Age-period cohort models

Zoë Fannon & Bent Nielsen

28 November 2018

Summary

Outcomes of interest often depend on the age, period, or cohort of the individual observed, where *cohort* and *age* add up to *period.* An example is consumption: consumption patterns change over the life-cycle (age) but are also affected by the availability of products at different times (period) and by birth cohort-specific habits and preferences (cohort). Age-period-cohort (APC) models are additive models where the predictor is a sum of three time effects, which are functions of age, period and cohort, respectively. Variations of these models are available for data aggregated over age, period, and cohort, and for data drawn from repeated cross-sections, where the time effects can be combined with individual covariates.

The age, period and cohort time effects are intertwined. Inclusion of an indicator variable for each level of age, period, and cohort results in perfect collinearity, which is referred to as "the age-period-cohort identification problem". Estimation can be done by dropping indicator variables. However, this has the adverse consequence that the time effects are not individually interpretable and inference becomes complicated. These consequences are avoided by decomposing the time effects into linear and non-linear components and noting that the identification problem relates to the linear components, whereas the non-linear components are identifiable. Thus, confusion is avoided by keeping the identifiable non-linear components of the time effects and the unidentifiable linear components apart. A variety of hypotheses of practical interest can be expressed in terms of the non-linear components.

Contents

Introduction

Age-period-cohort (APC) models are commonly used when individuals or populations are followed over time. In economics the models are most frequently used in labour economics and analysis of savings and consumption, but are also relevant to health economics, migration, political economy, and industrial organisation, among other sub-disciplines. Elsewhere the models are used in cancer epidemiology, in demography, in sociology, in political science, and in actuarial science. The models involve three time scales for age, period and cohort, which are linearly interlinked since the calendar period is the sum of the cohort and the age.

The APC time scales are typically measured discretely but can also be measured continuously. They can have various interpretations. The cohort often refers to the calendar year that a person is born, but it could also refer to the year an individual enters university or the year that a financial contract is written. The age is then the follow-up time since the birth, entry to university, or the signing of the contract. Period is the sum of the two effects, i.e. the point in calendar time at which follow-up occurs. Together the three APC time scales constitute two time dimensions that are tracked simultaneously.

APC data can take many shapes. Data may be recorded at the individual level in repeated cross sections, where age and time of recording (period) are known for each individual. It could be panel data, where for each individual age progresses with time (period). Data could be aggregated at the level of age, period and cohort. The empirical illustration in this chapter is concerned with US employment data aggregated by age and period, see Tables 2 and 3, so that the first entry in Table 2 indicates that 5.246 million 15-19 year olds were in the labour force in 1960. For this data, questions about age would consider the unemployment rates across different age groups while questions about period would relate to changes in the overall economy. A question about cohort effects might be whether workers entering the labour force during boom years face different unemployment rates to those entering during bust years.

APC models will have many different appearances depending on the data and the question at hand. At the core of the models is a linear predictor of the form

$$\mu_{age,coh} = \alpha_{age} + \beta_{per} + \gamma_{coh} + \delta. \qquad (1)$$

This is a non-parametric model for APC which is additively separable in the three time scales, age, per, and coh. Thus, the time effects, α_{age}, β_{per} and γ_{coh}, are functions of the respective time indices. The right-hand side of (1) has a well-known identification problem in that linear trends can be added to the period effect and subtracted from the age and cohort effect without changing the left hand side of (1). The time effects can be decomposed into linear and non-linear parts. Due to the identification problem the linear parts from the three APC effects cannot be disentangled. However, the non-linear parts are identifiable. As an example, suppose the age effect is quadratic

$$\alpha_{age} = \alpha_c + \alpha_\ell \times age + \alpha_q \times (age)^2; \qquad (2)$$

then $\alpha_c + \alpha_\ell \times age$ is the non-identifiable linear part and $\alpha_q \times (age)^2$ is the identifiable non-linear part.

Note that the identification problem is concerned with the right hand side of (1) in that different values of the time effects on the right hand side result in the same predictor on the left hand side. The premise for this feature is that the left hand side predictor is identifiable and estimable in reasonable statistical models. This highlights that the crucial aspect of working with APC models is to be clear about what can and cannot be learned.

In economics a common type of data is the repeated cross section with a continuous outcome variable. Such data could be modelled as follows. Suppose the observations for each individual are a continuous dependent variable Y_i and a vector of regressors Z_i, as well as age_i and coh_i for $i = 1, \ldots, N$. A simple regression model has the form

$$Y_i = \zeta' Z_i + \mu_{age_i,coh_i} + \varepsilon_i, \qquad (3)$$

where the APC predictor μ_{age_i,coh_i} is given in (1) and ε_i is a least square error term. The identification problem from (1) is embedded in regression (3). The appropriate solution to this problem depends on what the investigator is interested in. If the primary interest is the parameter ζ the problem can simply be addressed by restricting four of the time effect parameters to be zero, such as

$$\alpha_1 = \beta_1 = \beta_2 = \gamma_1 = 0. \qquad (4)$$

This restriction to the time effects is just-identifying and therefore untestable. The just-identified linear trends do not have any interpretation outside the context of the restriction (4), which makes it difficult to interpret results and draw inferences. The issue, and the reason that (4) does not solve the problem, is that the investigator could just as well have imposed that

$$\alpha_1 = \alpha_2 = \beta_1 = \gamma_1 = 0, \qquad (5)$$

resulting in time effects with very different appearance, see Figures 1 and 2 below. To appreciate the APC identification problem one has to go back to the original formulation (1) and ask if any inference drawn would be different if imposing (5) instead of (4). If there is a difference one has to be careful.

The identification problem has generated an enormous literature where solutions fall in three broad categories. The traditional approach is to identify the time effects by introducing non-testable constraints on the linear parts of time effects which are in principle like (4) or (5) (Hanoch & Honig 1985). A second approach is to abandon the APC model and either use graphs to get an impression of time effects (Meghir & Whitehouse 1996, Voas & Chaves 2016) or replace the time effects in the model with other variables (Heckman & Robb 1985). Finally, a more recent approach reparametrizes the model in terms of invariant, non-linear parts of the time effects (Kuang & al. 2008a). The latter approach clearifies the inferences that can be drawn from APC models.

It is possible to characterize precisely which questions can and cannot be addressed by APC models. Questions that can be addressed include any question relating to the linear predictor $\mu_{age,coh}$ on the left hand side of (1). This is valuable in forecasting. For instance, if it is of interest to forecast the resources needed for schools an APC model can be fitted to data for counts of school children at different ages and the predictor can then be extrapolated into the future. A different type of question may be how consumption changed from 2008 to 2009 as compared to how it changed from 2007 to 2008 so as to measure the effect of the financial crisis. This question is concerned with differences-in-differences and is identifiable from the non-linear parts of the time effects. Note that a consequence of the model is that this change in consumption affects all cohorts in the same way. If one suspects that different cohorts are differently affected an interaction term would be needed in model (1). Conversely, the questions that cannot be addressed by APC models can also be characterized. These are questions that relate to levels or slopes of the time effects. In the context of the quadratic age example (2) the level and slope are α_c and α_ℓ, respectively.

There are a variety of applications in economics for which APC modelling can be useful. In any setting where the passage of time is an explanatory factor, there is a risk of confused interpretation due to the APC problem. This has been recognised in studies of labour market dynamics (Hanoch & Honig, 1985; Heckman & Robb, 1985; Krueger & Pischke, 1992; Fitzenberger & al., 2004), life-cycle saving and growth (Deaton & Paxson, 1994a), consumption (Attanasio, 1998; Deaton & Paxson, 2000; Browning & al., 2016), migration (Beenstock & al., 2010), inequality (Kalwij & Alessie, 2007), and structural analysis (Schulhofer-Wohl, 2018). Yang & Land (2013) and O'Brien (2015) describe examples in criminology, epidemiology, and sociology.

The risk of confusion due to the identification problem is avoidable. For example, McKenzie (2006) exploits the non-linear discontinuity in consumption with respect to period to evaluate the impact of the Mexican peso crisis. Ejrnæs & Hochguertel (2013) are not directly interested in the time effects and so can use an ad-hoc identified APC model to control for time in their investigation of the effect of unemployment insurance on the probability of becoming unemployed in Denmark.

However, where the research question involves the linear part of a time effect, any attempt to answer this directly must involve untestable restrictions on the linear parts of other time effects. In this context the risk of confounding between time effects cannot be mitigated. One solution is to reformulate the question in terms of the non-linear parts of the time effects. Certain difference-in-difference questions naturally take this form, see for example McKenzie's (2006) analysis of the peso crisis. Otherwise, the researcher's only option is to argue for untestable restrictions using economic theory. Such restrictions may be explicit as in (4) or (5) or implicit if time effects are replaced with a proxy variable (Krueger & Pischke, 1992; Deaton and Paxson, 1994b; Attanasio, 1998; Browning & al., 2016).

The risks of confounding inherent in models involving any of age, period, or cohort can be avoided by beginning with a general model that allows for any possible combination of time effects, then gradually reducing the model by imposing testable restrictions. There is substantial scope for such testable restrictions: exclusion and functional form restrictions on the non-linear parts of each of age, period, and cohort can be tested, as can the replacement of time effects by proxy variables.

The remainder of the chapter elaborates on the main points raised above. The identification problem is explained in greater detail. A number of approaches taken to resolve or avoid the identification problem are discussed, including several variants of the traditional approach and the recent reparameterization. Interpretation of the parameters of the APC model is discussed. The idea of sub-models, which provide a systematic guide to testable reductions of the APC model, is introduced. There is some discussion of "hidden" identification problems, which can arise when the initial model is insufficiently general. This is followed by a section explaining the sorts of problems that the APC model is well-equipped to address. The final section contains a more detailed discussion of statistical models for APC analysis and an empirical illustration.

Background

Elements of the conceptual framework used in subsequent formalized discussions of APC models are introduced. In particular, the recording of time is discussed, the types of data structures for which APC models are used are described, and vector notation is defined.

Time

In reality time is recorded discretely in units such as years, days, or seconds. Throughout this discussion it is assumed that the time index is positive. The traditional calendar convention is adopted whereby there is no year zero and time is rounded up to the nearest whole number of units, rather than the time stamp method which has a year zero and where time is rounded down to the nearest whole number of units. Suppose a given sample has single-year units. Then $age = 1$ is assigned to the youngest person and $coh = 1$ is assigned to the earliest recorded birth year. This leads to the relation

$$age + coh = per + 1. \tag{6}$$

Typically only two of the three time scales, age, per and coh, are recorded. This leads to a slight inaccuracy due to mid-year birthdays. When age and period are the recorded values this is referred to as the problem of overlapping cohorts. Osmond & Gardner (1989) show that it does not matter for the identification problem whether two or three time scales are recorded. Carstensen (2007) shows how to handle the additional information from a third recorded time scale.

Data array

A range of data structures appear in the literature. The main types are age-period (AP) arrays, a common format for repeated cross sections; period-cohort (PC) arrays, used in prospective cohort studies; and age-cohort (AC) arrays. In 1875 Lexis referred to these arrays as the principal sets of death (Keiding, 1990). Another common data array is the age-cohort triangle used for reserving in general insurance (England & Verrall, 2002). The different data arrays can be unified by thinking of them as instances of generalized trapezoids (Kuang & al., 2008a) defined by the index set

$$\mathcal{J} = (1 \leq age \leq A \quad \text{and} \quad 1 \leq coh \leq C \quad \text{and} \quad L + 1 \leq per \leq L + P), \tag{7}$$

where L is a period offset. The age-period array arises when $L = A - 1$ and $L + P = C$ while the age-cohort array has $L = 0$ and $P = A + C - 1$. From a geometric view point it is useful to consider all of these in an age-cohort coordinate system, due to the symmetry of age and cohort in equation (6). This convention is followed henceforth.

