a} = 101.9 + 7.3\boldsymbol{a} - 5.4\boldsymbol{t} \equiv 101.9 + 1.9\boldsymbol{a} - 5.4\boldsymbol{c}.$$

Positively sloping and concave in $\boldsymbol{t}$, when $\boldsymbol{a}$ is controlled for:

$$m_{t|a} \equiv \partial\widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{t}-\boldsymbol{a},\boldsymbol{t})}/\partial\boldsymbol{t} = 42.4 - 76.0\boldsymbol{t} - 5.4\boldsymbol{a} \equiv 42.4 - 76.0\boldsymbol{c} - 81.4\boldsymbol{a}.$$

Strictly, $m_{a|t}$ and $m_{t|a}$ are *marginal age−cohort and year+cohort effects*; cf. (2) and (7).

[14] Recall that the sample mean corresponds to $\boldsymbol{a}=\boldsymbol{c}=0$, and that the coefficients of the quadratic terms are invariant to changing the origins from which the variables are measured from origo to the respective sample means.

[15] Note: $m_{c|a} \equiv m_{t|a}$ (cohort+year effect), $m_{a|c} \equiv m_{t|c}$ (age+year effect), $m_{c|t} \equiv -m_{a|t}$ (cohort−age/age−cohort effect).

The *interaction terms* in the three versions of (27) are clearly of different importance. Omitting the cohort-year (ct) interaction from the second equation and the year-age (ta) interaction from the third equation, we get equations whose coefficients are 'largely similar' to those obtained for Model 2.2 and Model 2.3 in Table 9. On the other hand, *the cohort-age* (ca) *interaction is important*, with a standard error of its estimate (81.4) of 0.7. This is consistent with Table 5, where columns 1–3 show that Models 2.0, 2.2 and 2.3 have approximately the same fit (although, as remarked, the R^2 of the former is significantly larger than for the two latter, according to F-tests, which reflects the very large sample size), which is markedly better than the fit of Model 2.1. A message from our data is thus that, with respect to fit, an additive quadratic model in cohort and age is inferior. Columns 1 and 2 of Table 11 show that cohort-age interactions are important also for cubic and fourth-order models: the coefficient estimates of ac, a^2c, ac^2 are all significantly non-zero in both Model 3.4 and 4.4, while the coefficient estimates of the fourth-order terms, a^3c, a^2c^2, ac^3, are also significant in Model 4.4.

Replicating the above calculations of marginal effects, using the estimated gender specific quadratic polynomials in Table 10, leads to largely the same qualitative conclusions. From Table 11, columns 3 through 6, we see, however, that for the gender specific cubic and fourth-order models, the effects of the *interactions* are not so sharply determined as when using the pooled data set. An example is Model 3.4, where a^2c comes out with a statistically insignificant coefficient estimate for females. For males, the fourth-order Model 4.4, shows signs of being 'slightly overparametrized'. The facts that c^4 in the more parsimonious Model 4.0, and t^3 in the still more parsimonious Model 4.2 come out with insignificant estimates when based on such a large data set, supports this conclusion.

Table 9: Additive polynomial models, both genders. OLS estimates

Standard errors below coefficient estimates. All coefficients multiplied by 100

	Model 2.0	*Model 2.1*	*Model 2.2*	*Model 2.3*
c	42.380340	60.585173	-101.940029	
	1.345203	1.336902	0.429819	
t			144.536227	41.477363
			1.396276	1.345445
a	144.231486	161.751842		101.924793
	1.396398	1.389262		0.429966
c^2	2.715114	1.971366	3.385272	
	0.054864	0.054524	0.030284	
t^2	-40.724031		-41.220886	-38.489725
	0.368281		0.366725	0.365603
a^2	0.946568	1.604667		3.613155
	0.064619	0.064432		0.035677

	Model 3.0	*Model 3.1*	*Model 3.2*	*Model 3.3*
c	76.724597	84.154151	-57.250737	
	3.502955	1.903298	0.785178	
t			172.191865	56.729322
			3.296704	3.272278
a	112.836102	120.779528		37.426928
	3.582489	2.040247		0.919296
c^2	2.606039	1.853019	3.251259	
	0.055052	0.054712	0.030333	
t^2	-41.479595		-41.732512	-39.388832
	0.379800		0.378365	0.377246
a^2	0.947838	1.615172		3.513737
	0.064734	0.064545		0.035675
c^3	-0.046064	-0.048655	-0.134421	
	0.002895	0.002897	0.001977	
t^3	-0.430361		-0.317642	-0.517828
	0.105797		0.105793	0.105763
a^3	0.173579	0.166503		0.222032
	0.004094	0.004098		0.002798

