MEMORANDUM

No 08/2013

Identifying Age-Cohort-Time Effects, Their Curvature and Interactions from Polynomials: Examples Related to Sickness Absence

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 07/13	Alessandro Corsi and Steinar Strøm *The Price Premium for Organic Wines: Estimating a Hedonic Farm-gate Price Equations*
No 06/13	Ingvild Almås and Åshild Auglænd Johnsen *The Cost of Living in China: Implications for Inequality and Poverty*
No 05/13	André Kallåk Anundsen *Econometric Regime Shifts and the US Subprime Bubble*
No 04/13	André Kallåk Anundsen and Christian Heebøll *Supply Restrictions, Subprime Lending and Regional US Housing Prices*
No 03/13	Michael Hoel *Supply Side Climate Policy and the Green Paradox*
No 02/13	Michael Hoel and Aart de Zeeuw *Technology Agreements with Heteregeneous Countries*
No 01/13	Steinar Holden, Gisle James Natvik and Adrien Vigier *An Equilibrium Model of Credit Rating Agencies*
No 32/12	Leif Andreassen, Maria Laura Di Tomasso and Steinar Strøm *Do Medical Doctors Respond to Economic Incentives?*
No 31/12	Tarjei Havnes and Magne Mogstad *Is Universal Childcare Leveling the Playing Field?*
No 30/12	Vladimir E. Krivonozhko, Finn R. Førsund and Andrey V. Lychev *Identifying Suspicious Efficient Units in DEA Models*

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

IDENTIFYING AGE-COHORT-TIME EFFECTS, THEIR CURVATURE AND INTERACTIONS FROM POLYNOMIALS: EXAMPLES RELATED TO SICKNESS ABSENCE

ERIK BIØRN
Department of Economics, University of Oslo,
P.O. Box 1095 Blindern, 0317 Oslo, Norway
E-mail: erik.biorn@econ.uio.no

Memo 08/2013-v1

Abstract: In the paper is considered identification of coefficients in equations explaining a continuous variable, say the number of sickness absence days of an individual per year, by cohort, time and age, subject to their definitional identity. Extensions of a linear equation to polynomials, including additive polynomials, are explored. The cohort+time=age identity makes the treatment of interactions important. If no interactions between the three variables are included, only the coefficients of the linear terms remain unidentified unless additional information is available. Illustrations using a large data set for individual long-term sickness absence in Norway are given. The sensitivity to the estimated marginal effects of cohort and age at the sample mean, as well as conclusions about the equations' curvature, are illustrated. We find notable differences in this respect between linear and quadratic equations on the one hand and cubic and fourth-order polynomials on the other.

Keywords: Age-cohort-time problem, identification, polynomial regression, interaction, age-cohort curvature, panel data, sickness absence.

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 thank Daniel Bergsvik for preparing the data file and other assistance and Knut Røed and Terje Skjerpen for comments.

1 Introduction

The 'Age-Cohort-Time (ACT) problem' in individual data, following from the identity cohort+age = time is much discussed in social and medical research. Ways of handling it are considered in 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 potential identification problems it induces has motivated inclusion of additional assumptions to reduce the parameter space. The nature of the ACT identification problem depends critically on the functional form chosen. It is notorious in linear models, but may also arise when using more flexible functional forms, say polynomials, logarithmic functions or 'non-parametric' specifications.

In this paper the ACT problem when explaining a continuous variable by age, cohort and time, also treated as continuous, is reconsidered. The challenges in quantifying marginal effects of age and cohort are, *inter alia*, related to the functional form chosen. Starting from a linear, benchmark model, we extend it to polynomials in age, cohort and time up to order four. The role of interactions between the three variables and their possible effect on identification is explored. Illustrations based on a large Norwegian data set containing sickness absence records, measured in sickness absence days, from 1.7 million persons in the Norwegian labour force are given.

A general specification of the theoretical regression, with a continuous endogenous variable y explained by (a, c, t), denoting age, cohort, time, and satisfying $a+c=t$, is

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

where the function f is parametric, but so far unspecified. Eliminating one of the three explanatory variables, we can write the equation alternatively, as

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

An additive subclass of (1) has the form

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

which can be rewritten as

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

Which of the parameters of f, or of f_a, f_c, f_t, can be identified depends on its functional form. If f is linear, or a monotonically increasing transformation of a linear function, not all parameters can be identified. This is, loosely speaking, due to the fact that 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 a non-linear function, we have for example

[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".

$g(a)+g(c) \neq g(t)$. If g is restricted to be a polynomial we can be more specific: while t^3 and (a^3, c^3) are not collinear, t^3 is collinear with (a^3, c^3, a^2c, ac^2), etc. This 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 how their coefficients are restricted. This is a main issues of the paper.

The paper proceeds as follows. In Section 2 the ACT problem for the simple model with f (and f_a, f_c, f_t) linear 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 second- and higher order coefficients of f_a, f_c, f_t can be identified. The identifiability of coefficients of higher-order terms when we turn to the more general polynomial version of (1) depends on which interactions between the ACT variables are included and on their parametrization. The distinction between *full polynomials* and *additive polynomials* is crucial. In Section 4, alternative definitions of marginal effects for such models are elaborated. Illustrations for polynomial of orders up to four, based on a large set of sickness absence records for individuals in the Norwegian labour force, are discussed next, in Sections 5, 6 and 7. We conclude that long-term sickness, in absence days, is clearly non-linear in cohort and age, and that the model fit is significantly improved and the curvature changed when polynomial additivity is relaxed by including interactions between cohort and age, at least for polynomials up to order four.

