OPEN ACCESS

University of BRISTOL

Link to published version (if available):
10.4054/DemRes.2014.30.11

Link to publication record in Explore Bristol Research
PDF-document

Another 'futile quest'? A simulation study of Yang and Land's Hierarchical Age-Period-Cohort model

Andrew Bell and Kelvyn Jones

School of Geographical Sciences and Centre for Multilevel Modelling, University of Bristol

Last Revised: 22nd February 2013

Draft – please do not cite without permission

Contact: andrew.bell@bristol.ac.uk

Abstract

BACKGROUND

Whilst some argue that a solution to the age-period-cohort (APC) 'identification problem' is impossible, numerous methodological solutions have been proposed, including Yang and Land's Hierarchical-APC (HAPC) model: a multilevel model considering periods and cohorts as cross-classified contexts in which individuals exist.

OBJECTIVE

To assess the assumptions made by the HAPC model, and the situations in which it does and does not work.

METHODS

Simulation study. Simulation scenarios assess the effect of (a) cohort *trends* in the Data Generating Process (DGP) (compared to only random variation), and (b) grouping cohorts (in both DGP and fitted model).

RESULTS

The model only works if either (a) we can assume that there are no linear (or non-linear) *trends* in period or cohorts, (b) we control any cohort trend in the model's fixed part and assume there is no period trend, or (c) we group cohorts in such a way that they exactly match the groupings in the (unknown) DGP. Otherwise, the model can arbitrarily reapportion APC effects, radically impacting interpretation.

CONCLUSIONS

Since the purpose of APC analysis is often to ascertain the presence of period and/or cohort trends, and since we rarely have solid (if any) theory regarding cohort groupings, there are few circumstances in which this model achieves what Yang and Land claim it can. The results bring into question findings of several published studies using the HAPC model. However, the structure of the model remains a conceptual advance that is useful when we can assume the DGP has no period trends.

Keywords

1 Introduction

Social scientists have long been interested in how social processes change over time. Since the 1960s (Ryder, 1965), this interest has developed into an understanding that change over time can occur in three conceptually distinct ways. First, individuals can *age*, such that change occurs over their individual life courses. Second, change can occur by *cohorts*, such that a new birth cohort will be intrinsically different to a previous one regardless of their age. And third, change can occur by *periods*, such that characteristics of a given occasion affect individuals living through that occasion, again regardless of when they were born and how old they are in that year. Which of these age, period and cohort (APC) effects are important for a given research question is of profound importance to researchers in many disciplines who are interested in how things change over time.

A problem arises when attempting to model all three of these effects simultaneously, because the three terms are exactly mathematically dependent. This 'identification problem' has been known about for decades, and potential solutions to the problem have been proposed sparking debate, particularly in sociology (Firebaugh, 1989, Mason et al., 1973, Sasaki and Suzuki, 1987, Yang et al., 2004) and medical sciences (Osmond and Gardner, 1989, Robertson and Boyle, 1986, 1998a, 1998b, Tu et al., 2011). Over the past 30 years there have been a plethora of solutions proposed, but none have been shown to be foolproof. In fact, many in social science (Glenn, 1976, 2005, Goldstein, 1979) see the separation of the three terms as not just difficult, but a "futile quest" (Glenn, 1976):

"The continued search for a statistical technique that can be mechanically applied always to correctly estimate the effects is one of the most bizarre instances in the history of science of repeated attempts to do the logically impossible."

Glenn (2005, p. 6)

This paper considers one of these solutions, proposed recently by Yang and Land (2006) with some additional methodological caveats discussed in subsequent papers (Yang, 2006, Yang et al., 2011, Yang and Land, 2008). The methodology has already been employed in a number of empirical applications, studying social trends in happiness (Yang, 2008a), voter turnout (Dassonneville, 2012), obesity (Reither et al., 2009), religious service attendance (Schwadel, 2010) and cannabis use (Piontek et al., 2012), to name a few. With the relatively speedy uptake of this method, understanding how and when it works (if it works at all) is of profound importance.