Vector notation

The time effect equation (1) has the linear predictor $\mu_{age,coh}$ on the left hand side. It varies on a surface indexed by age and cohort and where the shape is given by the combination of the time effects, α_{age}, β_{per}, and γ_{coh}. Stacking the linear predictors as a vector gives

$$\mu = (\mu_{age,coh})_{age,coh \in \mathcal{J}}, \tag{8}$$

which has dimension n, so that $n = AC$ for an AC array and $n = AP$ for an AP array, and where $\mathcal{J}$ refers to an index set of the form (7).

Collecting the time effects on the right hand side of (1) gives the vector

$$\theta = (\alpha_1, \dots, \alpha_A, \beta_{L+1}, \dots, \beta_{L+P}, \gamma_1, \dots, \gamma_C, \delta)', \tag{9}$$

of dimension $q = A + P + C + 1$. Thus the model (1) implies that the n-vector μ in (8) varies on a surface with dimension of at most q. When n is not too small the surface for μ is estimable so that μ can be identified up to sampling error. The APC identification problem is that the time effects are collinear, so that not all components in the q-vector θ are identified.

The identification problem, explained

The identification problem arising in the linear parts of the time effects is formally defined and illustrated in a simplified linear model.

Formal characterization

In equation (1) the predictor $\mu_{age,coh}$ is identifiable from data whereas the time effects on the right hand side of equation (1) are only identifiable up to linear trends. Indeed, the equation can, for any *a,b,c,d,* be rewritten as

$$\mu_{age,coh} = (\alpha_{age} + a + d \times age) + (\beta_{per} + b - d \times per)$$
$$+(\gamma_{coh} + c + d \times coh) + (\delta - a - b - c - d). \tag{10}$$

Since the four quantities *a,b,c,d* are arbitrary, only a $p = q - 4$ dimensional version of θ is estimable. The equation (10) also shows that the time effects, such as the age effect α_{age}, are only discoverable up to an arbitrary linear trend. It is therefore possible to learn about the non-linear part of the age effect only. The non-linearity captures the shape of the age effect which can be expressed through second and higher derivatives. The unidentified linear parts of the time effects combine to form a shared identifiable linear plane, which is explored in the next subsection. The unidentifiability of the linear components has a number of consequences with respect to interpretation, count of degrees of freedom, plotting, inference and forecasting.

In the literature the identification problem has been addressed in various ways. A popular approach is to impose four (or sometimes more) constraints on the time effects. This addresses the immediate problem of estimating a version of the time effects, but leaves the problems outlined above. A more recent approach, which will be presented first, is to parametrize the predictor in terms of elements of the time effect that are invariant to the transformations given in (10). This approach clarifies what can be learned from the model.

Illustration in a simple case: the linear plane model

The linear plane model is the simplest model where the APC identification problem is present. It arises when all the time effects are assumed to be linear. For instance, the age effect is parametrized as $\alpha_{age} = \alpha_c + \alpha_\ell \times age$, where α_c is a constant level and α_ℓ is a linear slope. Combining the three linear time effects results in

$$\mu_{age,coh} = (\alpha_c + \alpha_\ell \times age) + (\beta_c + \beta_\ell \times per) + (\gamma_c + \gamma_\ell \times coh) + \delta. \tag{11}$$

This model involves seven parameters but only a three-dimensional combination is identified due to the transformations in (10).

It is tempting to restrict the four intercepts in (11) and the three slopes to get a single intercept and two slopes by imposing constraints. This will not change the range of the predictor on the left hand side of (11) but it will change the interpretation of the unidentified time effects on the right hand side. Two researchers choosing different restrictions may end up drawing different inferences about the time effects if this is not kept in mind.

Model (11) implies that the predictor varies on a linear plane. A linear plane can be parametrized in many ways. For instance the plane could be parametrized in terms of age and cohort slopes anchored at $age = coh = 1$ as in

$$\mu_{age,coh} = \mu_{11} + (\mu_{21} - \mu_{11})(age - 1) + (\mu_{12} - \mu_{11})(coh - 1). \quad (12)$$

Equally, it could be parametrized in terms of age and period slopes using (6) as in

$$\mu_{age,coh} = \mu_{11} + (\mu_{21} - \mu_{12})(age - 1) + (\mu_{12} - \mu_{11})(per - 1). \quad (13)$$

The parametrizations (12), (13) both identify the variation of the predictor on the left hand side of (11). However, the slopes in (12) and (13) do not identify the slopes of the time effects. The age slopes in (12) and (13) are different and satisfy, within the linear plane model, $\mu_{21} - \mu_{11} = \alpha_\ell + \beta_\ell$ and $\mu_{21} - \mu_{12} = \alpha_\ell - \gamma_\ell$ respectively; evidently, neither is equal to α_ℓ.

The equation (12) parametrises the linear plane without reference to time effects. Time effects can only be identified by imposing restrictions on these. The constraint (4) is equivalent to $\alpha_c = \beta_c = \gamma_c = \beta_\ell = 0$, in the linear plane model (11). With this constraint identification is achieved in that $\mu_{21} - \mu_{11} = \alpha_\ell$ and $\mu_{21} - \mu_{12} = -\gamma_\ell$ and $\mu_{11} = \delta$. This identification gives a model in terms of age and cohort time effects. By imposing the constraint (5) a model in terms of period and cohort time effects could be obtained, and a similar set of constraints would result in a model in terms of age and period slopes. Each set of constraints appears to lead to information about the time effects, but clearly they cannot all be correct. In fact it is not possible to establish if any of these three sets of constraints lead to a correct impression of the unidentifiable time effects. Although the time effects cannot be identified it is still possible to answer any question that relates to the predictor $\mu_{age,coh}$, such as forecasting future values or testing for change in $\mu_{age,coh}$.

As a numerical example of the identification issue, suppose the linear plane (12) is

$$\mu_{age,coh} = 1 + 3(age - 1) + (coh - 1) \quad (14)$$

over an AC array with $A = C = 10$. The linear plane (14) does not specify the time effects and the over-parametrized time effect specification (11) cannot be identified.

Suppose it is not known that the data is generated by (14), but it is known that a model of the form (11) generated the data. Applying the constraints (4) and (5) to the model (11) in the context of the data-generating process (14) results in the slopes $\alpha_\ell = 3$, $\beta_\ell = 0$, $\gamma_\ell = 1$ and $\alpha_\ell = 0$, $\beta_\ell = 3$, $\gamma_\ell = -2$ respectively, as illustrated in Figures 1 and 2.

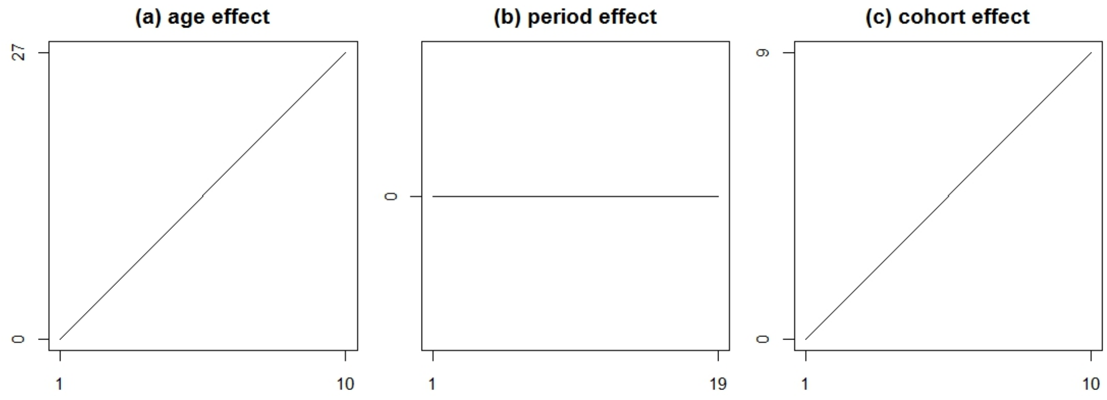

Figure 1. Time effect slopes under identification (4).

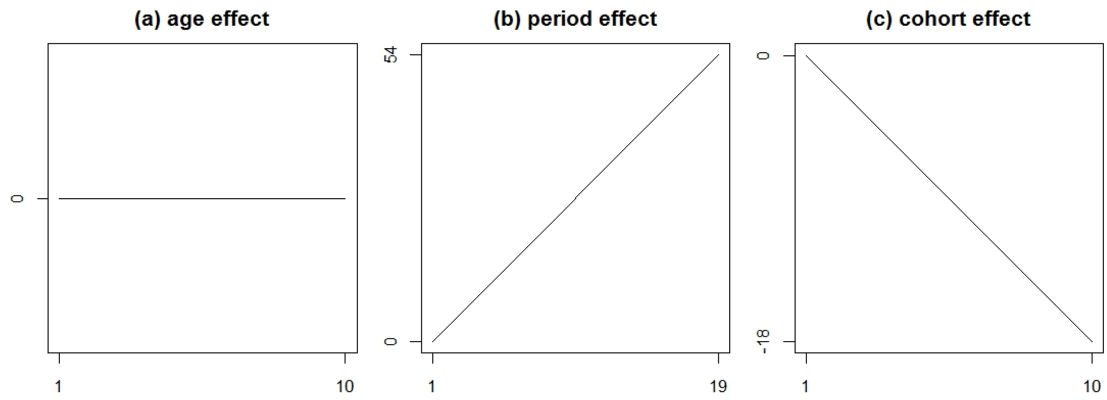

Figure 2. Time effect slopes under identification (5) for the same linear plane as in Figure 1.

The Figures 1 and 2 have a rather different appearance despite generating exactly the same linear plane. Three features are important. First, the signs of the slopes are not identified. The cohort effect is upward sloping in Figure 1(c) and downward sloping in Figure 2(c). Second, the units of the time effects have no meaning. The period scale is not defined in Figure 1(b) whereas it is defined in Figure 2(b). Further, the units of the cohort scales are very different in Figures 1(c) and 2(c) which have slopes of 1 and -2, respectively, yet they are observationally equivalent. Third, a subtler feature is that within each Figure the sub-plots are interlinked. For example, by setting the period slope to zero in Figure 1(b) the cohort slope in Figure 1(c) becomes upward sloping. But where the age slope is set to zero in Figure 2(a) the period is upward sloping in Figure 2(b) while the cohort is downward sloping in Figure 2(c). Thus, it is not possible to draw inferences from any sub-plot in isolation. This is a serious limitation in practice as the eye tends to focus on one sub-plot at a time.

Addressing the identification problem

An overview is given of the some of the most commonly encountered identification strategies in the APC literature. Each of three categories of solutions – identification by restriction, forgoing the formal APC model, and the canonical parametrization – is considered.

What to look for in a good approach

There are many proposed solutions and identification strategies in the literature on APC modelling, across several disciplines. This section provides guidance on assessing such identification strategies.

Invariance

It has long been recognized that it is useful to work with functions of the time effects that are invariant to the transformations in (10). Thus, there are some parallels to the theory for invariant reduction of statistical models (Lehman, 1986, §6; Cox & Hinkley, 1974, §5.3). In that vein Carstensen (2007) interpreted equation (10) as a group g of transformation from the collection of time effects θ in (9) to the collection of predictors μ in (8). Invariant functions of θ, say $f(\theta)$ are invariant if $f\{g(\theta)\} = f(\theta)$.

Double differences of the time effects are invariant (Fienberg & Mason, 1979; Clayton & Schifflers, 1987; McKenzie, 2006). To see this consider the double differenced age effect:

$$\Delta^2\alpha_{age} = \Delta\alpha_{age} - \Delta\alpha_{age-1} = \alpha_{age} - 2\alpha_{age-1} + \alpha_{age-2}. \tag{15}$$

Equation (10) shows that for any non-zero a, d the age effects α_{age} and $\alpha_{age} + a + d \times age$ are observationally equivalent but can differ substantially in value; this was demonstrated in Figures 1, 2. Now, the double differences of α_{age} and $\alpha_{age} + a + d \times age$ are both $\Delta^2\alpha_{age}$, which does not depend on a, d and is therefore invariant to the transformations in (10). In the context of the quadratic example (2) it can be shown that $\Delta^2\alpha_{age} = 2\alpha_q$. The double differences have an odds-ratio or difference-in-difference interpretation, that will be explored later, see (22).