2 The Age-Cohort-Time problem in a linear model revisited

Assume that observations from n individuals on the response variable y_i, and three explanatory variables, birth cohort, time and age of individual i, (c_i, t_i, a_i), are available. The regression equation is

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

Other explanatory variables are suppressed, but could be absorbed 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, there is reason to believe that say $\delta = 0$, γ and β will be identifiable as pure cohort and age affects. We have in general

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

which exemplifies (2).

The first-order conditions for the OLS problem for (5), subject to (6), exemplifies the mathematical problem of solving a system of linear equations subject to linear variable restrictions. We first comment on this problem, describe how the data type impacts identifiability and consider the regressor covariance matrix in some typical cases.

Formally, 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 $a_i + c_i = t_i$ has three independent first-order conditions: $\sum_i u_i = 0$ plus equations among $\sum_i u_i c_i = \sum_i u_i t_i = \sum_i u_i a_i = 0$. Therefore only two linear combinations of the slope coefficients can be identified: either $(\gamma+\delta), (\beta+\delta)$ or $(\delta+\gamma), (\beta-\gamma)$ or $(\gamma-\beta), (\delta+\beta)$. Boundary cases are:

1. Data from *one cohort*: $c_i = c, a_i = t_i - c$. Only $\beta+\delta$ can be identified, and *either a_i or t_i can be included as a regressor.*

2. Data from *one period*: $t_i = t, a_i = t - c_i$. Only $\beta-\gamma$ can be identified, and *either a_i or c_i can be included as a regressor.*

3. Data from individuals of *one age*: $a_i = a, c_i = t_i - a$. Only $\gamma+\delta$ can be identified, and *either c_i or t_i can be included as a regressor.*

In general, $X_i = (c_i, t_i, a_i)$ has a variance-covariance matrix of rank at most 2:

$$\mathbf{\Sigma}_X = \begin{bmatrix} \sigma_{cc} & \sigma_{ct} & \sigma_{ca} \\ \sigma_{ct} & \sigma_{tt} & \sigma_{ta} \\ \sigma_{ca} & \sigma_{ta} & \sigma_{aa} \end{bmatrix} = \begin{bmatrix} \sigma_{cc} & \sigma_{ct} & \sigma_{ct} - \sigma_{cc} \\ \sigma_{ct} & \sigma_{tt} & \sigma_{tt} - \sigma_{ct} \\ \sigma_{ct} - \sigma_{cc} & \sigma_{tt} - \sigma_{ct} & \sigma_{tt} - 2\sigma_{ct} + \sigma_{cc} \end{bmatrix},$$

where column (row) 3 is the difference between columns (rows) 2 and 1. The correlation pattern of (c_i, t_i, a_i) determines the kind of inference obtainable. Letting $\rho_{ct}, \rho_{ca}, \rho_{ta}$ be the respective correlation coefficients, the three cases, with $\text{rank}(\mathbf{\Sigma}_X) = 1$, can be described as:

Case 1. One cohort: $\sigma_{cc} = \sigma_{ct} = 0,\ \rho_{ta} = 1,\ \mathbf{\Sigma}_X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma_{tt} & \sigma_{tt} \\ 0 & \sigma_{tt} & \sigma_{tt} \end{bmatrix}$.

Case 2. One period: $\sigma_{tt} = \sigma_{ct} = 0,\ \rho_{ca} = -1,\ \mathbf{\Sigma}_X = \begin{bmatrix} \sigma_{cc} & 0 & -\sigma_{cc} \\ 0 & 0 & 0 \\ -\sigma_{cc} & 0 & \sigma_{cc} \end{bmatrix}$.

Case 3. One age: $\sigma_{tt} = \sigma_{ct} = \sigma_{cc},\ \rho_{ct} = 1,\ \mathbf{\Sigma}_X = \begin{bmatrix} \sigma_{cc} & \sigma_{cc} & 0 \\ \sigma_{cc} & \sigma_{cc} & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Generalizations for which $\text{rank}(\mathbf{\Sigma}_X) = 2$ are

Case 4. Cohort and Age uncorrelated: $\sigma_{tt} > \sigma_{cc} = \sigma_{ct},\ \rho_{ct}^2 + \rho_{ta}^2 = 1$,

$$\mathbf{\Sigma}_X = \begin{bmatrix} \sigma_{cc} & \sigma_{cc} & 0 \\ \sigma_{cc} & \sigma_{tt} & \sigma_{tt} - \sigma_{cc} \\ 0 & \sigma_{tt} - \sigma_{cc} & \sigma_{tt} - \sigma_{cc} \end{bmatrix},\ \rho_{ct} = \left[\frac{\sigma_{cc}}{\sigma_{tt}}\right]^{\frac{1}{2}},\ \rho_{ta} = \left[1 - \frac{\sigma_{cc}}{\sigma_{tt}}\right]^{\frac{1}{2}}.$$

Case 5. Period and Age uncorrelated: $\sigma_{cc} > \sigma_{tt} = \sigma_{ct},\ \rho_{ct}^2 + \rho_{ca}^2 = 1$,

$$\mathbf{\Sigma}_X = \begin{bmatrix} \sigma_{cc} & \sigma_{tt} & -(\sigma_{cc} - \sigma_{tt}) \\ \sigma_{tt} & \sigma_{tt} & 0 \\ -(\sigma_{cc} - \sigma_{tt}) & 0 & \sigma_{cc} - \sigma_{tt} \end{bmatrix},\ \rho_{ct} = \left[\frac{\sigma_{tt}}{\sigma_{cc}}\right]^{\frac{1}{2}},\ \rho_{ca} = -\left[1 - \frac{\sigma_{tt}}{\sigma_{cc}}\right]^{\frac{1}{2}}.$$