We thus use a simulation study to assess the model: the assumptions that it makes, the bias that occurs when those assumptions are broken, the situations in which it could be of use and the situations in which it is not. In doing so, we bring into question both the value of the model as a 'solution' to the identification problem, and the conclusions of a number of empirical studies that have used it. This is not to say that the conceptual distinction between age, period and cohort is useless, nor that the model could not be of use in a number of situations in finding substantive conclusions. However we hope our study will warn against the mechanical application of this, or any, age-period-cohort model, without any critical forethought about what the model may be hiding.

The paper proceeds as follows. The next section reviews the conceptual arguments that have accompanied the development of various attempts to solve the APC identification

problem. The paper then assesses the model suggested by Yang and Land, considering its conceptual structure, its technical specification and some problems that we see with Yang and Land's justification of it. Section 4 outlines the simulation study that we use to assess the model, and section 5 presents the results from this, before the implications of these results are discussed in section 6.

2   The Age-Period-Cohort identification problem

When studying longitudinal social processes, researchers are in essence looking at change. However the form of that change can be multifaceted. The distinction between ageing, change between periods and change over cohort groups has been considered for decades. Ryder (1965) was one of the first to make the distinction, by considering cohorts as a source of social change, rather than thinking of change as occurring over time through successive years. The division is well summarized by this fictional dialogue by Suzuki (2012 p. 1):

> A: I can't seem to shake off this tired feeling. Guess I'm just getting old. [Age effect]
> B: Do you think it's stress? Business is down this year, and you've let your fatigue build up. [Period effect]
> A: Maybe. What about you?
> B: Actually, I'm exhausted too! My body feels really heavy.
> A: You're kidding. You're still young. I could work all day long when I was your age.
> B: Oh, really?
> A: Yeah, young people these days are quick to whine. We were not like that. [Cohort effect]

Having made this distinction, it is worth noting that there are other ways in which change can be categorized. Firebaugh (2008) distinguishes between social change (studied with

repeated cross-sectional data) and individual change (studied with panel data). This can similarly be conceived as within- and between-individual change (Bell and Jones, 2012): an individual can age, or individuals can differ because they were born into a different cohort[1]. Thus individuals are nested within a context (their cohort group) and cohort change is considered across those contexts. Another conceptual distinction is between change within an individual and change by cohort replacement (Firebaugh, 1989, 2008). Here, cohort replacement effects are the result of old cohorts dying or leaving the sample, and young cohorts joining the sample (a mixture of age and cohort effects) and within-cohort change over time (which can be thought of as a mixture of age and period effects). Thus, whilst the age-period-cohort division remains conceptually attractive, and is linked in various ways to the divisions proposed here, it is not necessary to model all three in order to study social and individual change in a manner that is both robust and meaningful.

Furthermore, there is a problem in attempting to distinguish between the effects of these three sources of change in a statistical model. The three are perfectly correlated, such that:

$$Age = Period - Cohort$$

(1)

In such an equation we are always able to know the value of one of the three terms, if we know the value of the other two. As such, each of the following three (and an infinite number more) data generating processes (DGPs) will produce exactly the same data:

$$Y = Age + Period + Cohort$$

(2a)

---

[1] In a repeated cross-sectional study, the mean age of a cohort of individuals will be equivalent to the cohort variable itself (Freitas and Jones, 2012). Usually in such studies, period effects are assumed to be non-existent; this is done on the basis of theory regarding the processes that are thought to be operating.

Substituting $Period$ with $Age + Cohort$ gives us

$$Y = 2 * Age + 2 * Cohort$$

(2b)

And then substituting $Age + Cohort$ with $Period$ gives us

$$Y = 2 * Period$$

(2c)

As such, a DGP with equal age, period and cohort linear effects (3a) would create the same data as a DGP with (larger) age and cohort effects and no period effects (3b), and the same data as a DGP with just a single (larger) period effect (3c), because of the dependence expressed in equation 1.