	Model 4.0	*Model 4.1*	*Model 4.2*	*Model 4.3*
c	52.548887	86.106813	-55.494873	
	3.595653	1.905275	0.787608	
t			147.973838	31.604168
			3.392777	3.370314
a	87.286686	121.089294		36.231839
	3.673045	2.041224		0.919756
c^2	2.827748	3.561605	5.253592	
	0.117953	0.117917	0.079526	
t^2	-79.275911		-78.487626	-77.398579
	1.295017		1.295017	1.294423
a^2	4.284753	3.736192		6.736655
	0.146429	0.146550		0.105084
c^3	-0.048009	-0.055384	-0.141296	
	0.002917	0.002919	0.001991	
t^3	0.899624		0.992847	0.817227
	0.114244		0.114254	0.114216
a^3	0.178171	0.166682		0.228063
	0.004106	0.004110		0.002803
c^4	-0.000181	-0.002447	-0.003440	
	0.000162	0.000161	0.000126	
t^4	0.955973		0.957372	0.961728
	0.031188		0.031196	0.031195
a^4	-0.007641	-0.005475		-0.007123
	0.000274	0.000274		0.000218

Table 10: Additive polynomial models, by gender. OLS estimates
$n = 1\,925\,320$ male observations, $n = 2\,577\,671$ female observations

Standard errors below coefficient estimates. All coefficients multiplied by 100

	Males				Females			
	Model 2.0	*Model 2.1*	*Model 2.2*	*Model 2.3*	*Model 2.0*	*Model 2.1*	*Model 2.2*	*Model 2.3*
c	31.758144	54.771581	-115.664455		48.908553	64.057137	-90.961419	
	2.107399	2.089442	0.658028		1.752683	1.744434	0.569855	
t			149.584621	26.891806			139.502753	51.566550
			2.172239	2.100784			1.822877	1.751932
a	147.356020	169.461870		116.212245	139.808882	154.377164		90.459527
	2.177278	2.161711		0.657782	1.823456	1.816452		0.569991
c^2	2.371878	1.623362	3.422184		2.882900	2.155958	3.279112	
	0.083947	0.083514	0.045383		0.072751	0.072263	0.040810	
t^2	-44.115247		-44.845695	-42.250972	-38.288228		-38.588269	-35.890526
	0.571358		0.569276	0.567653	0.481510		0.479349	0.477839
a^2	1.466123	2.120466		3.809491	0.564440	1.210658		3.379057
	0.098584	0.098371		0.053304	0.085799	0.085518		0.048144

	Males				Females			
	Model 3.0	*Model 3.1*	*Model 3.2*	*Model 3.3*	*Model 3.0*	*Model 3.1*	*Model 3.2*	*Model 3.3*
c	108.765840	85.536541	-68.401344		52.141957	82.112596	-49.747876	
	5.421252	2.951421	1.210357		4.589224	2.493162	1.035054	
t			213.043639	79.779776			140.611404	39.608581
			5.103383	5.067528			4.316973	4.283502
a	156.749650	134.064345		50.339134	81.040013	111.414513		29.451108
	5.536130	3.153286		1.423234	4.698759	2.678112		1.208617
c^2	2.196869	1.448297	3.203897		2.827747	2.092146	3.226687	
	0.084661	0.084240	0.045602		0.072795	0.072305	0.040807	
t^2	-46.188503		-46.747857	-44.462844	-38.202607		-38.197185	-35.925127
	0.592667		0.590930	0.589332	0.494633		0.492521	0.490992
a^2	1.473527	2.123672		3.669646	0.573022	1.228847		3.329532
	0.098998	0.098782		0.053342	0.085837	0.085553		0.048121
c^3	-0.057928	-0.059146	-0.137191		-0.036372	-0.040038	-0.127723	
	0.004334	0.004338	0.002953		0.003905	0.003907	0.002678	
t^3	-1.673378		-1.532214	-1.804121	0.460897		0.556583	0.409808
	0.164161		0.164139	0.164095	0.138329		0.138328	0.138291
a^3	0.157809	0.151743		0.218089	0.178519	0.170771		0.216700
	0.006134	0.006142		0.004183	0.005518	0.005523		0.003786