Case 6. Cohort and Period uncorrelated: $\sigma_{cc} \neq \sigma_{tt},\ \sigma_{ct} = 0,\ \rho_{ca}^2 + \rho_{ta}^2 = 1$,

$$\mathbf{\Sigma}_X = \begin{bmatrix} \sigma_{cc} & 0 & -\sigma_{cc} \\ 0 & \sigma_{tt} & \sigma_{tt} \\ -\sigma_{cc} & \sigma_{tt} & \sigma_{cc} + \sigma_{tt} \end{bmatrix},\ \rho_{ca} = \left[\frac{\sigma_{cc}}{\sigma_{cc} + \sigma_{tt}}\right]^{\frac{1}{2}},\ \rho_{ta} = \left[\frac{\sigma_{tt}}{\sigma_{cc} + \sigma_{tt}}\right]^{\frac{1}{2}}.$$

Cases 4, 5 and 6 generalize Cases 1 & 3, 2 & 3 and 1 & 2, respectively.

3 Extension to polynomial models

We consider two extensions of (5), one with the additive form (3) and one with the more general form (1). In addition, a third hypothetical reference model is outlined.

Sum of Pth order polynomials in age, cohort and time

First, consider a sum of Pth order polynomials in a_i, c_i, t_i. We call this an *additive* Pth order polynomial. Eliminating, by (6), alternatively t_i, a_i and c_i, we get an equation, being a special case of (4), as respectively:

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

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

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

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 equivalent regressions with $3P-1$ regressors:

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

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

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

with coefficients defined as, respectively,

$$\begin{array}{llll}
\delta_p = \delta_p^*, & \beta_p = \beta_p^* + \delta_p^*, & \gamma_p = \gamma_p^* + \delta_p^*, & p = 2, \dots, P, \\
 & \beta_1 = \beta_1^* + \delta_1^*, & \gamma_1 = \gamma_1^* + \delta_1^*, &
\end{array} \tag{14}$$

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

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

All these coefficients are identifiable without additional conditions being needed.

Hence, although a sum of three Pth order polynomials seemingly has no interactions, its reparametrization that forms (11) from (8) etc., creates interactions between the (powers of the) two remaining variables and reduces the number of identified coefficients to $C_{1P} = 3P-1$. The interaction terms have a particular structure, however. The starred coefficients of the linear terms in (8)–(10), $(\beta_1^*, \gamma_1^*, \delta_1^*)$, cannot be identified unless restrictions are imposed, while $(\beta_p^*, \gamma_p^*, \delta_p^*)$ for $p \geq 2$ can be identified as follows:

> *$\delta_p = \delta_p^*$ is the coefficient of $\sum_{r=1}^{p-1} \binom{p}{r} a_i^r c_i^{p-r}$ in (11);*
> *(β_p^*, γ_p^*) can be derived from β_p and γ_p as prescribed by (14).*
>
> *$\bar{\beta}_p = \beta_p^*$ is the coefficient of $\sum_{r=1}^{p-1} \binom{p}{r} t_i^r (-c_i)^{p-r}$ in (12);*
> *(δ_p^*, γ_p^*) can be derived from $\bar{\delta}_p$ and $\bar{\gamma}_p$ as prescribed by (15).*

$\tilde{\gamma}_p = \gamma_p^*$ *is the coefficient of* $\sum_{r=1}^{p-1}\binom{p}{r}t_i^r(-a_i)^{p-r}$ *in (13);*
(δ_p^*, β_p^*) *can be derived from* $\tilde{\delta}_p$ *and* $\tilde{\beta}_p$ *as prescribed by (16).*

This describes compactly the ACT identification problem for an additive Pth order polynomial model of type (3)–(4).

FULL POLYNOMIALS IN TWO VARIABLES

The above additive ACT polynomials, which exemplify (3)–(4), have an 'asymmetry'. To obtain a model exemplifying (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 has the potential to improve the fit to data. We elaborate this extension only for (8), reparametrized as (11). Consider then

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

This increases the number of (identifiable) coefficients to $C_{2P} = 2P + \frac{1}{2}P(P-1) = \frac{1}{2}P(P+3)$, which, since $C_{2P} - C_{1P} = \frac{1}{2}P(P-3)+1$, is an effective increase when $P > 2$. Model (17) specializes to (11) for

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

A third model with all interactions between $(a, c), (a, t), (c, t)$ [although not the (a, c, t) interaction] included, would have had $C_{3P} = 3P + 3\frac{1}{2}P(P-1) = \frac{3}{2}P(P+1)$ coefficients. Hence $C_{3P} - C_{2P} = P^2$, and if $P > 2$ we have $C_{3P} > C_{2P} > C_{1P}$.[2] However, this model is *hypothetical*, since the inescapable restriction (6) precludes identification of all its coefficients.[3]

The number of coefficients in the three models is exemplified in Table 1, which also shows the difference between the number of coefficients (columns 5 and 8) and their increase when the polynomial order is increased by one (columns 2, 4 and 7):

$$\begin{array}{llll} C_{1P} = 3P-1, & C_{2P} = \frac{1}{2}P(P+3), & C_{3P} = \frac{3}{2}P(P+1) & \Longrightarrow \\ \Delta C_{1P} = 3, & \Delta C_{2P} = P+1, & \Delta C_{3P} = 3P & (P \geq 2). \end{array}$$

Table 1: THREE VERSIONS OF POLYNOMIAL MODELS. NO. OF COEFFICIENTS

Polyn. order	Additive polynomials		Full polynomials in 2 vars.			Full polynomials in 3 vars.		
P	C_{1P}	ΔC_{1P}	C_{2P}	ΔC_{2P}	$C_{2P}-C_{1P}$	C_{3P}	ΔC_{3P}	$C_{3P}-C_{2P}$
1	2	-	2	-	0	3	-	1
2	5	3	5	3	0	9	6	4
3	8	3	9	4	1	18	9	9
4	11	3	14	5	3	30	12	16
5	14	3	20	6	6	45	15	25
6	17	3	27	7	10	63	18	36

[2] Further extensions to *multinomial* models, which, in our three-variable case, would have included also terms in $a_i^r c_i^q t_i^{p-r-q}$ and therefore would have been 'still more hypothetical', are not considered.