Whilst different DGPs can be clearly conceived for linear effects, it is perhaps less obvious that a similar problem occurs with non-linear effects. For example, consider the following DGPs, each of which, again, will each produce identical data.

$$Y = Cohort^2$$

(3a)

Substituting $Cohort$ with $Period - Age$ and expanding gives us

$$Y = Period^2 - Period * Age + Age^2$$

(3b)

Then, substituting the second $Period$ term for $Age + Cohort$ and simplifying gives us

$$Y = Period^2 - Cohort * Age$$

(3c)

The same problem apples to other non-linear effects. Log effects are less easy to split apart because $\log(a + b)$ does not have an additive decomposition; so, for example, the term $\log(Period)$, whilst mathematically identical to $\log(Age + Cohort)$, cannot be split mathematically into separate age and cohort trends in a DGP. However, following Box and Draper (1987:424), who state that "all models are wrong, but some are useful", we argue that a real-life DGP is not the same as a mathematical approximation to it. As such, the presence in a DGP of a mathematical approximation that is 'un-confoundable' does not suggest that the real-life, non-mathematical process that it represents is correct.

It follows that, given a dataset, it is impossible to know which of the infinite possible effect combinations such as those listed above actually produced the data. Without the help of time machines or age accelerators (Suzuki, 2012), it is impossible to assess the effect of one of the terms independent of the others, because keeping any two of the APC terms constant automatically holds the third constant as well.

It is worth reiterating that this dependency lies not with a model that is fitted to the data, nor with the data itself, but with the underlying processes that created the data. The confounding is mathematical or logical, and as such cannot simply be solved by manipulation of the data or the model (Goldstein, 1979). Failing to realize this has led to a number of solutions being proposed which allow the model to be fitted but which produce arbitrary results. For example, Robertson and Boyle (1986) 'solved' the problem by aggregating the data in such a way to produce 'non-overlapping cohorts' which are not perfectly collinear. Similarly, Mason et al. (1973) propose a method which constrains two age groups, period groups or cohort groups to be aggregated together. However both of these methods, whilst able to produce *a* result, will not produce the correct result unless

that constraint or aggregation exactly matches constraints in the (unknown) DGP. Glenn (1976) shows that in the case of the Mason et al model, even non-linear effects will be biased where constraints are imposed arbitrarily (see also Glenn, 2005:14). And Osmond and Gardner (1989) show that different aggregations will produce wildly different results: unless those aggregations are present in the (unknown) DGP, they will produce incorrect results more often than not. Very solid theory about aggregations in the DGP is required for such methods to be of any use, and that theory is rarely forthcoming in applied research.

It is clear from the above that aggregation does not solve the identification problem; it simply hides it beneath coarser data. The fact that the model is able to estimate a solution simply means that the constraint or aggregation used is forcing the model to arbitrarily make a choice, on the basis of the researcher's arbitrary aggregation, not the data itself. Whatever parameterisation of the model are used to make the model fit, the problem of multiple possible underlying DGPs is not solved because it is the model that is being changed, not the DGP.

Despite these clear difficulties (impossibilities) in attempting to model age, period and cohort simultaneously, the desire to be able to explore all three terms remains, and as such there have been numerous attempts to 'solve' the problem, including Yang's Intrinsic Estimator (Yang, 2008b, Yang et al., 2004, Yang et al., 2008) and Tu's Partial Least Squares Regression (Tu et al., 2011). This paper looks at a single proposal, the Hierarchical Age Period Cohort (HAPC) model (Yang and Land, 2006). However, we hope that the results found in this paper should act as a cautionary tale to anyone considering using a model that claims to disentangle APC effects.

3 Yang and Land's HAPC model

The 'solution' to the APC identification model proposed by Yang and Land uses a multilevel model (also called a random effects, hierarchical linear, or mixed model) with a cross-classified structure. This structure treats individuals as nested within both periods and cohorts, and this must be conceived of as cross-classified because there is not exact nesting of periods into cohorts, or vice versa. As the data are (repeated) cross-sectional, an individual is only observed at one age and one period. Periods and cohorts are thus treated as random effect contexts at a higher level, in which individuals reside. Age is specified as an individual-level variable and is included in the fixed part of the model as a (potentially non-linear) function. The model structure can thus be expressed in a classification diagram as shown in figure 1.