The predictor $\mu_{age,coh}$ is also invariant (Kuang & al., 2008a). Indeed equation (10) shows that any transformation of that form applied to the time effects on the right hand side of (1) results in the same predictor. However, $\mu_{age,coh}$ alone may not be of great interest. The next step is therefore to represent the predictor μ exclusively in terms of invariant functions $\xi(\theta)$. That is, the desired outcome is to express μ as a bijective function of $\xi(\theta)$, where ξ is invariant so that $\xi(\theta) = \xi(g(\theta))$. The function ξ is then a maximal invariant and useful for parametrization of the model as it carries as much of the intended information from the time effects as possible while being invariant to the identification problem.

In the context of exponential family models, such as the linear model in (3) or logit or Poisson regressions, the predictor $\mu_{age,coh}$ enters linearly in the log-likelihood. If the maximal invariant parameter ξ is a linear function of the time effects and varies freely in an open parameter space then the exponential family model is regular with ξ as canonical parameter (Barndorff-Nielsen, 1978, §8). Such a canonical parameter is explicitly defined later.

Sub-sample analysis

An alternative way to think about invariance is sub-sample analysis. It is relevant in two ways. First, it can be used to check a claim that a particular identification strategy avoids the identification problem. Second, it can be used for specification testing in a practical analysis.

Suppose it is claimed that a proposed method for estimating the age effect or some structural parameter avoids the identification problem. In many cases it can be argued that the method should be, apart from estimation error, invariant to the choice of data array.

Specifically, suppose a data array $\mathcal{J}$ of the form (7) is available. A sub-set $\mathcal{J}'$ can be formed in various ways, for instance, by considering those age groups younger than some threshold A'. The claim that the method avoids the identification problem is then substantiated if the method gives the same result when applied to the full data array $\mathcal{J}$ and to the sub-set data array $\mathcal{J}'$.

Whatever method is applied, the specification of an estimated model can be checked by recursive analysis following common practice in time series analysis. The idea is to track the estimates of invariant parameters for different sub-sets $\mathcal{J}'$ with different choices of threshold A' and plotting these against the threshold values, following Chow (1960). Investigators can check the specification of models by recursive modelling along the three time scales. For a well-specified model those estimates should not vary substantially with the threshold apart from minor variation due to estimation error. Larger variation is indicative of structural breaks in the data generating process and calls for a more flexible model than (1).

Canonical Parametrization

Overview

The time effects can be decomposed into a linear and a non-linear part. The non-linear parts can be represented in terms of the double differences such as $\Delta^2\alpha_{age}$, introduced in (8). The linear parts of the three time effects along with the intercept in equation (1) combine to a linear plane where the slopes are identifiable (Holford, 1983). Combining these ideas the predictor can be given representations of the form

$$\mu_{age,coh} = linear\ plane + \Sigma\Sigma^{age}\Delta^2\alpha_s + \Sigma\Sigma^{per}\Delta^2\beta_s + \Sigma\Sigma^{coh}\Delta^2\gamma_s. \qquad (16)$$

The exact specification of the linear plane and the summation indices for the double sums of double differences depend on the index array for the age, period and cohort indices. Note that the linear terms are simply kept together as a linear plane without attempting to disentangle them into APC components. This circumvents the unsolvable identification problem.

Age-cohort index arrays

Kuang et al. (2008a) consider AC index arrays and show

$$\mu_{age,coh} = \mu_{1,1} + (\mu_{2,1} - \mu_{1,1}) \times (age - 1) + (\mu_{1,2} - \mu_{1,1}) \times (coh - 1)$$
$$+ \textstyle\sum_{t=3}^{age}\sum_{s=3}^{t}\Delta^2\alpha_s + \sum_{t=3}^{per}\sum_{s=3}^{t}\Delta^2\beta_s + \sum_{t=3}^{coh}\sum_{s=3}^{t}\Delta^2\gamma_s, \qquad (17)$$

with the convention that empty sums are zero. Here the linear plane has been parametrized as in (12). The plane is identified as it is invariant to the transformations (10) but the time effect slopes remain unidentified since the age, period, and cohort slopes remain interlinked, see (12), (13). A feature of the representation (17) is that the non-linear components are separated from the linear plane. The predictor in (17) can be summarised as $\mu_{age,coh} = \xi' x_{age,coh}$ where

$$\xi = (\mu_{11}, \mu_{21} - \mu_{11}, \mu_{12} - \mu_{11}, \Delta^2\alpha_3, \ldots, \Delta^2\alpha_A, \Delta^2\beta_3, \ldots, \Delta^2\beta_{A+C-1}, \Delta^2\gamma_3, \ldots, \Delta^2\gamma_C)' \qquad (18)$$

The design vector $x_{age,coh}$ is defined in terms of a function $m(t,s) = \max(t - s + 1, 0)$ as

$$x_{age,coh} = \{1, age - 1, coh - 1, m(age,3), \ldots, m(age,A),$$
$$m(per,3), \ldots, m(per,P), m(coh,3), \ldots, m(coh,C)\}' \qquad (19)$$

Theorem 1 of Kuang, Nielsen, & Nielsen (2008a) shows that ξ is a maximal invariant with respect to the transformations in (10) as it is composed of double differences and values of the predictor itself. The parameter ξ will be canonical in the context of exponential family models such as normal, logistic or Poisson regressions.

General index arrays including age-period arrays

General index arrays (7) are considered by Nielsen (2015). They exclude the age-cohort triangle where $age + coh \leq L$ and hence the anchoring point $age = coh = 1$ in (17) when $L \geq 1$. This will be the case for age-period arrays where $L = A - 1$. Thus, to achieve a unified representation that preserves the age-cohort symmetry the anchoring point is chosen in the middle of the first or second period diagonal, depending on whether L is even or odd. To find this point, the quantity $U = int\{(L+3)/2\}$ is defined and the anchoring point is chosen to be at $age = coh = U$. Note that $age = coh = U$ gives $per = 2U - 1$ and that $L + 3 = 2U$ for odd L while $L + 3 = 2U + 1$ for even L. The representation is then

$$\begin{aligned}\mu_{age,coh} = {} & \mu_{UU} + (\mu_{U+1,U} - \mu_{UU}) \times (age - U) + (\mu_{U,U+1} - \mu_{UU}) \times (coh - U) \\ & + 1_{(age<U)} \textstyle\sum_{t=age+2}^{U+1} \sum_{s=t}^{U+1} \Delta^2 \alpha_s + 1_{(age>U+1)} \sum_{t=U+2}^{age} \sum_{s=U+2}^{t} \Delta^2 \alpha_s \\ & + 1_{(L\, odd\, \&\, per=2U-2)} \Delta^2 \beta_{2U} + 1_{(per>2U)} \textstyle\sum_{t=2U+1}^{per} \sum_{s=2U+1}^{t} \Delta^2 \beta_s \\ & + 1_{(coh<U)} \textstyle\sum_{t=coh+2}^{U+1} \sum_{s=t}^{U+1} \Delta^2 \gamma_s + 1_{(coh>U+1)} \sum_{t=U+2}^{coh} \sum_{s=U+2}^{t} \Delta^2 \gamma_s \,. \end{aligned} \tag{20}$$

The representation (17) for age-cohort arrays arises in the special case where $L = 0$, implying $U = 1$. The general representation (20) can also be written as $\mu_{age,coh} = \xi' x_{age,coh}$ where

$$\begin{aligned}\xi = {} & (\mu_{UU}, \mu_{U+1,U} - \mu_{UU}, \mu_{U,U+1} - \mu_{UU}, \\ & \Delta^2 \alpha_3, \ldots, \Delta^2 \alpha_A, \Delta^2 \beta_{L+3}, \ldots, \Delta^2 \beta_{L+P}, \Delta^2 \gamma_3, \ldots, \Delta^2 \gamma_C)'. \end{aligned} \tag{21}$$

The design vector is defined in terms of the function $m(t,s) = \max(t - s + 1, 0)$ as

$$\begin{aligned}x_{age,coh} = {} & \{1, age - U, coh - U, \\ & m(1, age), \ldots, m(U-1, age), m(age, U+2), \ldots, m(age, A), x^{\beta}_{age,coh}, \\ & m(1, coh), \ldots, m(U-1, coh), m(coh, U+2), \ldots, m(coh, C)\}, \end{aligned} \tag{22}$$

where the period part $x^{\beta}_{age,coh}$ depends on whether L is even or odd:

$$x^{\beta}_{age,coh} = \begin{cases} 1_{(per=L+1)}, m(per, L+4), \ldots, m(per, L+P) & when\ L\ is\ odd, \\ m(per, L+3), \ldots, m(per, L+P) & when\ L\ is\ even. \end{cases} \tag{23}$$

This canonical parametrization captures all the identifiable variation in the predictor due to the time effects. The interpretation of the elements of ξ is discussed in a subsequent section.

Identification by restriction

The traditional approach to identification is to introduce restrictions of the type (4) and (5). Such restrictions give a parametrization that is not invariant to the transformations in (10). This leads to the kind of issues highlighted with Figures 1,2. The purpose of the restrictions is essentially to extract some version of the linear parts of the time effect from the linear plane. The linear plane only has one level and two slopes as seen in (12). There is no unique way to distribute these quantities on the three time effects. Various approaches have been suggested in the literature, some of which are reviewed below. Typically these approaches have two steps, where the levels are identified at first and then the linear slope is identified. This makes a formal analysis complicated, see Nielsen & Nielsen (2015).

Restrictions on levels

There are two main approaches to identifying the level: restricting particular coordinates of the time effects, or restricting the average level of the time effect. Neither approach is invariant to the transformations in (10).

Restricting coordinates of the time effects. A common restriction is to set individual coordinates of the time effects to zero as in (4) and (5). Ejrnæs & Hochguertel (2013) provide an example. In practice this works by first including a full set of APC dummies and then dropping the dummies where it is intended that time effects be set to zero. Such restrictions are not invariant to the transformations in (10). Indeed, the requirement $\alpha_1 = 0$ is violated when adding some non-zero number a to α_1. With this approach it is possible to ensure comparability between estimates for sub-samples as long as exactly the same restriction is imposed.

Restricting the average levels. A common restriction is to set the average of the time effects to zero so that $(1/A)\Sigma_{age=1}^{A}\alpha_{age} = (1/P)\Sigma_{per=L+1}^{L+P}\beta_{per} = (1/C)\Sigma_{coh=1}^{C}\gamma_{coh} = 0$. The level of the model is then picked up by the intercept δ in (1). This restriction is commonly used (Deaton & Paxson, 1994a; Schulhofer-Wohl, 2018). A feature of this type of restriction is that the unidentified levels and slopes are orthogonalized, but this comes at the cost of making the scale of the time effects dependent on the dimensions of the index array (7). As before, the zero average restriction is not invariant to the transformations in (10). Indeed, increasing all age effects by some non-zero number a violates the restriction.

Figures 3 and 4 applies this restriction to the plane (14) and demonstrates that the restriction is specific to the index array through a sub-sample argument. AC index arrays are chosen so that Figure 3 has $A = C = 10$ while Figure 4 has $A = C = 5$. In both figures the average level is set to zero while the period slope is set to zero as in (Deaton & Paxson, 1994a). Note that the absolute ranges for age (28) and cohort (10) are the same as in Figure 1. The intercepts are very different with $\delta = 19$ and $\delta = 9$, respectively. Further, the time effects are not comparable, for instance, $\alpha_{5.5} = 0$ in Figure 3, whereas $\alpha_3 = 0$ in Figure 3. Arguing, *ad absurdum*, the sub-sample analysis implies that by varying the data array while keeping the zero level constraint the time effects must be zero.