[3] Hall, Mairesse and Turner (2007, p. 162), with reference to Heckman and Robb (1985), comment on this kind of model as follows (apparently implying all interaction terms included when using the term 'polynomial') : "... for the linear model, only two of the three linear coefficients are identified. For a quadratic model, only three of the six quadratic coefficients are identified, and so forth. So, although low-order polynomials seem to be an attractive way to model these effects because of their smoothness, in practice, they have not been much used because the lack of identification is so obvious." Hall, Mairesse and Turner disregard the more parsimonious *additive* polynomial parametrization represented by (11)–(13), in which only one first-order coefficient is unidentified.

An example

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

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

When imposing, see (18),

$$\delta_{21} = 2\delta_2, \qquad \delta_{31} = \delta_{32} = 3\delta_3, \qquad \delta_{41} = \delta_{43} = 4\delta_4, \qquad \delta_{42} = 6\delta_4,$$

which implies the $C_{2P} - C_{1P} = 3$ effective restrictions $\delta_{31} = \delta_{32}, \delta_{41} = \delta_{43} = \frac{2}{3}\delta_{42}$ and replace $(\delta_{21}, \delta_{31}, \delta_{32}, \delta_{41}, \delta_{42}, \delta_{43})$ by $(\delta_2, \delta_3, \delta_4)$, we get the additive polynomial

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

Writing (17) as $F(a_i, c_i) = \mathsf{E}(y_i|a_i, c_i)$ and letting $F_a(a_i, c_i) \equiv \partial\mathsf{E}(y_i|a_i, c_i)/\partial a_i$, $F_c(a_i, c_i) \equiv \partial\mathsf{E}(y_i|a_i, c_i)/\partial c_i$, the corresponding partial derivatives become

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

In the additive polynomial case (11) we have

$$\begin{aligned}F_a(a_i,c_i) &= \beta_1 + 2\delta_2 c_i + 2\beta_2 a_i + 3\beta_3 a_i^2 + 4\beta_4 a_i^3 \\ &\quad + 3\delta_3(2a_i c_i + c_i^2) + 4\delta_4(3a_i^2 c_i + 3a_i c_i^2 + c_i^3), \\ F_c(a_i,c_i) &= \gamma_1 + 2\delta_2 a_i + 2\gamma_2 c_i + 3\gamma_3 c_i^2 + 4\gamma_4 c_i^3 \\ &\quad + 3\delta_3(a_i^2 + 2a_i c_i) + 4\delta_4(3a_i c_i^2 + 3a_i^2 c_i + a_i^3).\end{aligned}$$

Increase P from 1 to 4 implies replacing in (7) $\gamma+\delta$ and $\beta+\delta$ with $F_c(a_i, c_i)$ and $F_a(a_i, c_i)$:

$$\Delta\mathsf{E}(y_i|\Delta a_i, \Delta c_i) = F_c(a_i, c_i)\Delta c_i + F_a(a_i, c_i)\Delta a_i.$$

4 Marginal effects

In the OLS regressions to be considered, demeaned observations of cohort, year and age will be exploited. The primary intention of this transformation is to facilitate comparison of results across models with different polynomial orders.

A basis for interpreting the coefficient estimates is obtained by looking at some mathematical expressions for 'marginal effects' of cohort and age. Some notation for *central* moments will then be needed. Let $\boldsymbol{a} = a - \mathsf{E}(a)$, $\boldsymbol{c} = c - \mathsf{E}(c)$ and define

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

where, obviously

$$\begin{aligned}&\boldsymbol{\mu}_a(1) = \boldsymbol{\mu}_{ac}(1,0) = \boldsymbol{\mu}_c(1) = \boldsymbol{\mu}_{ac}(0,1) = 0, \\ &\boldsymbol{\mu}_{ac}(p,0) = \boldsymbol{\mu}_a(p),\ \boldsymbol{\mu}_{ac}(0,q) = \boldsymbol{\mu}_c(q).\end{aligned}$$

Corresponding to the non-additive polynomial equation (17), after having deducted from cohort and age their expectations, we have[4]

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

The law of iterated expectations gives

$$\mathsf{E}(y|\boldsymbol{a}) = \alpha + \beta_1\boldsymbol{a} + \textstyle\sum_{p=2}^{P}\beta_p\boldsymbol{a}^p + \sum_{p=2}^{P}\gamma_p\boldsymbol{\mu}_c(p) + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}\boldsymbol{a}^r\boldsymbol{\mu}_{c|a}(p-r), \tag{20}$$

$$\mathsf{E}(y|\boldsymbol{c}) = \alpha + \gamma_1\boldsymbol{c} + \textstyle\sum_{p=2}^{P}\beta_p\boldsymbol{\mu}_a(p) + \sum_{p=2}^{P}\gamma_p\boldsymbol{c}^p + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}\boldsymbol{\mu}_{a|c}(r)\boldsymbol{c}^{p-r}, \tag{21}$$

$$\mathsf{E}(y) = \alpha + \textstyle\sum_{p=2}^{P}\beta_p\boldsymbol{\mu}_a(p) + \sum_{p=2}^{P}\gamma_p\boldsymbol{\mu}_c(p) + \sum_{p=2}^{P}\sum_{r=1}^{p-1}\delta_{pr}\boldsymbol{\mu}_{ac}(r,p-r). \tag{22}$$

Two kinds of marginal effects 'at the mean' can be defined. They are obtained by taking the expectation and the differentiation operations in opposite succession.