[Figure 1 about here]

Conceptually, this presents a new angle to APC processes that is quite enticing. Treating periods and cohorts as contexts, and age as an individual characteristic, is intuitive to some degree because we move from one period *in*to another as time passes, and we *belong* to a cohort group that have common characteristics, whereas ageing is a process that occurs *within* an individual. Multilevel models in general assume that the higher level residuals are independently and identically distributed (IID). Thus, period effects are considered as, for example, 'the effect of being in 1990', independent of the periods around it, rather than 'the effect of moving from 1989 to 1990', a linear effect that is constant across all periods and so equal to the effect of moving from 1990 to 1991; similarly, cohort effects are considered in terms of 'the effect of being born in 1960' rather than 'the (linear) effect of

societal change between children born in 1959 and 1960.' Whilst periods and cohorts are

unlikely to be truly independent even if there is no trend (near cohorts are likely to be more

alike than cohorts that are far apart in time), techniques can be used to allow

autocorrelation to be taken into account in the model (Stegmueller, 2012). Crucially,

conceiving of separate periods and cohorts as quasi-independent contexts that have

(random) effects individually, rather than as part of an overall continuous linear fixed effect,

changes how the collinearity of APC works. In the cross-classified structure, because we are

modelling a different facet of period and cohort effects to that of the age effects (random

variation compared to a linear trend), we are able to model both a period and a cohort

effect for individuals, whilst still controlling for their age; the effects are not collinear in the

same way as they are for linear trends.

Given this conceptual structure, the model is specified[2] algebraically as follows:

$$y_{i(j_1 j_2)} = \beta_{0 j_1 j_2} + \beta_1 Age_{i(j_1 j_2)} + \beta_2 Age^2_{i(j_1 j_2)} + e_{i(j_1 j_2)}$$

$$\beta_{0 j_1 j_2} = \beta_0 + u_{1 j_1} + u_{2 j_2}$$

$$e_{i(j_1 j_2)} \sim N(0, \sigma_e^2), \qquad u_{1 j_1} \sim N(0, \sigma_{u1}^2), \qquad u_{2 j_2} \sim N(0, \sigma_{u2}^2)$$

$$(4)$$

The dependent variable, $y_{i(j_1 j_2)}$ is measured for individuals i in period j1 and cohort j2. The

'micro' model has linear and quadratic age terms, with coefficients β_1 and β_2 respectively, a

constant that varies across both periods and cohorts, and a level 1 residual error term. The

macro model defines the intercept in the micro model by a non-varying constant β_0, and a

---

residual term for each of period and cohort. The period, cohort and level-1 residuals are all assumed to follow a Normal[3] distributions, each with variances that are estimated.

Multilevel models also allow for additional levels to be included, where data is available. As previously stated, with repeated cross-sectional data a given individual is only observed at one age and in one period. However, if instead of repeated cross-sectional data we have panel data, an additional individual level could be included, thus allowing multiple observations (and so multiple ages and periods) per individual (Suzuki, 2012). Further, if the data has some kind of geographical indicator (such as neighborhoods, or countries), then these could also be considered as an additional level in this conceptual model. As such, Yang and Land's original 3-level model can (conceptually at least) be extended to account for other contexts in which measurements exist, leading to a more complex structure such as that shown in figure 2.

[Figure 2 about here]

In addition to extra levels in the model's random part, covariates can be added to the fixed part of the model, and these can vary at the individual level (in the micro model) or at the period or cohort levels (in the macro model). Random coefficients could be placed on any of the fixed part coefficients, so that, for example, the effect of age can be allowed to vary by cohort group. Whilst it is certainly true that you can have too complex a model, these possibilities make it seductive, where there is data that allow it and research questions that require such complexity.

---

[3] Non-Normal distributions are also possible. For example where the response variable is categorical, the level-1 residual would not be assumed Normal (see Jones and Subramanian, 2013).

However, such extensions are of little use if the model does not function as we wish it to. Conceiving of cohorts and periods as discrete temporal entities rather than as continuous temporal variables does not make them so in the DGP of real life. Any age, cohort or period *trends* (linear, or otherwise) in the DGP will still have the potential to be confounded by this process because multiple combinations of them will produce identical data, and no model is able to distinguish between identical datasets.

Yang and Land concur (Yang and Land, 2006) that the cross-classified structure does not solve the identification problem, as any age effect could in fact be a combination of period and cohort effects, and so on. They offer two solutions to this problem within the above framework; they suggest (1) specifying age as a quadratic polynomial, and (2) grouping cohorts into (for example) 5-year intervals. They argue that only one of these is necessary to allow identification and thus 'resolve' the problem outlined above:

> the underidentification problem of the classical APC accounting model has been resolved by the specification of the quadratic function for the age effects.