	Males				Females			
	Model 4.0	*Model 4.1*	*Model 4.2*	*Model 4.3*	*Model 4.0*	*Model 4.1*	*Model 4.2*	*Model 4.3*
c	82.700016	89.380737	-64.177988		29.654602	82.868936	-49.256908	
	5.593129	2.961020	1.221103		4.694992	2.493587	1.035630	
t			187.797695	52.541112			117.955136	16.617844
			5.277807	5.244442			4.427865	4.396966
a	126.067810	132.492564		46.049813	58.443651	112.105654		29.361707
	5.702334	3.158035		1.427940	4.802604	2.678450		1.208468
c^2	2.255944	3.232719	6.062680		3.145343	3.748540	4.468093	
	0.182853	0.182692	0.121247		0.155075	0.155064	0.105791	
t^2	-84.234909		-83.218515	-82.524897	-75.243878		-74.594408	-73.233143
	2.006220		2.006638	2.005199	1.694794		1.694565	1.694010
a^2	6.924148	6.167961		8.861912	2.113462	1.680311		4.854743
	0.224401	0.224549		0.159849	0.194024	0.194193		0.140149
c^3	-0.059685	-0.067393	-0.150240		-0.037801	-0.044921	-0.130744	
	0.004382	0.004386	0.002993		0.003923	0.003925	0.002685	
t^3	-0.244280		-0.147421	-0.376444	1.701115		1.789498	1.655038
	0.178787		0.178812	0.178731	0.148461		0.148469	0.148429
a^3	0.170544	0.159813		0.232005	0.179445	0.167663		0.218822
	0.006157	0.006164		0.004199	0.005532	0.005536		0.003788
c^4	0.000201	-0.002339	-0.004799		-0.000432	-0.002533	-0.002178	
	0.000243	0.000241	0.000188		0.000219	0.000217	0.000171	
t^4	0.963787		0.969761	0.969139	0.935572		0.933612	0.939849
	0.048432		0.048451	0.048442	0.040750		0.040758	0.040760
a^4	-0.012130	-0.009735		-0.011205	-0.003664	-0.001655		-0.003444
	0.000407	0.000407		0.000325	0.000372	0.000372		0.000297

Table 11: Cubic and fourth-order polynomials in cohort and age. OLS estimates

Models 3.4 and 4.4. All admissible interactions. $n = 4\,502\,991$

Standard errors below coefficient estimates. All coefficients multiplied by 100

	Both genders		*Males*		*Females*	
	Model 3.4	*Model 4.4*	*Model 3.4*	*Model 4.4*	*Model 3.4*	*Model 4.4*
c	80.688492	56.505550	111.785778	87.713282	56.339426	31.705002
	3.516609	3.610363	5.443383	5.624745	4.606265	4.716606
a	122.216542	95.949950	163.704856	134.194692	91.221186	65.825553
	3.656929	3.745537	5.650399	5.815533	4.796330	4.904079
c^2	-38.845045	-76.273126	-43.737948	-80.109273	-35.594895	-73.248320
	0.377275	1.339883	0.590993	2.074577	0.491469	1.754509
ac	-83.049804	-160.563765	-91.993013	-166.122777	-76.993718	-156.201652
	0.759619	2.696105	1.186967	4.173005	0.990809	3.530790
a^2	-40.638468	-77.181161	-44.576257	-76.662045	-37.983520	-77.843211
	0.390953	1.390827	0.608975	2.150675	0.510883	1.822723
c^3	-0.434736	0.885162	-1.695345	-0.322790	0.467804	1.758356
	0.105785	0.114250	0.164174	0.179303	0.138308	0.148542
a^2c	-1.934940	2.210794	-5.478432	-1.178388	0.668550	4.747733
	0.321363	0.346461	0.498084	0.541302	0.420441	0.451723
ac^2	-1.534567	2.520752	-5.183824	-0.970427	1.108163	5.084168
	0.317957	0.343230	0.493197	0.537537	0.415791	0.446826
a^3	-0.614528	0.802044	-1.773503	-0.299884	0.244140	1.639669
	0.109428	0.117710	0.169405	0.183405	0.143284	0.153755
c^4		0.950453		0.962878		0.924521
		0.031204		0.048469		0.040766
a^3c		4.059271		4.029632		4.053828
		0.126792		0.196631		0.165793
a^2c^2		5.961552		5.968617		5.894444
		0.188265		0.292170		0.246064
ac^3		3.888539		3.918741		3.811717
		0.124868		0.193886		0.163156
a^4		1.028033		1.004982		1.042547
		0.032210		0.049907		0.042145

6 Heterogeneity attached to individual rather than to cohort

So far, heterogeneity has been treated as attached to the observable variables cohort, year and age, sometimes also to gender, through linear, quadratic, cubic or fourth-order polynomials. In this respect, the modeling of heterogeneity differs from the modeling adopted in the binary choice' analysis of Biørn *et al.* (2013), where heterogeneity is attached to the individual.[16] We will in this section, as a kind of sensitivity analysis, extend parts of the polynomial regression analysis in Section 5 – still measuring sickness absence, cohort, age and time continuously – in that direction. Unobserved *individual-specific* heterogeneity will be allowed for in two alternative ways: random and modeled (Tables 12, 15 and 16) and as fixed and non-modeled (Tables 13 and 14).