Expected marginal effects: Definition 1 (Differentiation prior to expectation): The marginal expectations of the derivatives of y, with respect to (demeaned) age, $\boldsymbol{a}$, and with respect to (demeaned) cohort, $\boldsymbol{c}$ – taken across the age-cohort distribution – can be expressed in terms of population moments as[5]

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

Since the coefficients of the quadratic terms in (11), β_2 and γ_2, do not enter these expressions, linear and quadratic functions simply give $\mathsf{E}[\partial y/\partial \boldsymbol{a}] = \beta_1$ and $\mathsf{E}[\partial y/\partial \boldsymbol{c}] = \gamma_1$. If $P \geq 3$ second and higher-order moments of $\boldsymbol{a}$ and $\boldsymbol{c}$ will interact with the coefficients of the cubic and higher-order terms. When $P=3$, (23) gives for example

$$\begin{aligned} \mathsf{E}[\partial y/\partial \boldsymbol{a}] &= \beta_1 + 3\beta_3\boldsymbol{\mu}_a(2) + \delta_{31}\boldsymbol{\mu}_c(2) + 2\delta_{32}\boldsymbol{\mu}_{ac}(1,1), \\ \mathsf{E}[\partial y/\partial \boldsymbol{c}] &= \gamma_1 + 3\gamma_3\boldsymbol{\mu}_c(2) + \delta_{32}\boldsymbol{\mu}_a(2) + 2\delta_{31}\boldsymbol{\mu}_{ac}(1,1). \end{aligned}$$

Expected marginal effects: Definition 2 (Expectation operation prior to differentiation): Two versions of the effects thus defined can be obtained from (19). First, by conditioning on both age and cohort and differentiating with respect to one of them, we get

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

Second, if we condition the expectation on the variable which is subject to differentiation only, (20) and (21) yield

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

[4]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.

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

In (24) and (25), unlike (23), the second-order coefficients β_2 and γ_2 always occur, except when the derivatives are evaluated at the expected cohort and age ($\boldsymbol{a}=\boldsymbol{c}=0$). When $P=3$, we have for example

$$\begin{aligned}
\partial \mathsf{E}(y|\boldsymbol{a},\boldsymbol{c})/\partial \boldsymbol{a} &= \beta_1 + 2\beta_2\boldsymbol{a} + 3\beta_3\boldsymbol{a}^2 + \delta_{21}\boldsymbol{c} + \delta_{31}\boldsymbol{c}^2 + 2\delta_{32}\boldsymbol{ac},\\
\partial \mathsf{E}(y|\boldsymbol{c},\boldsymbol{a})/\partial \boldsymbol{c} &= \gamma_1 + 2\gamma_2\boldsymbol{c} + 3\gamma_3\boldsymbol{c}^2 + \delta_{21}\boldsymbol{a} + \delta_{32}\boldsymbol{a}^2 + 2\delta_{31}\boldsymbol{ac},
\end{aligned}$$

$$\begin{aligned}
\partial \mathsf{E}(y|\boldsymbol{a})/\partial \boldsymbol{a} &= \beta_1 + 2\beta_2\boldsymbol{a} + 3\beta_3\boldsymbol{a}^2 + \delta_{21}\mathsf{E}(\boldsymbol{c}|\boldsymbol{a}) + \delta_{31}\mathsf{E}(\boldsymbol{c}^2|\boldsymbol{a}) + 2\delta_{32}\boldsymbol{a}\mathsf{E}(\boldsymbol{c}|\boldsymbol{a}),\\
\partial \mathsf{E}(y|\boldsymbol{c})/\partial \boldsymbol{c} &= \gamma_1 + 2\gamma_2\boldsymbol{c} + 3\gamma_3\boldsymbol{c}^2 + \delta_{21}\mathsf{E}(\boldsymbol{a}|\boldsymbol{c}) + \delta_{32}\mathsf{E}(\boldsymbol{a}^2|\boldsymbol{c}) + 2\delta_{31}\boldsymbol{c}\mathsf{E}(\boldsymbol{a}|\boldsymbol{c}).
\end{aligned}$$

5 An illustration: Sickness absence

We now illustrate the above results by using a large panel data set for long-term sickness absence of individuals in the Norwegian labour force. Sickness absences of length at least 16 days are recorded in the data set, while shorter absences, labeled short-term sickness absence are (for institutional reasons) recorded as a zero number.[6] The full data set, including individuals with no recorded 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. The individuals in the full sample have, on average, 12.6 absence days, while the mean number of absence days in the truncated sample, with zero absence entries removed, is 112.7. Only for 1 786 105 individuals at least one sickness absence of at least 16 days is recorded during these 14 years. It is the truncated sample, which has 4 502 991 observations, that will be used in the illustrations. We restrict the illustrations to polynomials of order at most $P=4$.

Table 2: Correlation matrices

	All observations				Observations with abs > 16			
	abs	coh	yea	age	abs	coh	yea	age
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

Cohort and year are measured from the year 1920, giving `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 years.

Correlation matrices are given in Table 2. Unsurprisingly, `abs` is positively correlated with `age` and negatively correlated with `coh` (correlation coefficients 0.0456 and -0.0376 in the full sample, 0.1123 and -0.1004 in the truncated sample). The correlation is stronger after truncation than before because of the omission of all zero absence spells. As expected, `age` and `coh` show strong negative correlation, with correlation coefficients -0.9630 and -0.9509 before and after truncation, respectively.[7] We are very far from having a data set like Case 4 in Section 2, although, with corr(`age`, `yea`) = 0.0367, it is not far from resembling Case 5.