> Yang and Land (2006:84)

However, as the argument in section 2 hopefully made clear, we regard the ability of either of these techniques to solve the identification problem with some skepticism. Including a quadratic term in the model does not put it in the DGP, nor does not remove any linear term from the DGP. In any case, a quadratic term in the DGP can itself be confounded. Similarly, grouping the data does not put those groupings in the DGP and so it will not solve the identification problem either.

Yang and Land argue that the model can be estimated using maximum likelihood, but that, where there are few cohorts or few periods, the model should be estimated using Monte Carlo Markov Chain (MCMC) methods (Yang, 2006). In a later paper (Yang and Land, 2008), they also argue that there is a choice to be made between fixed and random effects for period and cohort terms. The advantages of random effects as specified above is that it uses fewer parameters, it is likely to be more efficient, and it is more easily extendable than the fixed effects version. The downside is that the model estimates will be biased if covariates are correlated with the higher level residuals. This is likely to be problematic here because if there is a cohort trend, with repeated cross-sectional data the age variable will always be correlated with the cohort residuals because older cohorts will have a higher age than newer cohorts, over the time period being studied (this point is not addressed in Yang and Land's paper[4]). A solution to this could be to include the group (cohort[5]) mean of the age coefficient (Bell and Jones, 2012, Mundlak, 1978); however in this case, the cohort mean of age is identical to the cohort variable itself. As with any model that has both age and cohort in the fixed part of the model, it must assume that there is no period trend.

4   Simulation design

The preceding discussion raises a number of issues regarding the situations in which Yang and Land's model works, and the situations in which it does not. We are particularly interested in how the model treats linear period or cohort trends, compared to randomly distributed effects that are assumed by the HAPC model. For the model to work in the

---

[4] Indeed, the paper uses an example where the FE and RE solutions produce quite clearly different results (the FE result is nine times the size of the RE result!). However by running a Hausman Test (Hausman, 1978) on all coefficients simultaneously rather than individually, they ignore the differences in the age coefficient and claim erroneously that there is no problem in this regard (Yang and Land, 2008:317-318, and table 4 on p.319).
[5] With repeated cross-sectional data, the period group mean will in general not vary a great deal, because of the way the data is sampled.

presence of linear effects, it must be able to accurately convert these linear[6] cohort or period trends into a separate random effect for each cohort or period. Further, we are interested in whether the grouping of cohorts aids in this conversion, as is claimed by Yang and Land. This includes not just assuming the presence of certain groups in the fitted model, but also in understanding how the model copes with actually occurring groupings in the DGP. This might occur if people born in a time-period of longer than a year share particular characteristics (for example, those born in the 1960s, or 'baby-boomers'). Finally, if the model is able to do the above successfully, we want to know the extent of bias caused by correlation between the age variable and the cohort residuals, and the ability of the methods suggested by Mundlak (1978) and Bell and Jones (2012) to solve these problems within the HAPC framework.

These issues can be reduced to the following five questions:

i. Does the HAPC model work when periods and cohorts are Normally distributed in the DGP?

ii. Does the model work when there is a linear trend in the period or cohort effects in the DGP?

iii. Does grouping of cohorts in the fitted model help to achieve the correct answer?

iv. What happens if there are groupings in the DGP as well as the fitted model (both matching those in the fitted model, and not)

v. Does including a linear cohort (mean age) term in the fixed part of the model remove bias from the random part, and solve any biased caused by correlation between the age variable and the cohort residuals.

---

[6] The same applies to non-linear period and cohort *trends* in the DGP, which would also need to be converted into discrete effects, and not reassigned into other APC combinations as in equation 3. However, for the sake of simplicity, in these simulations we focus on linear trends.

Thus, these five questions inform the design of the simulation scenarios outlined in table 1. The first scenario tests the model's capability where there are no period or cohort trends, thus answering question (i) above. The second scenario adds a cohort trend, allowing us to find an answer to question (ii). Scenario 3 will find an answer to question (iii) by grouping cohorts in the fitted model. Scenarios 4 and 5 will provide an answer to question (iv), assessing the affect of grouping in the DGP and the fitted model, both when those groupings match (4) and when they do not (5). Finally, scenarios 6 and 7 aim to answer to question (v), including a cohort term in the fixed part of the fitted model, and assessing the performance of the model in the presence of a cohort trend (6), and then a period trend (7).