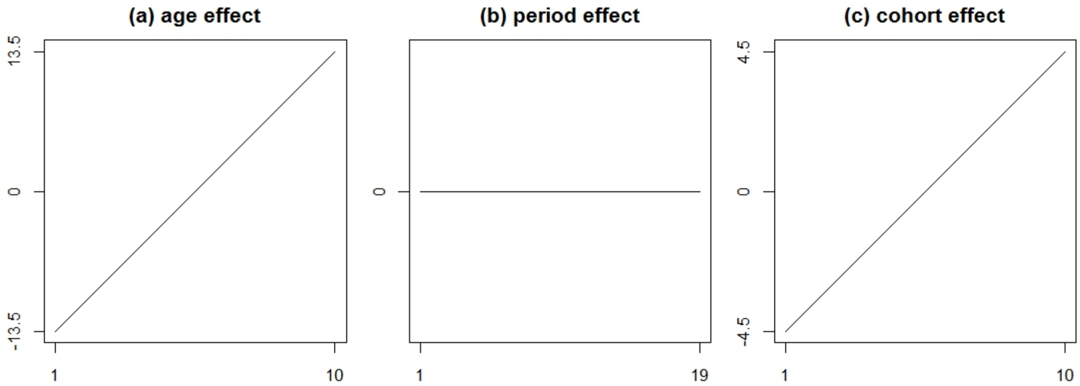

Figure 3. Time effect slopes under average level identification
For an AC array with $A = C = 10$

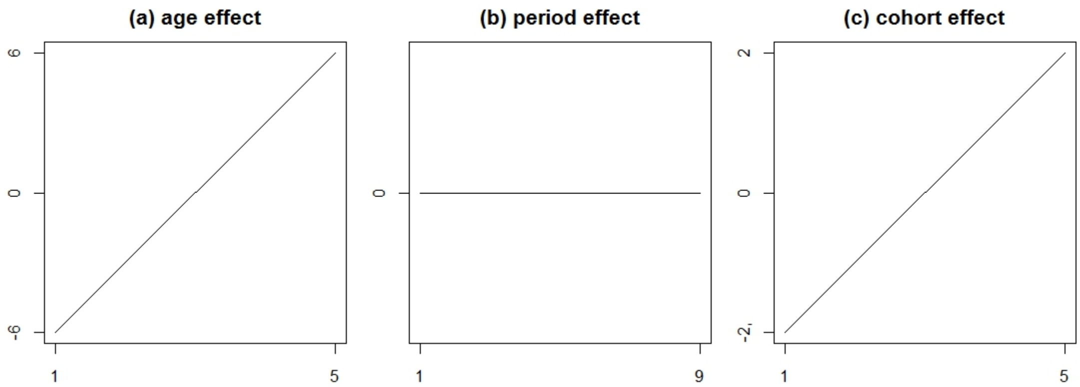

Figure 4. Time effect slopes under average level identification
for an AC array with $A = C = 5$

The APC slopes are the same in Figures 3 and 4. This is not a general feature of the zero average restriction but a consequence of working with a linear plane predictor of the form (14). To illustrate this point introduce a non-linear effect into (14) to get

$$\mu_{age,coh} = 1 + 3(age - 1) + 2(coh - 1) + 1_{(per \geq 10)}. \tag{24}$$

On the smaller AC array with $A = C = 5$ this reduces to the linear plane in (14) so that for zero average levels and a zero period slope Figure 4 emerges. On the larger AC array with $A = C = 10$ the non-linearity matters. Keeping the zero average level constraint and setting the period slope to zero through $\Sigma_{per=1}^{19} per \times \beta_{per} = 0$ results in Figure 5. Comparing Figures 4 and 5 it is seen that all slopes are different. The age slope are 3 and 3.02, respectively, and the cohort slopes are 1 and 1.02 respectively. The period slopes for $per \leq 9$ are zero and -0.08, respectively.

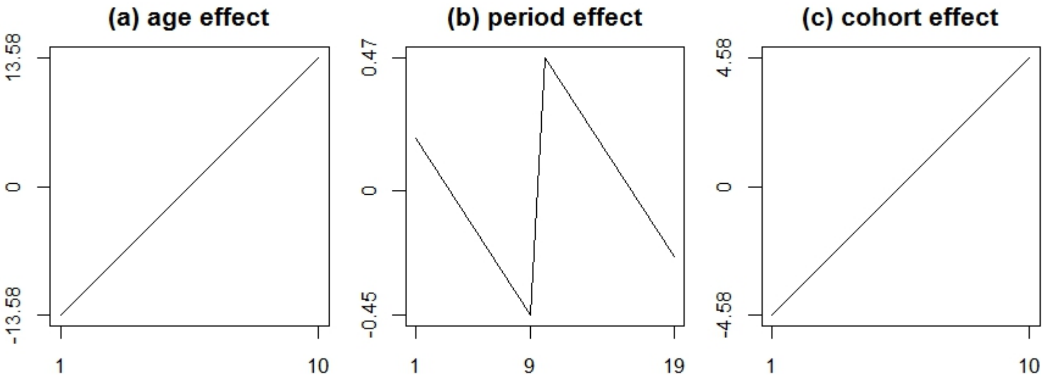

Figure 5. Time effect slopes for (24) under average level identification and the slope constraint $\Sigma_{per=1}^{19} per \times \beta_{per} = 0$ for an AC array with $A = C = 10$.

Restrictions on slopes

Once the level is attributed between the time effects and the intercept, the slopes have to be restricted. This approach necessarily binds the slopes of the three time effects together. Graphically, this can have dramatic consequences as seen in Figures 1 and 2.

Restricting a pair of adjacent time effects. The slope can be identified by restricting a pair of adjacent time effects to be equal. An example would be to let $\beta_1 = \beta_2$ as in (4). Fienberg & Mason (1979) propose this method combined with a zero average restriction. This restriction is not invariant to (10). Indeed, adding a linear trend with non-zero slope d to the age effect violates the restriction.

Orthogonalizing a time effect with respect to a time trend. A more complicated version of the previous approach is to pin down one of time effects by orthogonalization with respect to a time trend so as to constrain the slope to be zero. An example would be to require that $\Sigma_{per} per \times \beta_{per} = 0$. Deaton and Paxson (1994a) apply this approach in conjunction with an average restriction on the level of the period effect and zero restrictions of the first coordinates of the age and cohort effects. The lack of invariance is commented upon above in connection with (24) and Figures 4 and 5.

Generalized inverses. The identification problem can be thought of as a collinearity problem that can be addressed using generalized inverses. This would be implemented as follows. First a design matrix D with a full set of APC dummies is created. Zero average constraints are imposed and three columns of D are dropped to implement this constraint, leaving the selected matrix DS with a rank deficiency of one. The time effects are then estimated using least squares while applying a Moore-Penrose generalized inverse for $S'D'DS$. This method was proposed by Kupper & al. (1985) and is called the "intrinsic estimator" by Yang & al. (2004). It has been criticised by Holford (1985), O'Brien (2011) and Luo (2013). Nielsen & Nielsen (2014, Theorem 8) analyse how the estimator depends on the choice of level constraint, selection matrix S, and choice of generalized inverse.

Forgoing APC models

Some researchers take the position that since formal modelling of the linear time effects is plagued by problems of identification, the attempt to construct a statistical model which allows for all three of age, period, and cohort effects should be abandoned. Two approaches are followed: either to use a combination of graphs and discipline-specific knowledge to build a story about the time effects, or to replace the time effects with other explanatory variables.

Graphical analysis

Most research involving APC effects will include some preliminary graphical analysis of the data by age, by period, and by cohort. Some researchers believe that due to the identification problem there is little to gain by going beyond the graphical analysis. Kupper & al. (1984) were early proponents of this view. A clear articulation of the position and an illustration of how conclusions might be drawn from graphs can be found in Voas & Chaves (2016). Their Figure 1 shows trends in religious affiliation against time, which can be read as age or period, for several British cohorts. The lines are broadly parallel and horizontal, with the line for each cohort successively lower than the next. They argue that such a graph could be generated by only two models: either a model containing only cohort effects, or a model with perfectly balanced age and period effects. Since the latter is implausible they decide that the data must have been generated by the first. Meghir & Whitehouse (1996) also use this sort of graphical analysis in their analysis of wage trends.

The graphical approach can be helpful when the common features and appropriate interpretation of them are clear as they are in Voas & Chaves (2016). However, without parallel trends it is difficult to draw inferences, and of course there is no scope for formal testing.

Alternative variables

Another way of side-stepping the APC identification problem, advocated by Heckman & Robb (1985), is to reconceptualize the model. They argue that researchers are rarely interested in pure APC time effects; rather, these variables are "proxies" for the true "latent" variable of interest. Their solution is to replace one or all of age, period, and cohort with a latent variable. For example, they suggest using a physiological measure of aging in place of age and indicators reflecting macroeconomic conditions in place of period in a model for earnings.

An example of this approach is the model of life cycle demand for consumer durables in Browning & al. (2016). The idea is to retain age and cohort time effects, but replace the period time effect with a measure of the user cost of durables. This gives a sub-model of the APC model, which is analysed in the below section on sub-models. As such it is a testable restriction on the APC model. The linear period effect remains unidentifiable but is present in part as an unidentified contributor to the linear plane generated by the age and cohort time effects and in part as the linear component of the observed period variable.

Bayesian identification

Bayesian methods are also used for identification. By and large the problems are the same as with identification by restriction. Bayesian models are set up as follows. The likelihood is denoted $p(Y|\theta)$ where θ is the q-vector of time effects in (9) and Y is the data. The prior is $p(\theta)$. Decompose $\theta = (\xi, \lambda)$, where ξ is the p-dimensional canonical parameter and λ is of dimension $q - p = 4$ and represents the unidentifiable part of θ. The likelihood thus satisfies $p(Y|\theta) = p(Y|\xi)$. Now, decompose the prior $p(\xi, \lambda) = p(\xi)p(\lambda|\xi)$, so that $p(\xi)$ is the prior for the identifiable parameter and $p(\lambda|\xi)$ is the conditional prior for the unidentified parameter given the identified parameter. Finally, the posterior distribution decomposes as $p(\theta|Y) = p(\xi|Y)p(\lambda|\xi, Y)$ so, by Proposition 2 of Poirier (1998),

$$p(\xi|Y) = p(Y|\xi)p(\xi)/p(Y) \text{ and } p(\lambda|\xi, Y) = p(\lambda|\xi). \tag{25}$$

This shows that the likelihood updates the canonical parameter but cannot update any prior information about the unidentified parameter, in line with the earlier argument concerning identification by restriction. Consequences for forecasting are analysed in Nielsen & Nielsen (2014).

In Bayesian APC models it is common to choose a prior where the double differences are independent identically normal. Berzuini & Clayton (1994) suggested using a uniform prior for the linear parts of the time effects, which include the unidentified parameter λ. More recently, Smith & Wakefield (2016) have suggested a model with a prior for the linear plane, but avoiding formulation of a prior for the unidentified parameter.

Some concluding remarks on the identification problem.

To summarize, the identification problem is that the linear parts of the time effects cannot be identified because of transformations in (10). Instead, what can be identified are the non-linear parts of the time effects and a linear plane for the predictor that combines the linear parts of the time effects. In practice one has to keep these non-linear and linear features apart. The approach of identification by restriction does not achieve this, as demonstrated in Figures 1 to 5. It creates problems with interpretation, formulation of hypotheses, and counts of degrees of freedom. In contrast, the canonical parametrization keeps non-linear and linear features apart and it is therefore suitable for estimation, formulation of hypotheses, and counts of degrees freedom. The interpretation of the APC model and its elements is addressed in the next section.