[6]Sickness absence are, for part-time workers, measured in full-time equivalents, and the number of absence days recorded in a year refers to absence spells *starting in that year and possibly extending to the next year.* For more details on definitions and institutional setting See Biørn (2013) and Biørn *et al.* (2013).

[7]In a corresponding cross-section this correlation would, of course, have been -1; confer Case 2 in Section 2.

Model-tree: Table 3 specifies 18 polynomial models of orders 1 through 4. They are labeled as $d.k$, where d and k indicate, respectively, the polynomial order and the collection of power terms included (for models with $k=1,2,3$) and interaction terms (for models with $k=4$). From now on we let c, t, a symbolize *demeaned* variables. The model-tree can be described as follows:

- The three linear models, 1.1, 1.2 and 1.3, are equivalent, which exemplifies the ACT problem outlined in Section 2.
- Models $2.k$, $3.k$ and $4.k$ ($k = 1,2,3$) include linear and power terms in two 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) – see also (8) and (14) – and have as special cases, respectively, $2.k$, $3.k$, and $4.k$ ($k = 1,2,3$).
- Models 2.4, 3.4, and 4.4 extend Models $2.k$, $3.k$, and $4.k$ ($k = 1,2,3$), respectively, by adding interaction terms to the power terms. This extension exemplifies (17) and increases the number of coefficients to 6, 10, and 15 (including intercept), respectively.[8]

Table 3: 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 4.4 nests 4.1 & 3.4. Model 4.0 nests 4.1, 4.2, 4.3 & 3.0.
Model 3.4 nests 3.1 & 2.4. Model 3.0 nests 3.1, 3.2, 3.3 & 2.0.
Model 2.4=2.0 nests 2.1, 2.2, 2.3 & 1.1=1.2=1.3.

While Model 2.4 reparametrizes Model 2.0, Model 3.0 imposes one restriction on Model 3.4, and Model 4.0 imposes 3 restrictions on Model 4.4. See the example with $P=4$ 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.[9]

Goodness of fit: Table 4 contains *fit statistics* for OLS estimation of the 18 models based on the (truncated) data set: sum of squared residuals (`SSR`), standard error of regression (σ_u) and squared multiple correlation (R^2). The fit, measured by the σ_u estimate, is about 1.1×10^{-4} in all models (column 2). Measured by R^2 the fit varies between 0.013 and 0.020 (column 3). Hence, even with as many as 14 coefficients and intercept,

[8]The equivalent models obtained from (12) and (13) for $P=4$ are omitted from this survey. Restricting attention to (11) and (17) in estimation, has the advantage that no sign-shifts for the binomial coefficients will have to be dealt with.

[9]See the footnote of Table 3.

Table 4: Estimated models. OLS fit statistics

Observations with abs > 16 *only*

Model	SSR$\times 10^{-14}$	$\sigma_u \times 10^{-4}$	R^2
1.1	5.4667	1.1018	0.013041
1.2	5.4667	1.1018	0.013041
1.3	5.4667	1.1018	0.013041
2.0	5.4385	1.0990	0.018146
2.1	5.4532	1.1005	0.015480
2.2	5.4387	1.0990	0.018099
2.3	5.4414	1.0993	0.017612
3.0	5.4306	1.0982	0.019567
3.1	5.4457	1.0997	0.016848
3.2	5.4331	1.0984	0.019111
3.3	5.4338	1.0985	0.018989
4.0	5.4279	1.0979	0.020049
4.1	5.4438	1.0995	0.017188
4.2	5.4311	1.0982	0.019474
4.3	5.4314	1.0983	0.019427
2.4	5.4385	1.0990	0.018146
3.4	5.4304	1.0982	0.019602
4.4	5.4276	1.0979	0.020104

Table 5: Equivalent linear models. OLS estimates

Standard errors below coefficient estimates.
All coefficients and standard errors multiplied by 100

Observations with abs > 16 *only*

	Model 1.1	*Model 1.2*	*Model 1.3*
c	58.788114	-102.576460	
	1.337988	0.430708	
t		161.364574	58.788114
		1.390470	1.337988
a	161.364574		102.576460
	1.390470		0.430708

Table 6: Additive quadratic models. OLS estimates

Standard errors below coefficient estimates.
All coefficients and standard errors multiplied by 100

Observations with abs > 16 *only*

	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 4.4, the most parameter-rich model, which includes as covariates only power and interaction terms in the ACT variables, has an explained variation in sickness absence that accounts for a minor part of the total.

Among the models with linear and power terms in two of the three variables, those including (c, a) (omitting t) [Models 2.1, 3.1, and 4.1] give somewhat poorer fit than the corresponding models which include (c, t) [Models 2.2, 3.2, and 4.2] or (t, a) [Models 2.3, 3.3, and 4.3]. The improved fit, indicated by a reduced SSR, when the regressors include both second, third, and fourth powers *of all the three variables* – also 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, Model 3.1 against 3.0, and 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 4), is 'compensated' by the large number of observations in the F-statistics, and together they lead to a clear rejection of the restrictive model.[10]

Coefficient estimates: Tables 5–9 give coefficient estimates for the 18 polynomial models listed in Table 3.[11] For the *linear models* we find, when *controlling for cohort*, 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 1.61 days (Table 5). Likewise, *controlling for calendar year*, an increase in birth-year by one (equivalent to being one year younger) gives an estimated reduction of long-term absence length of 1.03 days. Equivalently, *controlling for age*, an increase in birth-year by one (equivalent to increasing calendar time by one year) gives an estimated increase in long-term absence of 0.59 days. The interpretation of these estimates in terms of the coefficients in (5) follows from (7).