[Table 1 about here]

For each scenario, 1000 datasets were randomly generated and the scenario model was fitted to each of these datasets. Each dataset consisted of 20,000 individuals, with a random uniform distribution of ages (between 20 and 60), and periods (between 1990 and 2010). Cohorts were calculated on the basis of these values, and the dependency expressed in equation 1. The dependent variables were generated with trends and residuals[7] as specified by the DGPs in table 1. The data were generated in Stata, and the models were estimated using MCMC[8] Gibbs sampling in MLwiN v2.25 (Rasbash et al., 2011) with the runmlwin (Leckie and Charlton, 2013) command in Stata. When using MCMC methods, a number of issues need to be considered. First, we must choose what starting values should

---

[7] Where cohorts were grouped, residuals were calculated [with the variance (of 1) as stated in table 1] and then averaged over those groupings, before being included in the DGP. For grouped cohort trends, cohort groups were assigned the value of their earliest year, before being included (centred) in the DGP [with the coefficient (of 0.1) as stated in table 1].

[8] We use MCMC rather than maximum likelihood methods because of the small number of periods that are in our sample (and which is typical of the data we use. Stegmueller (2013) finds that having only a small number of higher level units will significantly bias variance estimates using ML, whilst affecting MCMC estimations only slightly. See also Browne and Draper (2006).

be used. Here, as we are simulating our data, we know what the true values are. As such, we use these values as starting values. Whilst we would not have this information in reality, doing the simulations this way allows us to assess biases in the model unencumbered by any issues of convergence and bad starting values. If the model fails to work when we are telling it the answer, we can be pretty sure that something is seriously wrong!

Despite this, it is still important to test for convergence of each model that we run. In order to do this, we create a version of the Potential Scale Reduction Factor (PSRF) (Brooks and Gelman, 1998, Gelman and Rubin, 1992). The original version of this compares the variance of 5 different chains (from 5 different starting values) compared to the variance of the pooling of all five of these chains. As we are not interested in starting values here (and know that the starting values we are using could not be better), instead we use a single chain divided in 5, and compare the 95% coverage intervals of each of the five chains to that of the whole chain. In addition to this, we use the Effective Sample Size (ESS) for each parameter (which is automatically calculated by MLwiN) to assess whether the chain has been run for long enough for sufficient 'independent' draws to characterize the distribution of the parameter. Finally, we use hierarchical centering (Browne, 2009:401) at the cohort level to reduce autocorrelation in the chains. We have found that a chain of 100,000 iterations, with a 5,000 iteration burn-in, is in general more than sufficient to achieve good values for both the PSRF and the ESS for the vast majority of the simulations.

5   Simulation results

The full simulation results (and the Stata code used to create them) can be found online, and graphs representing the results of each of the seven simulation scenarios can be found

in figure 3. The first row of graphs shows the relative bias[9] of the estimates (the medians[10] of the monitoring chain) for each of the parameters in the model. If the model is fitting correctly, the median bias of each parameter should be zero. The second row shows the cohort shrunken residuals, where each line represents one of the 1000 simulation runs for each scenario. We expect these to appear as random white noise, except where there is a cohort linear trend in the DGP (i.e. scenarios 2-5), in which case we would hope to see that trend in the residuals (given the size of the linear trends in comparison to the size of the random variances in the DGPs). Similarly, the final row shows the period level residuals, and again, with the exception of scenario 7 where there is a period trend in the DGP, we would hope that these appear as random white noise.

[Figure 3 about here]

Here, the results are structured by the five questions outlined above.

5.1   Does the HAPC model work when periods and cohorts are normally distributed in the DGP?

Scenario 1 illustrates a situation in which there is no cohort or period trend, but there is random cohort and period variation. The model does not group cohorts, and nor is there any grouping of the data in the DGP. As can be seen in figure 3, this model estimates have with very little bias[11], and the model has not put any erroneous trends in the cohort or period residuals (they appear as white noise). As such in situations where there is no period

---

[9] Relative bias is calculated as the difference between the estimated and actual values, divided by the actual value

[10] The median was found by Browne and Draper (2006) to be the least biased point summary (rather than mean, or mode) of the MCMC parameter chains. However the results were not substantively different when the mean or model was used.