When including *fixed effects* in the linear and polynomial regressions, any transformations of cohort and the linear term in time have to be excluded. This is because the fixed effects individual dummies, to escape perfect collinearity, have to capture all explanatory variables that are time-invariant while year, equalling cohort+age, also is linearly related to the continuously measured age and the individual dummies. Modeling heterogeneity via random effects may be interpreted as supplementing the fixed cohort effects, represented parametrically by c and its powers, by an *additional* individual-specific sickness absence component. These additional components are modeled as independent

[16] For an extended discussion of censoring issues in this context, see Biørn (2010).

draws from a *normal* distribution with zero expectation and constant variance and *by construction, uncorrelated with c and any of its powers specified.*[17]

For this modified approach, two reduced-size data sets will be used. The intention of the reduction is to delete individuals with very few observations, since estimation of heterogeneity characteristics from few observations may 'disturb' also estimation of the coefficients of the polynomials. The first and largest of these reduced panels contains observations from individuals observed in *at least 5 years* (before truncation) while the second, which is a subset of the former, is the sub-panel of individuals observed in *all 14 years* (before truncation).[18]

Linear random effects models: Estimates: Results for the linear random effects models – the counterpart to Models 1.1, 1.2 and 1.3 in Table 8 – obtained by feasible Generalized Least Squares (FGLS), are given in Table 12). The coefficients of cohort, year, and age are, on the whole, markedly larger that the OLS results reported in the upper left panel of Table 8. *Controlling for cohort* and using the sub-panel *observed in at least 5 years* (Table 12, left half), we find that a one year increase in age (equivalent to a one year increase in the calendar time) gives an estimated increase in long-term absence of 2.68 days. This is about one day longer than the estimate when neglecting individual heterogeneity. *Controlling for calendar year*, and increasing birth-year by one (equivalent to being one year younger) gives an estimated reduction in long-term absence of 1.26 days. This is about 1/4 day more than the estimate when neglecting individual heterogeneity. Equivalently, controlling for age, and increasing birth-year by one (equivalent to increasing calendar time by one year) gives an estimated increase in long-term absence of 1.42 days, which is about 3/4 day longer than the estimate when neglecting individual heterogeneity. Estimation from the sub-panel *observed in 14 years* (Table 12, right half) gives somewhat larger cohort effect, 1.66 days, when controlling for calendar year, and a smaller age effect, 1.03 days, when controlling for cohort.

Non-linear random effects models: Estimates: The estimates obtained when including random effects in the quadratic, cubic and fourth-order polynomials, are collected in Table 15. The coefficient estimates of the linear and the quadratic terms are substantially magnified relative to those in Table 9. Again, the (γ_1, β_1) estimates from Model 2.1 and Model 1.1 are close: $(1.43, 2.67)$ versus $(1.43, 2.68)$ when using the sub-panel observed in at least 5 years and $(1.68, 2.67)$ versus $(1.66, 2.70)$ when using the sub-panel observed in 14 years. This again may be interpreted as an empirical counterpart to the expected marginal effects, expressed in (23) for $P=1$ and $P=2$. As was also the finding from the model with individual heterogeneity neglected, the (γ_1, β_1) estimates from Model 4.1 are close to those from Model 3.1: $(1.98, 1.59)$ versus $(1.99, 1.59)$ when using the sub-panel observed in at least 5 years and $(1.91, 1.65)$ versus $(1.92, 1.67)$ when using the sub-panel observed in 14 years. Still, there is a systematic difference between the (γ_1, β_1) estimates from Models 2.1 and 1.1 on the one hand and Models 4.1 and 3.1 on the other.

[17] For a further discussion of the distinction between 'random' and 'systematic' heterogeneity in this context, see Biørn (2010, Section 3).

[18] The latter is a balanced 14 year subpanel within the full data set which becomes unbalanced with a maximum of 14 observations of each individual when all observations with `abs`=0 have been deleted.