Interpretation

The previous section demonstrated that non-linear and linear features of the APC model must be kept apart. The canonical parametrization (20) combines the linear features in a single, common linear plane and records the non-linear features as double differences. The representation (20) is therefore well-suited for estimation and statistical inference. In terms of interpretation two issues remain: how to interpret double differences of the time effect directly, and whether any interpretation in terms of the original time effects in (1) is feasible.

Interpretation of double differences of time effects

The double differences have an odds ratio or difference-in-difference interpretation. A double difference in age is defined by

$$\Delta^2\alpha_{age} = \mu_{age,coh} - \mu_{age-1,coh} - \mu_{age-1,coh+1} + \mu_{age-2,coh+1}. \tag{26}$$

As a numerical example, let $age = 18$ and $coh = 2001$. Then the first two terms in (26) give the effect of ageing from 17 to 18 for the 2001 cohort, while the last two terms give the effect of ageing from 16 to 17 for the 2002 cohort. Both of these effects happen over the period 2017 to 2018, with the time convention in (6). Indeed, writing (26) in AP coordinates gives

$$\Delta^2\alpha_{age} = \mu_{age,per} - \mu_{age-1,per-1} - \mu_{age-1,per} + \mu_{age-2,per-1}. \tag{27}$$

On the right hand sides of (26) and (27) any pair of consecutive cohorts or periods, respectively, could be used. Thus $\Delta^2\alpha_{age}$ equals the average difference-in-difference effect for all cohorts or periods. For binary outcomes the double difference $\Delta^2\alpha_{age}$ has a log odds interpretation.

In the same vein, the period and cohort double differences are interpretable through

$$\Delta^2\beta_{per} = \mu_{age,coh} - \mu_{age-1,coh} - \mu_{age,coh-1} + \mu_{age-1,coh-1}, \tag{28}$$

$$\Delta^2\gamma_{coh} = \mu_{age,coh} - \mu_{age,coh-1} - \mu_{age+1,coh-1} + \mu_{age+1,coh-2}. \tag{29}$$

The equations (26), (28), (29) are illustrated with Figure 6, which is a modification of a figure in Martínez Miranda & al. (2015). A major advantage of the double differences is their invariance as explored above. However, estimated double differences will inevitably be somewhat erratic. Therefore it is often desirable for interpretation to generate a representation of the time effects by cumulating the double differences. This procedure is discussed below.

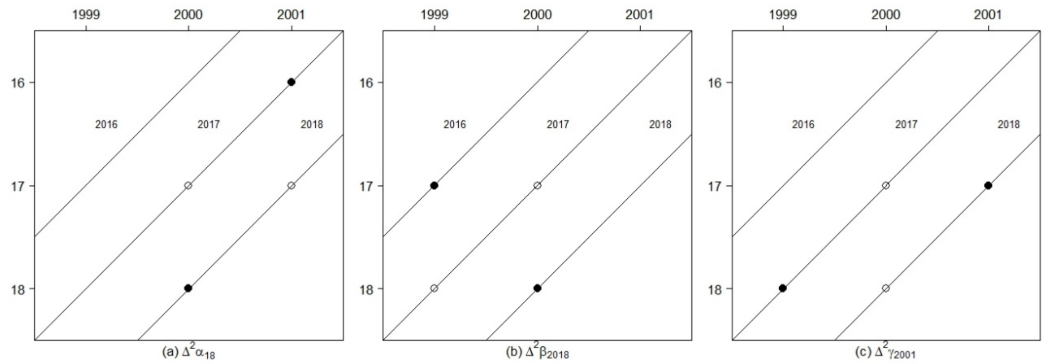

Figure 6. Illustration of double differences.
Solid/open circles represent predictors taken with positive/negative sign.

Interpretation of time effects

The original time effects are not fully identifiable and thus they are also not fully interpretable. Yet, the APC model (1) is composed of the time effects so it remains of interest to seek to interpret them as far as possible. Since the non-linear parts of the time effects are identifiable the focus should be on illustrating these. This can be done through detrending. The three plots of age, period, and cohort time effects should be interpretable individually, so the detrending must be applied to each time effect. Linear trends are absorbed into linear plane.

In the representations (17) and (20) the double differences are double cumulated with respect to the plane anchored at μ_{UU}, $\mu_{U,U+1}$, and $\mu_{U+1,U}$. This representation is useful for estimation as it immediately leads to design vectors in (19) and (22). However, the cumulations of the double differences are not ideally suited for graphical representation of the non-linear time effect. On the one hand, it is easy to see that these double sums have the same degrees of freedom as the double differences and are disentangled, in contrast to the time effects identified by restriction. On the other hand, they will often be strongly trending in practice which does not allow for an easy interpretation. The last issue can be addressed through detrending.

The double sums of double differences can be detrended in various ways. One approach would be to orthogonalize each of the three sets of double sums with respect to an intercept and a time trend. Consider for instance the age effects from (20). The double sum of double differences is $\alpha_{age}^{\Sigma\Sigma\Delta\Delta} = \sum_{t=3}^{age}\sum_{s=3}^{t}\Delta^2\alpha_s$, for $1 \leq age \leq \mathrm{A}$ so that $\alpha_1^{\Sigma\Sigma\Delta\Delta} = \alpha_2^{\Sigma\Sigma\Delta\Delta} = 0$. Then $\alpha_{age}^{ortho} = \alpha_{age}^{\Sigma\Sigma\Delta\Delta} - a - d \times age$ is the orthogonalized age effect if a, d are chosen so that $\Sigma_{age=1}^{A}\alpha_{age}^{ortho}(1, age) = 0$. This is in spirit with the approach of Deaton & Paxson (1994a), but with the difference that the orthogonalization is applied to each of the three double sums, so that the time trends are disentangled. A drawback of this approach is that it is no longer evident that the degrees of freedom are the same as for the double differences.

Another approach to detrending is to impose that the double sums start and end in zero (Nielsen, 2015). Defining $\alpha_{age}^{detrend} = \alpha_{age}^{\Sigma\Sigma\Delta\Delta} - a - d \times age$ this entails the choices $a = -d$ and $d = \alpha_A^{\Sigma\Sigma\Delta\Delta}/(A-1)$ so that $\alpha_1^{detrend} = \alpha_A^{detrend} = 0$. With this approach it is apparent that the degrees of freedom are the same as for the double differences. The graph of $\alpha_{age}^{detrend}$ visually emphasises the non-linearity as the start and end points are anchored at zero. At the same time, the detrending clearly depends on the particular index array with its particular choice of minimal and maximal age. From the graph it may be possible to identify a U or S-shaped curve which can be tested for consistency with a quadratic or higher-order polynomial.

Sub-models

A common empirical question is whether all components of the APC model are needed. While this question is often formulated in terms of the time effect formulation (1) it is actually easier to appreciate the restrictions and the associated degrees of freedom using the canonical parametrization (20) and the associated canonical parameter ξ in (21).

Age-cohort models

The hypothesis of no period effect illustrates the identification issues very well. The hypothesis results in age-cohort (AC) models, which are commonly used in economics; see for instance Browning & al. (1985), Attanasio (1998), Deaton & Paxson (2000), and Browning & al. (2016). AC models can arise through reduction of the general APC model or they may be postulated at the outset. From the perspective of the time effect formulation (1) the hypothesis is that $\beta_{L+1} = \cdots = \beta_{L+P} = 0$. This leaves the model (1) as an age-cohort model of the form

$$\mu_{age,coh} = \alpha_{age} + \gamma_{coh} + \delta. \qquad (30)$$

This formulation gives the impression of a P-dimensional restriction. However, it is in fact observationally equivalent to imposing a hypothesis of no non-linear effect in period. Under the canonical parametrization this is $\Delta^2\beta_{L+3} = \cdots = \Delta^2\beta_{L+P} = 0$, which is a restriction of dimension $P-2$. Nielsen & Nielsen (2014, §5.3) present a formal algebraic analysis of the relation between restrictions of time effects and double differences. The intuition is that because the period effect is only identified up to a linear trend, imposing the hypothesis $\beta_{L+1} = \cdots = \beta_{L+P} = 0$ in (1) does not actually restrict the common linear plane at all. Any linear effect of period will still be present in the restricted model (30).

The feature that the linear time effects are not identifiable from the AC model is perhaps best understood in the special case where all time effects are linear as in (11). It was explained in a previous section that (11) can be written equivalently as a combination of APC, AC, AP or CP effects. The model (30) is analogous to the model (12). At first glance it may appear natural to attribute the linear plane in (12) to age and cohort effects, but in fact the linear effect of period is not constrained. Rather it is absorbed into the slopes in the age and cohort dimensions, with $\mu_{21} - \mu_{11} = \Delta\alpha_2 + \Delta\beta_2$ and $\mu_{12} - \mu_{11} = \Delta\beta_2 + \Delta\gamma_2$.

Linear sub-models

Apart from the AC model, there are many other sub-models of the APC model. Table 1 gives a range of sub-models that may be of interest. It is taken from Nielsen (2015), with similar tables appearing in Holford (1983) and Oh & Holford (2015). The first model, denoted APC, is the unrestricted APC model.

Restricting one set of double differences. The three models, AP, AC, and PC each have one set of double differences or non-linearities eliminated, that is the cohort, period, and age double differences, respectively. The preceding remarks pertaining to the AC model apply to any of the three models.

Restricting two sets of double differences. The three models Ad, Pd, and Cd are known as drift models. For instance, the age-drift model has both period and cohort double differences eliminated, so that $\Delta^2\alpha_3 = \cdots = \Delta^2\alpha_A = 0$ and $\Delta^2\beta_{L+3} = \cdots = \Delta^2\beta_{L+P} = 0$, while the linear plane is unrestricted. The identification problem remains, as pointed out by Clayton & Schiffler (1987), because the linear plane can be parametrized either in terms of age and cohort linear trends or in terms of age and period linear trends.

Restricting two sets of double differences and the linear plane. The three models A, P, and C are the first to include restrictions on the linear plane. For instance, in the A model period and cohort double differences are eliminated and the linear plane is restricted to have just one slope, in age. Consequently, the A model can be written as $\mu_{age,coh} = \alpha_{age}$.

Linear plane model. Finally, the linear plane model arises when all non-linear effects are absent. In this case $\Delta^2\alpha_3 = \cdots = \Delta^2\alpha_A = 0$ and $\Delta^2\beta_{L+3} = \cdots = \Delta^2\beta_{L+P} = 0$ and $\Delta^2\gamma_3 = \cdots = \Delta^2\gamma_C = 0$. This model was used earlier to illustrate the simplest example of the APC identification problem of entangled linear age, period and cohort effects.

Model	Linear Plane	double differences $\Delta^2\alpha_{age}$	$\Delta^2\beta_{per}$	$\Delta^2\gamma_{coh}$	total
APC	*3*	*A-2*	*P-2*	*C-2*	*A+P+C-3*
AP	*3*	*A-2*	*P-2*		*A+P-1*
AC	*3*	*A-2*		*C-2*	*A+C-1*
PC	*3*		*P-2*	*C-2*	*P+C-1*
A-drift	*3*	*A-2*			*A+1*
P-drift	*3*		*P-2*		*P+1*
C-drift	*3*			*C-2*	*C+1*
A	*2*	*A-2*			*A*
P	*2*		*P-2*		*P*
C	*2*			*C-2*	*C*
linear plane	*3*				*3*

Table 1: *Sub-models with degrees of freedom.*

Functional form sub-models

Another set of sub-models arises by imposing a specific functional form on the time effects.

Quadratic polynomials. The age effect, in particular, often has a concave or convex appearance. In that case the age effect may be described parsimoniously by a quadratic polynomial. The hypotheses of a quadratic age effect, $\alpha_{age} = \alpha_c + \alpha_\ell \times age + \alpha_q \times age^2$ as in (2), and of constant double differences,

$$\Delta^2\alpha_3 = \cdots = \Delta^2\alpha_A, \tag{31}$$

are equivalent since the linear trends are not identified. Thus, the hypothesis can be imposed as a linear restriction on the canonical parameter. The degrees of freedom are $A-3$. Similarly, restricting a time effect to be a polynomial of order k is equivalent to restricting the corresponding double differences to be a polynomial of order $k-2$. For instance a slightly skew concave or an S-shape appearance could potentially be captured by a third order polynomial in the time effects, or equivalently a first order polynomial in the double differences.