For the *quadratic, cubic, and fourth-order* polynomial regressions, we conclude from Tables 6, 7 and 8 that the estimated marginal cohort and age effects at the empirical mean – corresponding to γ_1 and β_1 in (24) at the expected age and cohort – are not invariant to the polynomial degree assumed. A certain pattern is visible, however: The estimates of γ_1 and β_1 from the *quadratic* Model 2.1 are close to the estimates from the *linear* Model 1.1 (year omitted): $(0.61, 1.61)$ days versus $(0.59, 1.61)$ days. This may be interpreted as an empirical counterpart to Definition 2 of Expected marginal effects, (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 are obtained. Likewise, 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 (year omitted). The estimates are $(0.86, 1.21)$ days versus $(0.84, 1.20)$ days. On the other hand, contrasting Model 4.2 with 3.2 (age omitted) and Model 4.3 with 3.3 (cohort omitted), larger discrepancies emerge. The closeness of the γ_1 and β_1 estimates for the third and fourth order polynomials as well as the discrepancies between the γ_1 and β_1 estimates from Models 3.1 and 2.1 cannot, however, be easily explained from either of the definitions of the expected marginal effects, (23) or (25).

[10] The rejection conclusion is also indicated from Tables 6–8 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.

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

6 Curvature and interactions in the quadratic case

In this section we take a closer look at the curvature and interactions implied by the quadratic models. The illustrations specifically contrasts *the quadratic additive three-variable model, Model 2.0* – which is equivalent to the quadratic *non-additive* two-variable model, Model 2.4 – and *three quadratic additive two-variable models, Models 2.1, 2.2 and 2.3*, which are all special cases of Model 2.0=Model 2.4.

Curvature inferred from Model 2.0: Consider first the estimates from the full quadratic model, first column of Table 6. This estimated model can be written in several forms:

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

By manipulating the second-order terms, eliminating, respectively, $\boldsymbol{t}^2$, $\boldsymbol{a}^2$, $\boldsymbol{c}^2$, and, by implication, including interaction terms in the two remaining variables, we get the following equivalent expressions

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

Since for neither version of (27) the Hessian matrix is positive or negative definite, *neither of the estimated regressions derived from this equation are convex or concave in the two variables included.*

However, *controlling for one variable*, the curvature and marginal effect of the other variable can from (27) be described as follows:[12]

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

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

Strictly, this is a *marginal cohort+year effect;* confer (2) and (7).

Positively sloping (around mean) 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, this is a *marginal age+year effect;* confer (2) and (7).

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

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

Strictly, this is a *marginal cohort−age effect;* confer (2) and (7).

Positively sloping (around mean) 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, this is a *marginal year+age effect;* confer (2) and (7).

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

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

Strictly, this is a *marginal age−cohort effect;* confer (2) and (7).

Positively sloping (around mean) 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, this is a *marginal year+cohort effect;* confer (2) and (7).

[12] 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.

We see that $m_{c|a} \equiv m_{t|a}$, $m_{a|c} \equiv m_{t|c}$ and $m_{c|t} \equiv -m_{a|t}$.

The *interaction terms* in the three versions of (27) are not equally important. We have:

1. *Omitting the cohort-age* (ca) *interaction* in the first version of (27), we get a result very different from that obtained from Model 2.1 (Table 6, column 4). The omitted variable bias seems to be large. The interaction has coefficient estimate 81.4 with standard error 0.7.

2. *Omitting the cohort-year* (ct) *interaction* from the second version of (27), we get a result largely similar to that obtained from Model 2.2 (Table 6, column 3). The omitted variable does not seem essential.

3. *Omitting the year-age* (ta) *interaction* from the third version of (27), we get a result largely similar to that obtained from Model 2.3 (Table 6, column 2). The omitted variable bias does not seem essential.

Conclusions 1–3 concur with the results in Table 4 that Models 2.0, 2.2 and 2.3 have approximately the same fit (although, as remarked, the R^2 of the former, according to F-tests, is significantly larger than that of the two latter, which reflects the large sample size). This fit is markedly better than the fit of Model 2.1. A message from our data is thus that an additive quadratic model in cohort and age is inferior in terms of fit.

Curvature inferred from Models 2.1, 2.2, and 2.3: Consider now the estimated curvature when instead using, respectively, Models 2.1, 2.2 and 2.3. We find from Table 6:

Model 2.1 is obtained from Model 2.0 by deleting the squared time variable,
i.e., deleting from Model 2.4 the cohort-age interaction.
Relying on *Model 2.1* we would get an absence equation showing *convexity in both* $\boldsymbol{c}$ *and* $\boldsymbol{a}$:

$$\widetilde{m}_{c|a} \equiv \partial \widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c})}/\partial \boldsymbol{c} = 60.6 + 3.9\boldsymbol{c},$$

$$\widetilde{m}_{a|c} \equiv \partial \widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{c})}/\partial \boldsymbol{a} = 161.8 + 3.2\boldsymbol{a}.$$

Model 2.2 is obtained from (the reparametrized) Model 2.0 by deleting the squared age,
i.e., deleting from Model 2.4 (reparametrized) the cohort-time interaction.
Relying on *Model 2.2* we would get an absence equation showing *convexity in* $\boldsymbol{c}$*, concavity in* $\boldsymbol{t}$:

$$\widetilde{m}_{c|t} \equiv \partial \widehat{\mathsf{E}(y|\boldsymbol{c},\boldsymbol{t})}/\partial \boldsymbol{c} = -101.9 + 7.8\boldsymbol{c},$$

$$\widetilde{m}_{t|c} \equiv \partial \widehat{\mathsf{E}(y|\boldsymbol{c},\boldsymbol{t})}/\partial \boldsymbol{t} = 144.5 - 82.4\boldsymbol{t}.$$