Table 12: Linear models with heterogeneity represented as **random** effects

FGLS estimates. Two sub-panels. Observations with `abs` > 16 only

Standard errors below coefficient estimates. All coefficients multiplied by 100

	Individuals observed at least 5 times			*Individuals observed in all 14 years*		
	Model 1.1	*Model 1.2*	*Model 1.3*	*Model 1.1*	*Model 1.2*	*Model 1.3*
c	142.6212	-125.5736		166.7959	-103.2787	
	1.3702	0.5109		1.5120	0.6214	
t		268.1948	142.6212		270.0746	166.7959
		1.3971	1.3702		1.4487	1.5120
a	268.1948		125.5736	270.0746		103.2787
	1.3971		0.5109	1.4487		0.6214

Table 13: Linear and quadratic models with heterogeneity represented as **fixed** effects

OLS estimates. Two sub-panels. Observations with `abs` > 16 only

Standard errors below coefficient estimates. All coefficients multiplied by 100

	Inds. obs. at least 5 times		*Inds. obs. in all 14 years*	
a	579.6145	578.3416	548.8104	545.9721
	1.7406	1.7370	1.7740	1.7712
a^2		8.9940		8.1642
		0.0842		0.0889

Table 14: Polynomial models in age and year with individual **fixed** effects

OLS estimates. Observations with `abs` > 16 *only*

Standard errors below coefficient estimates. All coefficients multiplied by 100

	Individuals observed in at least 5 years				Individuals observed in 14 years			
a	377.372099	377.377919	360.109748	381.696239	379.354625	380.055312	402.411870	426.194449
	2.525332	2.524786	4.390597	4.289115	2.618011	2.615502	4.611818	4.508477
a^2	8.662031	8.643476	8.751903	8.509869	8.798711	7.676804	7.312783	8.278119
	0.187251	0.084114	0.186674	0.083875	0.204524	0.088919	0.204348	0.088836
a^3	0.599213	0.599179	0.612254	0.599213	0.531269	0.527526	0.528431	0.532179
	0.005482	0.005474	0.005466	0.005482	0.006157	0.006127	0.006143	0.006113
a^4	-0.000041		-0.000518		-0.002689		0.002363	
	0.000369		0.000368		0.000441		0.000442	
t^2			-88.641588	-56.159106			-91.233783	-53.430282
			1.486173	0.447723			1.557727	0.468120
t^3			0.088218	-1.061174			-1.135111	-2.454241
			0.132516	0.122690			0.138990	0.128761
t^4			0.820367				0.951599	
			0.035797				0.037601	

Table 15: Additive polynomial models with individual **random** effects

FGLS estimates. Two sub-panel. Observations with `abs` > 16 *only*
Standard errors below coefficient estimates. All coefficients multiplied by 100

	Individuals observed in at least 5 years				Individuals observed in 14 years			
	Model 2.0	*Model 2.1*	*Model 2.2*	*Model 2.3*	*Model 2.0*	*Model 2.1*	*Model 2.2*	*Model 2.3*
c	125.113992	142.923506	-125.286022		150.302287	168.242666	-100.151922	
	1.374993	1.368335	0.509263		1.518202	1.510585	0.621195	
t			250.266019	125.908336			251.075251	150.555962
			1.401148	1.375446			1.454861	1.518178
a	250.209467	266.932440		125.454935	249.631625	266.950204		99.929247
	1.400933	1.395511		0.511080	1.454854	1.447706		0.623343
c^2	2.009051	1.379126	3.876123		1.458629	1.899043	4.416996	
	0.059117	0.058940	0.036695		0.084212	0.084164	0.060257	
t^2	-44.004294		-45.012907	-42.950756	-42.474665		-41.683506	-42.812243
	0.363092		0.362307	0.361580	0.385494		0.385356	0.384974
a^2	2.636979	3.172288		4.389779	3.584462	3.253804		4.456342
	0.065380	0.065324		0.040682	0.071161	0.071215		0.051030

	Individuals observed in at least 5 years				Individuals observed in 14 years			
	Model 3.0	*Model 3.1*	*Model 3.2*	*Model 3.3*	*Model 3.0*	*Model 3.1*	*Model 3.2*	*Model 3.3*
c	161.272255	158.618650	-69.829545		184.346354	166.770640	-47.235116	
	3.539701	2.005788	0.987517		3.950440	2.385029	1.404267	
t			284.255223	145.790241			298.536287	167.410746
			3.255563	3.231344			3.423117	3.472377
a	200.885244	198.991022		43.287048	210.860656	192.367911		34.696490
	3.558600	2.045873		1.056088	3.824070	2.157608		1.160972
c^2	1.912377	1.273346	3.731148		1.243286	1.680162	4.073588	
	0.059305	0.059128	0.036744		0.084730	0.084685	0.060810	
t^2	-45.228011		-45.792555	-44.258694	-43.840507		-42.898913	-44.186905
	0.373937		0.373409	0.372498	0.397306		0.397355	0.396772
a^2	2.620599	3.163003		4.290284	3.354001	3.019470		4.115372
	0.065454	0.065398		0.040715	0.071315	0.071371		0.051317
c^3	-0.038890	-0.042231	-0.164516		-0.061783	-0.069398	-0.239322	
	0.003332	0.003334	0.002510		0.006688	0.006687	0.005700	
t^3	-0.789528		-0.672400	-0.820273	-1.352429		-1.507776	-1.342679
	0.103860		0.103954	0.103829	0.110739		0.110773	0.110681
a^3	0.247382	0.237228		0.280694	0.252530	0.246774		0.278243
	0.004159	0.004165		0.003146	0.004881	0.004885		0.004170