A more elaborate quadratic model. Suppose now that all three time effects are quadratic so that equation (1) becomes

$$\mu_{age,coh} = \{\alpha_c + \alpha_\ell \times age + \alpha_q \times (age)^2\} + \{\beta_c + \beta_\ell \times per + \beta_q \times (per)^2\}$$
$$+\{\gamma_c + \gamma_\ell \times coh + \gamma_q \times (coh)^2\} + \delta. \qquad (32)$$

The identifiable non-linear parameters are $\alpha_q, \beta_q, \gamma_q$, while the remaining parameters combine to a linear plane as in (11). A sub-model is the quadratic AC model

$$\mu_{age,coh} = \{\alpha_c + \alpha_\ell \times age + \alpha_q \times (age)^2\}$$
$$+\{\gamma_c + \gamma_\ell \times coh + \gamma_q \times (coh)^2\} + \delta, \qquad (33)$$

which is a special case of (30). The linear parts $\alpha_c + \alpha_\ell \times age, \gamma_c + \gamma_\ell \times coh,$ and δ combine to a linear plane and the identification problem remains. Only the absence of β_q is an over-identifying constraint. Thus, a tests of (33) against (32) would have 1 degree of freedom.

Replacing a time effect by an observed variable. It is often of interest to replace the period effect, in particular, with an observed time series, T_{per} say. The time series T_{per} decomposes into a linear part and a non-linear part. Thus, in the context of an APC model it is equivalent to imposing $\beta_{per} = T_{per}$ for $1 \leq per \leq P$ and $\Delta^2\beta_{per} = \Delta^2 T_{per}$ for $3 \leq per \leq P$. Thus, this restriction has $P-3$ degrees of freedom. Since there is already a linear plane in the model the linear effect of T_{per} remains unidentified.

When to use APC models

It is important to recognise that the APC models described above do not "solve" the identification problem. The identification problem still limits the range of questions that can be answered using formal statistical analysis. The following sections explain the questions that can and cannot be answered with APC models, given that the non-linear parts of the time effects are identified but the linear parts are not.

Questions that can be answered

The questions that APC models can answer fall into the following categories: certain difference-in-difference questions; questions related to the non-linear effects of age, period, or cohort; exploratory analysis; forecasting; and questions where APC effects appear in the model as control variables.

Difference-in-difference analysis can be done using the APC model. For example, McKenzie (2006) uses data from the Mexican ENIGH household survey, collected at two-year intervals, to investigate the effect of the 1995 peso crisis on consumption. He compares the change in consumption from 1994 to 1996 with the change in consumption from 1992 to 1994, and that from 1996 to 1998. This is equivalent to tests on the parameters $\Delta^2\beta_{1996}$ and $\Delta^2\beta_{1998}$.

Non-linearities implied by economic theory can be investigated with APC models. For example, the life cycle hypothesis of consumption implies decelerating saving in old age, which is a testable non-linearity in the age effect. An analysis could start by first estimating an APC model for the stock of savings. Next, the age non-linearity would be isolated from the linear plane and tested for significance. If significant, the shape could be inspected for consistency with the life cycle hypothesis in consumption either through visual inspection or through a formal test for instance for a concave, quadratic age effect, see (31).

Exploratory analysis. APC models are well-suited to exploratory analysis. Diouf & al. (2010) conduct such an analysis of the dynamics of the obesity epidemic in France from 1997 to 2006. They find significant curvature in the cohort dimension, with deceleration among those who were children during the second world war and acceleration post-1960s, but little evidence for non-linearities in either age or period. These findings correspond to an cohort-drift model, see Table 1, and are interpreted as evidence that early life conditions are important determinants of obesity.

Forecasting. APC models are effective forecasting tools. When forecasting it is usually necessary to extrapolate one or more time effects. Identification assumptions will impact the forecast unless the extrapolation method is chosen to be invariant to transformations in (10), see Kuang & al. (2008b). Extrapolation can avoided altogether if an AC model is adequate and only current cohorts are forecasted. Mammen & al. (2015) refer to this as in-sample forecasting. One example is the Chain-Ladder model used in general insurance (England & Verrall, 2002) with distribution forecasts by bootstrap (England, 2002) or by asymptotic theory (Harnau & Nielsen, 2017). Another example is the forecast of future rates of mesothelioma, a cancer resulting from exposure to asbestos, in Martinez Miranda & al. (2015). Extrapolation is needed when the model involves period non-linearities. In that case techniques from econometric forecasting of non-stationary time series can be applied with advantage, see Clements & Hendry (1999). An application to general insurance is given by Kuang & al. (2011).

Questions that do not involve time effects. Often, a researcher is interested in the effect of some policy intervention or treatment but is concerned about possible confounding with pure time effects; in this case, the APC model is included as a statistical control. For example, Ejrnæs & Hochguertel (2013) are interested in the effect of a change to unemployment insurance in Denmark on employment and use a model incorporating APC effects identified by restriction to ensure that their results are not contaminated by pure time effects.

There are many variations and extensions of the question types outlined above. One possibility is to include interactions with other covariates; for example, allowing for an interaction between age and level of education in a model for earnings. Another is to use two or more samples and test cross-sample restrictions: comparing estimated period non-linearities in savings between pairs of countries to assess macroeconomic interdependence. Some extensions are discussed further in the section on using APC models.

Questions that cannot be answered

Any question relating to the linear parts of any of the time effects is unanswerable. This is true regardless of the nature of the dataset. If the data is a single slice in any one time dimension it is not possible to separate the effects of the other two. For example, with a cross-section of adults in 2018 it is not possible to determine whether the old have higher savings because savings increase with age or because later cohorts exhibit declining financial responsibility.

Having a repeated cross-section, containing data from 2008-2018, does not help. There is now a possible period trend to contend with: savings may be decreasing over the period range due to a rising gap between real wages and the cost of living. An APC model cannot separate these effects, except by imposing a substantive and untestable assumption. More subtly, it is not possible to identify the linear part of the effect in a single time dimension even if the other time dimensions are excluded from the model.

Given the above, it is recommended that hypotheses in terms of the linear parts of any of the three time effects be avoided, as any test of these is necessarily biased. Instead, it is advised to formulate hypotheses primarily in terms of the non-linear parts of time effects.

Using APC models

This section introduces the reader to the practicalities of APC modelling. The different data contexts in which APC models have been used are described. Possible extensions of the APC models are discussed. Finally, a fully-worked example of an APC analysis is provided.

Data types

APC models have primarily been used with aggregate or repeated cross-section data. The most commonly used models are least squares, Poisson, and logistic regressions. These are all examples of generalized linear models (GLMs); the GLM framework was developed by Nelder & Wedderburn (1972) and an introduction can be found in Dobson (1990).

Aggregate data

The simplest form of APC data is a table where each age-cohort combination is a single cell. Information is aggregated over individuals within each cell. The APC literature using this form of data has focused on point estimation and point forecasting. The information recorded in each cell will take one of the following forms:

Counts of both exposure and outcomes. An example is the size of the labour force and the number of unemployed. This format is common in epidemiology, where exposure is the population size and the outcome is the number of deaths from a particular disease, such as cancer. Clayton & Schifflers (1987) provide an overview of the use of APC models for this form of epidemiological data. Such data are analysed using logistic regression or by Poisson regression with the log exposure as an offset.

Rates can be calculated from counts of outcomes and exposure. The unemployment rate is a clear example. In demography, fertility and mortality rates are of substantial interest. Rates are often modelled by (log) least squares regression.

Counts of outcomes without a measure of exposure. While outcomes may be clearly defined, the exposure is sometimes ill-defined or poorly measured. Forecasts of the counts alone may be of interest in this situation. An example from epidemiology is the number of AIDS cases classified by time of diagnosis (cohort) and reporting delay (age), where only an unknown subset of the population is exposed (Davison & Hinkley, 1997, Example 7.4). Another example is the number of mesothelioma deaths, caused by exposure to asbestos fibres, classified by age and year of death (period). Proxies for exposure may be constructed (Peto & al., 1995), or the counts can be modelled directly using Poisson regression with no offset (Martínez Miranda & al., 2015).

Values of outcomes without a measure of exposure. An example is the insurance reserving problem, where the data consists of the total value of payments from an insurance portfolio classified by insurance year (cohort) and reporting delay (age). The objective is to forecast unknown liabilities (i.e. incurred but not yet reported). A commonly-used modelling approach is the chain ladder (England & Verrall, 2002), which is equivalent to a Poisson regression with an AC predictor.

Inference for aggregate data

For conducting inference, classical exact normal theory may be applied. Some thought is required concerning the repetitive structure. Two frameworks have been considered for asymptotic analysis: expanding array asymptotics and fixed array asymptotics.

Expanding array asymptotics. Fu & Hall (2006) consider a least squares approach to modelling aggregate values of outcomes. The time effects are identified by restricting averages in each dimension to zero. Consistency is investigated with increasing period dimension. Fu (2016) gives further consistency results for the age effects for the same least squares model and for a Poisson regression with exposure.

Fixed array asymptotics. By holding the time dimensions fixed, asymptotic analysis can be related to the analysis of contingency tables (Agresti, 2013) with the difference that rows and columns are ordered by the APC structure. The analysis of models without exposure has been studied. Martínez Miranda et al. (2015) considered a Poisson model for counts, while Harnau & Nielsen (2017) analysed an over-dispersed Poisson model for values of outcomes.

Specification tests. For aggregate, discrete data the model fit can be assessed by a deviance test against a saturated model where the cells have unrelated predictors $\mu_{age,coh}$. Harnau (2018a) suggests a Bartlett test for constant over dispersion in an over-dispersed Poisson model. Harnau (2018b) suggests an encompassing test comparing over-dispersed Poisson and log normal specifications.

Repeated cross sections

Repeated surveys can be used to form repeated cross section data. A basic regression model would be of the form (3). Ejrnæs & Hochguertel (2013) estimate a model of this form and address the identification problem by the restriction method. Yang & Land (2006) propose a hierarchical APC model where age is quadratic and where cohort and period are treated as random effects. Fannon & al (2018) propose models involving the canonical parametrization. This includes a least squares regression as in (3) and a logistic regression of the form

$$\log\{P(Y_i = 1|Z)/P(Y_i = 0|Z)\} = \zeta' Z_i + \mu_{age_i,coh_i} \tag{34}$$

Asymptotic inference is conducted by allowing the number of individuals in the sample to increase while holding the array fixed. Likelihood ratio tests are used to assess restrictions imposed on the APC model. In both models the fit can be tested by saturating the data array with indicators for each age-cohort cell.

Extensions

Continuous time data

There is a budding literature on non-parametric models for continuous time data. Ogate et al. (2000) develop an empirical Bayes model for the incidence of diabetes. Martínez Miranda et al. (2013) develop a continuous time version of the chain ladder model. This is extended to in-sample density forecasting methods by Lee et al. (2015) and Mammen et al. (2015).

Models with unequal intervals

The theoretical framework used in this chapter is primarily concerned with data where each time dimension is recorded in the same units. This is often not the case.

Regular intervals. It is common that data are recorded annually, but age is grouped at a coarser level; this is seen in the empirical example in this chapter. There are two approaches when working with such data. The first and easy option is to coerce the data into a single unit framework by grouping periods, either by taking averages or by dropping certain periods. This of course implies a loss of information. The second option is to construct a model allowing for different interval lengths. This may actually create more identification issues, as discussed by Holford, 2006. He proposed an approach based on finding the least common multiple of the interval lengths, using this least common multiple to split the data into blocks, and treating within-block micro trends separately from between-block macro trends.