Model 2.3 is obtained from (the reparametrized) Model 2.0 by deleting the squared cohort,
i.e., deleting from Model 2.4 (reparametrized) the age-time interaction.
Relying on *Model 2.3* we would get an absence equation showing *convexity in* $\boldsymbol{a}$*, concavity in* $\boldsymbol{t}$:

$$\widetilde{m}_{a|t} \equiv \partial \widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{t})}/\partial \boldsymbol{a} = 101.9 + 7.2\boldsymbol{a},$$

$$\widetilde{m}_{t|a} \equiv \partial \widehat{\mathsf{E}(y|\boldsymbol{a},\boldsymbol{t})}/\partial \boldsymbol{t} = 41.5 - 77.0\boldsymbol{t}.$$

To conclusions **1–3** we can therefore add:

4. *Curvature in year:* The conclusion that the number of sickness-absence days is *concave* in *year* is robust: The coefficient estimate of $\boldsymbol{t}$ in $m_{t|a}$, $m_{t|c}$, $\widetilde{m}_{t|a}$ and $\widetilde{m}_{t|c}$ is around -77 to -82 (i) irrespective of which is the other conditioning variable in the equation (Model 2.0) and (ii) irrespective of whether age or cohort is excluded and hence not controlled for (Models 2.2 and 2.3).

5. *Curvature in cohort:* Model 2.0 implies that the number of sickness-absence days is *concave* in *cohort* when age is controlled for, while it is *convex* when year is controlled for. However, from Model 2.1 we find *convexity* in cohort (age included, year omitted and hence not controlled for), and the finding from Model 2.2 also suggests *convexity* in cohort (year included, age omitted and hence not controlled for). The omitted variables bias in Models 2.1 and 2.2 relative to Model 2.0 is severe: convexity of absence in cohort is spurious.

6. *Curvature in age:* Model 2.0 implies that the number of sickness-absence days is *concave* in *age* when cohort is controlled for, while it is *convex* when year is controlled for. However, from Model 2.1 we find *convexity* in age (cohort included, year omitted and hence not controlled for), and the finding from Model 2.3 also suggests *convexity* in age (year included, cohort omitted and hence not controlled for). The omitted variables bias in Models 2.1 and 2.2 relative to Model 2.0 is severe: convexity of absence in age is spurious.

7 Higher-order polynomials – a few remarks

Table 9 shows that cohort-age interactions are important also for cubic and fourth-order models. On the one hand, 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. On the other hand, the coefficient estimates of a^3, c^3 and a^4, c^4 are severely distorted when the interaction terms are omitted from the regression. In Table 10, columns 1–3, the results for the full cubic Model 3.4 and the cubic age-cohort Model 3.1 are contrasted with Model 3.0 when the power terms of t and their coefficient estimates are 'translated into' interaction terms in c and a. A similar comparison for the fourth-order Models 4.4, 4.1 and 4.0 is given in columns 4–6. Also here notable discrepancies emerge. Since the coefficient estimates of all second and higher-order coefficients in Models 3.4 and 3.0 are negative, the conclusion of concavity of sickness absence in age and in cohort is supported from the cubic models. The Model 3.1 estimates seem spurious.

Table 7: Additive cubic models. OLS estimates

Standard errors below coefficient estimates.

All coefficients multiplied by 100. $n = 4\,502\,991$

	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

Table 8: Additive fourth-order models. OLS estimates

Standard errors below coefficient estimates.

All coefficients multiplied by 100. $n = 4\,502\,991$

	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 9: Cubic and fourth-order full polynomials in cohort and age

OLS estimates

Standard errors below coefficient estimates.

Coefficients multiplied by 100

	Model 3.4	*Model 4.4*
c	80.688492	56.505550
	3.516609	3.610363
a	122.216542	95.949950
	3.656929	3.745537
c^2	-38.845045	-76.273126
	0.377275	1.339883
ac	-83.049804	-160.563765
	0.759619	2.696105
a^2	-40.638468	-77.181161
	0.390953	1.390827
c^3	-0.434736	0.885162
	0.105785	0.114250
a^2c	-1.934940	2.210794
	0.321363	0.346461
ac^2	-1.534567	2.520752
	0.317957	0.343230
a^3	-0.614528	0.802044
	0.109428	0.117710
c^4		0.950453
		0.031204
a^3c		4.059271
		0.126792
a^2c^2		5.961552
		0.188265
ac^3		3.888539
		0.124868
a^4		1.028033
		0.032210

Table 10: Additive vs. non-additive cubic and fourth-order models in cohort and age

Pairwise comparison of coefficients multiplied by 100

	Model 3.4	*Model 3.0*	*Model 3.1*	*Model 4.4*	*Model 4.0*	*Model 4.1*
c	80.688	76.725	84.154	56.506	52.549	86.107
a	122.217	112.836	120.780	95.950	87.286	121.089
c^2	-38.845	-38.874	1.853	-76.273	-76.452	3.562
ac	-83.050	-82.959		-160.564	-158.559	
a^2	-40.638	-40.532	1.615	-77.181	-74.995	3.736
c^3	-0.435	-0.476	-0.049	0.885	0.852	-0.055
a^2c	-1.935	-1.291		2.211	2.699	
ac^2	-1.535	-1.291		2.521	2.699	
a^3	-0.615	-0.257	0.162	0.802	1.078	0.167
c^4				0.950	0.956	-0.002
a^3c				4.059	3.824	
a^2c^2				5.962	5.736	
ac^3				3.889	3.824	
a^4				1.028	0.943	-0.005

References

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.

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.

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.