	Individuals observed in at least 5 years				Individuals observed in 14 years			
	Model 4.0	*Model 4.1*	*Model 4.2*	*Model 4.3*	*Model 4.0*	*Model 4.1*	*Model 4.2*	*Model 4.3*
c	135.659699	158.872113	-68.683955		156.764178	164.822885	-47.431939	
	3.625635	2.007926	0.990071		4.038856	2.394229	1.417176	
t			260.275601	121.205340			272.388764	140.988306
			3.346407	3.323883			3.520775	3.568027
a	174.685219	198.331320		42.460716	184.480206	190.987298		34.561697
	3.643733	2.046851		1.055993	3.911465	2.160169		1.163782
c^2	0.675775	1.545318	5.067815		-0.259428	-0.574552	3.920815	
	0.137490	0.137410	0.104034		0.225933	0.225932	0.191532	
t^2	-83.297779		-82.619623	-81.594349	-84.624570		-83.620837	-85.017442
	1.266764		1.267675	1.264973	1.344082		1.344168	1.343449
a^2	6.041510	5.477589		6.926121	3.769465	4.871978		4.457734
	0.147833	0.148041		0.116097	0.167162	0.167177		0.135652
c^3	-0.036007	-0.044169	-0.169305		-0.052560	-0.057789	-0.238149	
	0.003358	0.003360	0.002528		0.006817	0.006816	0.005828	
t^3	0.526938		0.613212	0.477038	0.065922		-0.091606	0.083834
	0.112005		0.112109	0.111963	0.119373		0.119429	0.119316
a^3	0.253474	0.239343		0.285180	0.253921	0.253850		0.279121
	0.004170	0.004176		0.003149	0.004911	0.004916		0.004201
c^4	0.002054	-0.000438	-0.002366		0.004206	0.005888	0.000455	
	0.000204	0.000203	0.000173		0.000588	0.000587	0.000545	
t^4	0.949058		0.940285	0.939489	1.035212		1.032219	1.037326
	0.030521		0.030553	0.030509	0.032521		0.032550	0.032509
a^4	-0.007312	-0.005222		-0.005839	-0.000831	-0.004229		-0.000864
	0.000278	0.000278		0.000241	0.000353	0.000353		0.000323

Fixed effects results: Representing heterogeneity by fixed effects, we find for equations which are linear and quadratic in `age`, results exemplified in Table 13. Here no linear term in time is included, for reasons explained above. Results when also cubic and fourth-order terms in age as well as quadratic, cubic and fourth-order terms in time are included are given in Table 14. On the whole, the age variables obtain coefficient estimates that are substantially larger in the fixed effects models than in corresponding random effects models. In the equations with cubic and fourth-order terms in time, the effect of the squared time (negative) is substantially larger in absolute value than the corresponding equations with random effects and with no individual heterogeneity accounted for. This may capture effects on sickness absence of omitted variables that are correlated with age and time and not properly accounted for by the way we have specified heterogeneity in the previous models.

We note a substantial drop in the coefficient estimate of the age variable in the fixed effects models when cubic and fourth-order terms are added to the linear and quadratic terms. On the other hand, inclusion of higher-order terms in time has no strong impact on the estimates of the coefficients of the age variable. There are, however, some discrepancies in this respect between the two sub-panels; see the first row of Table 14.

Random and fixed effects models: Heterogeneity and goodness of fit: Table 16 gives, for 13 polynomial random effects models, the estimated ρ, *i.e.*, the variance of the individual-specific effect as a share of the 'gross disturbance' variance (the variance of the sum of the individual-specific effect and the genuine disturbance). The estimates are 18–19% when the sub-panel of individuals observed in at least 5 years is used, and slightly lower (17–18%) for the sub-panel observed in 14 years.

Overall, we may then conclude that inclusion of individual random or fixed effects gives a better fit of the polynomial models in age and time. The R^2 measures, in Table 17, exceed those of comparable models in Table 5.