Irregular intervals. This can arise with repeated survey data. In some cases one is interested in an outcome variable that is irregularly recorded; for instance, a variable recorded in 1997, 1999, 2002, 2009, and annually thereafter. One solution is to use a subsample with a single frequency. An alternative possibility may be to use interpolation to regularise the intervals or to use continuous time scales.

Two-sample-model

A further extension involves combining data for two samples, for instance women and men or data from two countries. The model (1) for the predictor then becomes

$$\mu_{age,coh,s} = \alpha_{age,s} + \beta_{per,s} + \gamma_{coh,s} + \delta_s, \tag{35}$$

where the index s indicates the sample. Tests could then be performed for common parameters between the two samples, for instance a common period effect such that $\beta_{per,1} = \beta_{per,2}$. Riebler & Held (2010) present a Bayesian estimation method. The identification is discussed further by Nielsen & Nielsen (2014).

Software

Various software packages are available for APC analysis. For R these include *epi* (Carstensen, 2013) and *apc* (Nielsen, 2018). For Stata these include *st0245* (Sasieni, 2012), *apc* (Schulhofer-Wohl & Yang, 2006), and *apcd* (Chauvel, 2012).

Empirical illustration using US employment data

Consider US data for employment for 1960-2015, retrieved from the OECD's online database. Age is recorded in five-year intervals. Data from every fifth year is used to get an AP dataset with base unit five. There are 12 periods and 11 ages, thus 22 cohorts. Table 2 shows the size of the labour force in each age-cohort cell while Table 3 shows the number of unemployed, each in thousands of persons.

Various questions could be answered with this data. Expected non-linearities could be checked: for example, a U-shape in age, or discontinuities in period consistent with known periods of recession. Difference-in-difference hypotheses could be tested: for instance, was there a significant difference between the increase in unemployment from 2000 to 2005 and that from 2005 to 2010? This could indicate how quickly the effects of the financial crisis were felt in the labour market.

	1960	**1965**	**1970**	**1975**	**1980**	**1985**	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**
15-19	5246	6350	7249	8870	9380	7901	7792	7765	8271	7164	5905	5700
20-24	7679	9301	10597	13750	15922	15717	14700	13687	14251	15127	15028	15523
25-29	7186	7582	9241	12698	15400	17265	17677	15913	15800	16049	17300	17494
30-34	7884	7407	7795	10165	13827	16285	18253	18285	16955	16291	16313	17153
35-39	8474	8341	7774	8560	11161	14371	16927	18633	18616	17124	16271	16267
40-44	8173	8887	8664	8343	9303	11702	15218	17118	18950	18905	17095	16337
45-49	8011	8326	8980	8675	8478	9270	11557	14667	16907	18562	18460	16640
50-54	6903	7520	7968	8409	8433	8052	8691	10555	14164	15841	17500	17262
55-59	5464	6138	6768	6866	7388	7240	6902	7423	9267	12289	14145	15394
60-64	3927	4217	4515	4480	4597	4751	4673	4437	5090	6691	9152	10559
65-69	1798	1794	1922	1757	1828	1719	2076	2123	2322	2846	3796	5125

Table 2: *US labour force in 1000s*

	1960	**1965**	**1970**	**1975**	**1980**	**1985**	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**
15-19	711	874	1105	1768	1668	1467	1211	1346	1082	1186	1527	966
20-24	583	557	866	1864	1836	1738	1299	1244	1022	1335	2329	1501
25-29	380	288	427	1091	1234	1299	1056	916	651	933	1883	1057
30-34	372	241	290	685	791	1043	938	925	556	728	1501	848
35-39	354	272	250	514	548	769	739	864	582	694	1320	708
40-44	317	275	265	437	392	572	589	686	550	705	1383	644
45-49	328	237	261	452	362	448	443	503	422	675	1441	616
50-54	286	199	214	440	313	364	279	342	340	520	1328	643
55-59	221	189	197	308	246	327	241	266	220	416	995	576
60-64	174	133	113	212	153	191	145	159	134	214	667	402
65-69	83	68	75	114	66	62	67	91	73	98	286	198

Table 3: *US unemployed in 1000s*

Preliminaries

The package *apc* for R is used (Nielsen, 2015). The first step of the analysis is to visualize the data. Employment rates are found by dividing the unemployment numbers in Table 3 with the labour force numbers in Table 2. Line plots of within-period changes in employment with respect to age, or within-cohort evolution of unemployment over time, can be informative; see Figure 7. To aid the visualization the numbers are averaged over 10 or 20 year groups. The curves in panel (a) correspond to the columns in the AP table for unemployment rates. Panel (b) shows the same columns, but plotted against cohort which is period minus age. In panel (c) the curves correspond to the cohort diagonals in the AP table plotted against age. Finally, in panel (d) the rows of the AP table are plotted against cohort.

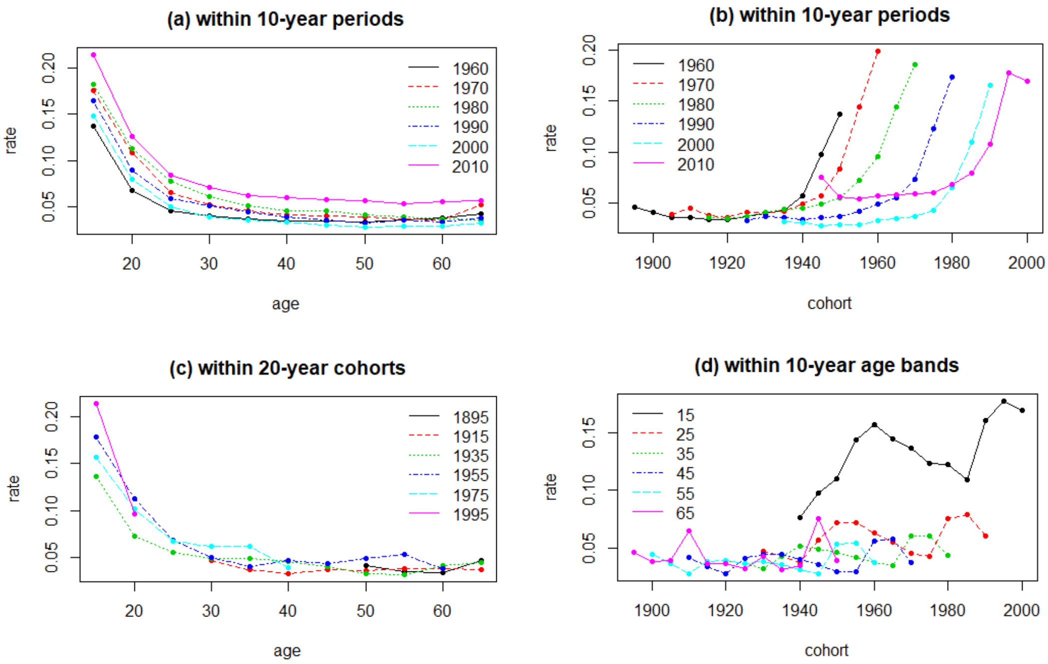

Figure 7. Plots of unemployment data.

Model estimation

To answer the questions proposed above an econometric model that isolates the identifiable non-linear parts of the time effects from the non-identifiable linear parts is required. A logit model is used where

$$\log\{\pi_{age,coh}/(1-\pi_{age,coh})\} = \mu_{age,coh}. \tag{36}$$

Here $\pi_{age,coh}$ is the probability of unemployment for a given age-cohort combination and $\mu_{age,coh} = \xi' x_{age,coh}$, where ξ and $x_{age,coh}$ are given in (21) and (22). Since the canonical parametrization is identified and embedded in a GLM framework it can be estimated uniquely.

The individual double-differences at this point have a diff-in-diff or log odds interpretation. Where it is of interest to study the general shape of the non-linearities in each time dimension, the double differences may be double cumulated and detrended, following the discussion in the earlier section on *interpretation of time effects*. This fully separates the linear and non-linear parts of the time effects.

Figure 8 visualizes the estimated APC model for the US unemployment data using the canonical parametrization and detrending. Panels (a)-(c) show the estimated double-differences in each of age, period, and cohort. Panels (d)-(f) show the level and slopes of the linear plane, calculated after the detrending. Panels (g)-(i) show the non-linear parts of time effects. These are found by double cumulating and detrending the double differences so that the first and last value in each plot is anchored at zero. There is evidence for a U-shaped relationship between age and unemployment. The non-linear parts of the period effect show the discontinuous effects of macroeconomic conditions, with accelerations in unemployment in the early 1970s and late 2000s. There is weak evidence for discontinuities in cohort which may reflect hysteresis; the cohorts of the late 1950s (who came of age in the 1970s) are relatively underemployed compared to those before and after them.

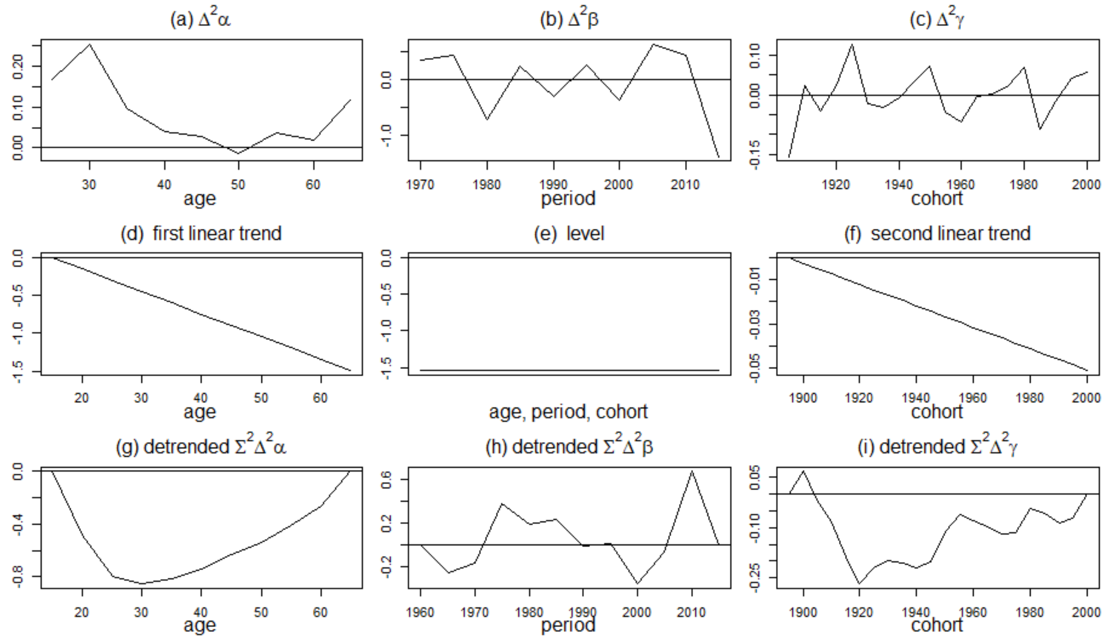

Figure 8. APC model for US unemployment data in terms of the canonical parametrization.

Conclusions

The existence of an identification problem between age, period, and cohort is widely recognised by economists. Many papers have grappled with the problem, particularly in the contexts of consumption, savings, and labour market dynamics. The problem is not unique to economics; it is also discussed by sociologists, demographers, political scientists, actuaries, epidemiologists, and statisticians. A comprehensive account of the problem therefore requires a survey of a broad literature, much of it outside economics.

The APC identification problem arises due to the identity $age + coh = per + 1$ which links the time scales. This chapter has focused exclusively on the linear APC model, but the problem also arises in the non-linear Lee-Carter (Lee & Carter 1992) model and in extensions thereof such as Cairns et al. (2009). The main features of the APC identification problem are the following. First, it is a problem affecting the linear parts of the time effects only; the levels and slopes specific to each dimension cannot be identified, whereas higher-order effects can be. Second, a model including only one or two of the three remains afflicted by the problem. Finally, the problem is fundamentally one in continuous time; changing the observation unit for the APC scales will not resolve it.