Table 16: VARIANCE RATIO (ρ) ESTIMATES IN **random** EFFECTS MODELS

ρ = var(indiv. effect)/var(indiv. effect+disturbance) *Observations with* `abs` > 16

Model	Observed $\geq$ 5 years	Observed 14 years
1.1	0.1863	0.1744
2.0	0.1865	0.1747
2.1	0.1840	0.1724
2.2	0.1854	0.1735
2.3	0.1899	0.1777
3.0	0.1889	0.1764
3.1	0.1863	0.1739
3.2	0.1853	0.1734
3.3	0.1924	0.1794
4.0	0.1880	0.1763
4.1	0.1853	0.1739
4.2	0.1849	0.1734
4.3	0.1917	0.1791

Table 17: R^2 FIT INDEX. SELECTED MODELS

Observations with abs > 16. *Individuals observed* $\geq$ *5 years.*

Random *individual effects models*

Regressors	
c,a	0.0394
c,a, c^2,a^2	0.0432
c,a, c^2,t^2,a^2	0.0460
c,a, c^2,a^2, c^3,a^3	0.0473
c,a, c^2,t^2,a^2, c^3,t^3,a^3	0.0505
c,a, c^2,a^2, c^3,a^3, c^4,a^4	0.0470
c,a, c^2,t^2,a^2, c^3,t^3,a^3, c^4,t^4,a^4	0.0500
t,a, t^2,a^2	0.0470
t,a, t^2,a^2, t^3,a^3	0.0515
t,a, t^2,a^2, t^3,a^3, t^4,a^4	0.0511

Fixed *individual effects models*

Regressors	
a	0.0394
a, a^2	0.0434
a, a^2, t^2	0.0489
a, a^2, a^3	0.0476
a, a^2, a^3, t^2, t^3	0.0533
a, a^2, a^3, a^4	0.0476
a, a^2, a^3, a^4, t^2, t^3, t^4	0.0535

R^2 = fit measure 'R^2_{within}' in the Stata output

7 CONCLUSION

The conclusions from the empirical part of this study can be summarized as follows:

1. The dependence of sickness absence on age, cohort and time – all measured continuously – is clearly non-linear in all variables.

2. The coefficient estimates of terms up to (at least) order four are statistically significant at usual levels.

3. There are clear gender effects in sickness absence. Not only does the gender dummy come out with a significant coefficient estimate in the linear models, there are also notable differences between the coefficient estimates in the polynomial models.

4. The improvement in fit, when we let the regressor set includes both second, third, and fourth powers of all the three variables – including powers of the variable omitted from the linear part to escape the ACT identification problem – is clearly significant. The overall fit of the polynomial regressions, measured by R^2 and related statistics, is still poor, but not worse than the fit obtained from a more parameter rich discrete response (sick/non-sick) model with age, cohort and time represented by dummies.

5. Including interaction terms in additive polynomials improves the fit significantly. However, the interactions of the ACT variables are not of equal importance. In *e.g.* a quadratic model, the cohort-age interaction seems to be particularly important.

6. Marginal effects of cohort and age – at the sample mean – come out with approximately the same estimates from the linear and the quadratic models. This concurs with the theoretical definition expectation of first-derivatives. Also the estimates from the cubic and the quadratic models are fairly equal, but there is a marked discrepancy between between the two pairs of estimates.

7. Representing heterogeneity by individual, fixed or random, effects gives markedly larger estimates of the age effects than when attaching heterogeneity to cohort only. However, it should be recalled that, after truncation, on average, only 2.5 observations per individual, are available for models which put this kind of heterogeneity in focus.

References

Baetschmann, G. (2013): Heterogeneity in the Relationship Between Happiness and Age: Evidence from the German Socio-Economic Panel. *German Economic Review*, forthcoming.

Biørn, E. (2010): Identifying Trend and Age Effects in Sickness Absence from Individual Data: Some Econometric Problems". Memorandum No. 20/2010. Department of Economics, University of Oslo.

Biørn, E. (2013): Identifying Age-Cohort-Time Effects, Their Curvature and Interactions from Polynomials: Examples related to Sickness Absence. Memorandum No. 08/2013. Department of Economics, University of Oslo.

Biørn, E., Gaure, S., Markussen, S., and Røed, K. (2013): The Rise in Absenteeism: Disentangling the Impacts of Cohort, Age and Time. *Journal of Population Economics*, forthcoming.

Fisher, F.M. (1961): Identification Criteria in Non-linear Systems. *Econometrica* **20**, 574–590.

Hall, B.H., Mairesse, J., and Turner, L. (2007): Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists. *Economics of Innovation and New Technology* **16**, 159–177.

Heckman, J., and Robb, R. (1985): Using Longitudinal Data to Estimate Age, Period, and Cohort Effects in Earnings Equations. In Mason, W., and Fienberg, S. (eds.) *Cohort Analysis in Social Research: Beyond the Identification Problem.* New York: Springer.

Mason, K.H., Mason, M.M., Winsborough, H.H., and Poole, W.K. (1973): Some Methodological Issues in Cohort Analysis of Archival Data. *American Sociological Review* **38**, 242–258.

McKenzie, D.J. (2006): Disentangling Age, Cohort and Time Effects in the Additive Model. *Oxford Bulletin of Economics and Statistics* **68**, 473–495.

Portrait, F., Alessie, R., and Deeg, D. (2002): Disentangling the Age, Period, and Cohort Effects Using a Modeling Approach. Tinbergen Institute Discussion Paper, TI 2002-120/3.

Ree, J. de, and Alessie, R. (2011): Life Satisfaction and Age: Dealing With Underidentification in Age-Period-Cohort Models. *Social Science & Medicine* **73**, 177–182.

Rodgers, W.L. (1982): Estimable Functions of Age, Period, and Cohort Effects. *American Sociological Review* **47**, 774–787.

Winship, C., and Harding, D.J. (2008): A Mechanism-Based Approach to the Identification of Age-Period-Cohort Models. *Sociological Methods & Research* **36**, 362–401.

Yang, Y., and Land, K.C. (2008): Age-Period-Cohort Analysis of Repeated Cross-Section Surveys. Fixed or Random Effects? *Sociological Methods & Research* **36**, 297–326.

Appendix Tables

Table A.1: Correlation matrix including powers of variables

4 502 991 observations after truncation

A. Powers of (demeaned) ACT-variables

	c	t	a	c^2	t^2	a^2	c^3	t^3	a^3	c^4	t^4	a^4
c	1.0000											
t	0.2744	1.0000										
a	**-0.9509**	0.0367	1.0000									
c^2	-0.0325	-0.0175	0.0281	1.0000								
t^2	-0.0490	-0.1245	0.0108	0.0852	1.0000							
a^2	-0.0304	-0.0018	0.0310	**0.8324**	0.0190	1.0000						
c^3	**0.8502**	0.3324	**-0.7765**	-0.0626	-0.0611	-0.0428	1.0000					
t^3	0.2537	**0.9103**	0.0293	-0.0240	-0.2156	-0.0014	0.3122	1.0000				
a^3	**-0.8420**	0.0287	**0.8843**	0.0410	0.0174	0.0441	**-0.8638**	0.0210	1.0000			
c^4	-0.0479	-0.0346	0.0387	**0.9241**	0.1283	**0.7216**	-0.0978	-0.0452	0.0584	1.0000		
t^4	-0.0668	-0.1996	0.0052	0.0824	**0.9538**	0.0175	-0.0838	-0.3220	0.0115	0.1257	1.0000	
a^4	-0.0387	-0.0025	0.0395	**0.7945**	0.0230	**0.9407**	-0.0582	-0.0024	0.0600	**0.7663**	0.0214	1.0000

B. `abs` *and* `fdum` *versus powers of (demeaned) ACT-variables*

	`abs`	`fdum`
c	-0.1004	0.0580
t	0.0247	0.0220
a	0.1123	-0.0532
c^2	0.0507	-0.0240
t^2	-0.0493	0.0035
a^2	0.0498	-0.0315
c^3	-0.0976	0.0439
t^3	0.0261	0.0212
a^3	0.1170	-0.0386
c^4	0.0421	-0.0305
t^4	-0.0447	0.0006
a^4	0.0433	-0.0407

Symbols a, c, t in this table correspond to `age,coh,yea` in Tables 3 and A.2

Table A.2: Standard deviations for subsamples

		Non-truncated:			*Truncated:*		
		All.	Males	Females	All	Males	Females
`abs`	overall	51.1401	47.8121	54.2816	110.9083	112.7852	109.4792
	between	19.3747	18.6246	20.0012	94.5701	98.1903	91.3440
	within	47.5608	44.3741	50.6283	80.5409	79.1534	81.5618
`coh`	overall	14.9596	14.8772	15.0416	12.5367	12.6974	12.3780
	between	18.0487	17.9254	18.1711	13.6686	13.8124	13.4984
	within	0	0	0	0	0	0
`yea`	overall	4.0382	4.0371	4.0390	3.8833	3.8746	3.8882
	between	2.4929	2.4876	2.4984	3.3038	3.3662	3.2464
	within	3.7720	3.7718	3.7722	2.7512	2.6497	2.8246
`age`	overall	14.5812	14.5084	14.6533	12.0635	12.2585	11.8856
	between	16.1962	16.0862	16.3050	12.6251	12.8686	12.3737
	within	3.7720	3.7718	3.7722	2.7512	2.6497	2.8246