A range of identification strategies have been proposed to deal with the APC problem, some of which are outlined in this chapter. The key question to ask of any such strategy is: Would a different identification strategy lead to the same conclusions? This is a question of invariance to the transformations in (10). Of those parametrizations discussed in this chapter, only the canonical parametrization is invariant as it does not attempt the impossible by seeking to separate the linear effects, but rather focuses on the identifiable non-linear effects. This brings clarity to interpretation and inference.

Acknowledgements

Funding was received from ESRC grant ES/J500112/1 (Fannon) and ERC grant 694262, DisCont (Nielsen).

References

Attanasio, O.P. (1998) Cohort analysis of saving behaviour by U.S. households. *Journal of Human Resources 33*, 575-609.

Barndorff-Nielsen, O.E. (1978) *Information and Exponential Families.* New York: Wiley.

Beenstock, M., Chiswick, B.R., & Paltiel, A. (2010) Testing the immigrant assimilation hypothesis with longitudinal data. *Review of Economics of the Household 8,* 7-27.

Berzuini, C. & Clayton, D. (1994) Bayesian analysis of survival on multiple time scales. *Statistics in Medicine 13*, 823-838.

Browning, M., Crossley, T.F. & Lührmann, M. (2016) Durable purchases over the later life cycle. *Oxford Bulletin of Economics and Statistics 78*, 145-169.

Browning, M., Deaton, A. & Irish, M. (1985) A profitable approach to labor supply and commodity demands over the life-cycle. *Econometrica 53*, 503-544.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A. & Balevich, I. (2009) A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal 13*, 1-35.

Carstensen, B. (2007) Age-period-cohort models for the Lexis diagram. *Statistics in Medicine 26*, 3018-3045.

Carstensen, B., Plummer, M., Laara, E., & Hills, M. (2018) Epi: A Package for Statistical Analysis in Epidemiology. R package version 2.32. URL https://CRAN.R-project.org/package=Epi

Chauvel, L. (2012) APCD: Stata module for estimating age-period-cohort effects with detrended coefficients, *Statistical Software Components S457440*, Boston College Department of Economics.

Chow, G.C. (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*, 591-605.

Clayton, D. and Schifflers, E. (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine 6*, 469-481.

Clements, M.P. & Hendry, D.F. (1999). *Forecasting Non-Stationary Time Series.* Cambridge, MA: MIT Press.

Cox, D.R. & Hinkley, D.V. (1974) *Theoretical Statistics.* London: Chapman & Hall.

Davison, A.C. & Hinkley, D.V. (1997) *Bootstrap Methods and their Applications.* Cambridge University Press.

Deaton, A.S. & Paxson, C.H. (1994a) Saving, growth, and aging in Taiwan. In Wise, D.A. (ed.) *Studies in the Economics of Aging.* Chicago University Press, 331-361.

Deaton, A.S. & Paxson, C.H. (1994b) Intertemporal choice and inequality. *Journal of Political Economy 102*, 437-467.

Deaton, A & Paxson, C. (2000) Growth and saving among individuals and households. *Review of Economics and Statistics 82*, 212-225.

Diouf, I., Charles, M., Ducimetière, P., Basdevant, A., Eschwege, E. & Heude, B. (2010) Evolution of obesity prevalence in France: An age-period-cohort analysis. *Epidemiology 21,* 360-365.

Dobson, A. (1990) *An introduction to generalized linear models.* Boca Raton, Fl: London: Chapman and Hall.

Ejrnæs, M. & Hochguertel, S. (2013) Is business failure due to lack of effort? Empirical evidence from a large administrative sample. *Economic Journal 123*, 791-830.

England, P.D. (2002) Addendum to "Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics 31*, 461-466.

England, P.D. and Verrall, R.J. (2002) Stochastic claims reserving in general insurance. *British Actuarial Journal 8*, 519-544.

Fahrmeir, L. & Kaufmann, H. (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics 13*, 342-368.

Fannon, Z., Monden, C. & Nielsen, B. (2018) Age-period-cohort modelling and covariates, with an application to obesity in England 2001-2014. Mimeo, University of Oxford.

Fienberg, S.E. and Mason, W.M. (1979) Identification and estimation of age-period-cohort models in the analysis of discrete archival data. *Sociological Methodology 10*, 1-67.

Fitzenberger, B., Schnabel, R. & Wunderlich, G. (2004) The gender gap in labor market participation and employment: a cohort analysis for West Germany. *Journal of Population Economics 17,* 83-116.

Fu, W.J. (2016) Constrained estimators and consistency of a regression model on a Lexis diagram. *Journal of the American Statistical Association 111*, 180-199.

Fu, W.J. & Hall, P. (2006) Asymptotic properties of estimators in age-period-cohort analysis. *Statistics & Probability Letters 76*, 1925-1929.

Fu, W.J., Land, K.C. & Yang, Y. (2011) On the intrinsic estimators and constrained estimators in age-period-cohort models. *Sociological Methods & Research 40*, 453-466.

Hanoch, G. & Honig, M. (1985) "True" age profiles of earnings: adjusting for censoring and for period and cohort effects. *The Review of Economics and Statistics 67*, 384-394.

Harnau, J. (2018a) Misspecification tests for log-normal and over-dispersed Poisson chain-ladder models. *Risk 6*, article 25.

Harnau, J. (2018b) Log-normal or over-dispersed Poisson. *Risk 6*, article 70.

Harnau, J. & Nielsen, B. (2017) Over-dispersed age-period-cohort models. To appear in *Journal of the American Statistical Association*.

Heckman, J. & Robb, R. (1985) Using longitudinal data to estimate age, period and cohort effects in earnings equations. In Mason, W.M. & Fienberg, S.E. (eds.) *Cohort Analysis in Social Research*, New York: Springer, 137-150.

Holford, T.R. (1983) The estimation of age, period and cohort effects for vital rates. *Biometrics 39*, 311-324.

Holford, T.R. (1985) An alternative approach to statistical age-period-cohort analysis. *Journal of Chronic Diseases 38*, 831-836.

Holford, T.R. (2005) Age-period-cohort analysis. In *Encyclopedia of Biostatistics.* J. Wiley & Sons. Reissued in *Wiley StatsRef: Statistics Reference Online*, 2014.

Holford, T.R. (2006) Approaches to fitting age-period-cohort models with unequal intervals. *Statistics in Medicine 25*, 977-993.

Kalwij, A.S. & Alessie, R. (2007) Permanent and transitory wages of British men, 1975-2001: year, age, and cohort effects. *Journal of Applied Econometrics 22,* 1063-1093.

Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philosophical Transactions of the Royal Society of London A332*, 487-509.

Krueger, A.B. & Pischke, J. (1992) The effect of social security on labor supply: a cohort analysis of the notch generation. *Journal of Labor Economics 10,* 412-437.

Kuang, D., Nielsen, B. & Nielsen, J.P. (2008a) Identification of the age-period-cohort model and the extended chain ladder model. *Biometrika 95*, 979-986.

Kuang, D., Nielsen, B. & Nielsen, J.P. (2008b) Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika 95*, 987-991.

Kupper, L.L., Janis, J.M., Karmous, A. & Greenberg, B.G. (1985) Statistical age-period-cohort analysis: A review and critique. *Journal of Chronic Diseases 38*, 811-830.

Lee, R.D. & Carter, L.R. (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association 87*, 659-671.

Lee, Y.K., Mammen, E., Nielsen, J.P. & Park, B.U. (2015) Asymptotics for in-sample density forecasting. *Annals of Statistics 43*, 620-651.

Lehman, E.L. (1986) *Testing Statistical Hypotheses*, 2nd ed. New York: Springer.

Luo, L. (2013) Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem. *Demography 50*, 1945-1967.

Mammen, E., Martínez Miranda, M.D & Nielsen, J.P. (2015) In-sample forecasting applied to reserving and mesothelioma mortality. *Insurance: Mathematics and Economics 61*, 76-86.

Martínez Miranda, M.D., Nielsen, B. & Nielsen, J.P. (2015) Inference and forecasting in the age-period-cohort model with unknown exposure with an application to mesothelioma mortality. *Journal of the Royal Statistical Society A178*, 29-55.

Martínez Miranda, M.D., Nielsen, J.P., Sperlich, S. & Verrall, R. (2013) Continuous Chain Ladder: Reformulating and generalizing a classical insurance problem. *Expert Systems with Applications 40*, 5588-5603.

Mason, K.O., Mason, W.M., Winsborough, H.H. & 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.

Meghir, C. & Whitehouse, E. (1996) The evolution of wages in the United Kingdom: Evidence from micro-data. *Journal of Labor Economics 14*, 1-25.

Moffitt, R. (1993) Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics 59*, 99-123.

Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of the Royal Statistical Society Series A 135,* 370-384.

Nielsen, B. & Nielsen, J.P. (2014) Identification and forecasting in mortality models. *The Scientific World Journal*. Article ID 347043, 24 pages. doi:10.1155/2014/347043

Nielsen, B. (2015) apc: An R package for age-period-cohort analysis. *R Journal 7*, 52-64.

Nielsen, B (2018). apc: Age-Period-Cohort Analysis. R package version 1.4. URL https://CRAN.R-project.org/package=apc

O'Brien, R.M. (2011) Constrained estimators and age-period-cohort models (with discussion). *Sociological Methods & Research 40*, 419-470.

O'Brien, R.M. (2015) *Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data.* Boca Raton, FL: CRC Press.

OECD (2018) Labour Force Statistics. https://stats.oecd.org/index.aspx?queryid=36499# (accessed 09 Nov 2018 12:13 GMT)

Ogate, Y., Katsura, K., Keiding, N., Holst, C. & Green, A. (2000) Empirical Bayes age-period-cohort analysis of retrospective incidence data. *Scandinavian Journal of Statistics 27*, 415-432.

Oh, C. & Holford, T.R. (2015) Age-period-cohort approaches to back-calculation of cancer incidence rate. *Statistics in Medicine 34*, 1953-1964.

Osmond, C. & Gardner, M.J. (1982) Age, period and cohort models applied to cancer mortality rates. *Statistics in Medicine 1*, 245-259.

Osmond, C. & Gardner, M.J. (1989) Age, period, and cohort models: Non-overlapping cohorts don't resolve the identification problem. *American Journal of Epidemiology 129*, 31-35.

Peto, J., Hodgson, J.T., Matthews, F.E. & Jones, J.R. (1995) Continuing increase in mesothelioma mortality in Britain. *Lancet 345*, 535-539.

Poirier, D. (1998) Revising belief in nonidentified models. *Econometric Theory 14*, 483-509.

Riebler, A. and Held, L. (2010) The analysis of heterogeneous time trends in multivariate age-period-cohort models. *Biostatistics 11*, 57-69.

Schulhofer-Wohl, S. & Yang, Y. (2006) APC: Stata module for estimating age-period-cohort effects, *Statistical Software Components S456754*, Boston College Department of Economics.

Schulhofer-Wohl, S. (2018) The age-time-cohort problem and the identification of structural parameters in life-cycle models. *Quantitative Economics 9*, 643-658.

Smith, T.R. & Wakefield, J. (2016) A review and comparison of age-period-cohort models for cancer incidence. *Statistical Science 31*, 591-610.

Stasieni, P.D. (2012) Age-period-cohort models in Stata. *The Stata Journal 12*, 45-60.

Voas, D. & Chaves, M. (2016) Is the United States a counterexample to the secularization thesis? *American Journal of Sociology 121,* 1517-56.

Yang, Y. & Land, K.C. (2006) Age-period-cohort analysis of repeated cross-section surveys. *Sociological Methodology 36*, 297-326.

Yang, Y. & Land, K.D. (2013) *Age-Period-Cohort Analysis: New Models, Methods and Empirical Applications.* Boca Raton, FL: CRC Press.

Yang, Y., Fu, W.J. & Land, K.C. (2004) A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models. *Sociological Methodology 34*, 75-110.