[11] There is some positive bias in the period variance term. This is caused by the small number of periods in the data.

or cohort *trend* in the data, the HAPC model seems to work, and this applies even when cohorts are ungrouped.

5.2 Does the model work when there is a trend in the period or cohort effects in the DGP?

Scenario 2 is the same as scenario 1, but with an additional linear cohort trend in the DGP. As can be seen the model suffers from severe bias. The age and cohort trends have effectively been combined by the model to create a period trend, as shown in the period residuals, and eliminating the age and cohort trends (which do not appear in the model estimates at all). This is unsurprising because the data would be identical if it had been created by a simple period trend. To reiterate, we know that the MCMC starting values we have used are as good as they can be, so these cannot be blamed for the poor performance of the model.

5.3 Does grouping of cohorts in the fitted model help to achieve the correct answer?

Scenario 3 is the same as scenario 2, except that cohorts are grouped into 5-year groups prior to estimation. Yang and Land (2006) argue that this can solve the identification problem, but as the graphs in column 3 of figure 3 show, this is not the case. The model suffers from exactly the same average bias as in scenario 2. What is different is that the model is much less reliable in how it assigns effects to age period and cohort. Whilst scenario 2 always (erroneously) assigned the entire trend to period instead of cohort and age, the results from scenario 3 produced a range of combinations of APC linear effects. Each of these combinations produces the same data so the combination that a given model

chooses cannot be based on the data or the DGP that created it. There is also some positive bias in the level 1 variance, but this is not hugely surprising – when the cohorts are grouped, the between cohort, within cohort-group cohort variance transfers to level 1.

5.4 What happens if there are groupings in the DGP as well as the fitted model?

Scenario 4 fits the same model as scenario 3, but the data was generated with cohorts grouped in the same 5-year intervals used to define cohorts by the model being fitted. As can be seen the model works relatively well. The model correctly assigns a trend to the cohort residuals and not to the period residuals. There is some bias (in this case, about 9%) in the estimation of the age coefficient; this is caused by correlation between the age variable and the cohort level residuals. But in comparison to the bias present in the other scenarios, this is relatively minor.

However, it is not enough to simply know that the cohorts are grouped in some way; it is necessary to know exactly how those groups are formed. Scenario 5 investigates mismatched groupings between the DGP and the fitted model, and it has the same problems of bias found for scenarios 2 and 3. In some cases we may be able to make educated guesses as to how cohorts are grouped – baby boomers may share characteristics, for example – but it would be very rare to be able to delineate accurate cohort groupings because we rarely have theory that is so exact in the social sciences. Whilst we may agree that the baby boomers share characteristics, it would be difficult to reliably say exactly where the baby boomers started and finished. As such, the success of the model in scenario 4 is rarely going to be of much use to an applied researcher, using real data and an unknown DGP.

5.5 What does including a linear cohort (mean-age) term in the fixed part of the model do?

We have seen in scenario 4 that, when we have solved the identification problem (albeit in a way that is rarely practical with real-life data) there remains some bias because the age variable is correlated with the cohort residuals. In their work comparing Fixed and Random Effects models, Bell and Jones (2012) argue that this bias can be solved by including the group mean of the biased variable, decomposing the variable's effect into a 'within' and a 'contextual' effect. However with the age variable here, its cohort mean is exactly collinear with the cohort variable itself. As such, including it in the model as a linear fixed effect is equivalent to including the cohort variable, and in doing so assuming that there is no linear period effect in the DGP.

When such a model is fitted (scenario 6), we see that the problem of bias in the age variable is solved. The age and cohort trends are correctly estimated in the fixed part of the model, whilst the random period and cohort variation is correctly estimated in the random part without the trend. The problem with this model is that it assumes there is no period trend. When this assumption is violated (as in scenario 7), the period trend is redistributed into age and cohort effects, which are then overestimated. As such this model is not able to correctly assign APC *trends*. Indeed, it is clear to us that no model can.

6 Discussion

We hope that this paper will function as a warning to those hoping to disentangle APC effects. The results make clear that no technical solution can break the logical or mathematical relationship of age, period and cohort without strong a priori assumptions

being imposed and being correct. Whilst we have only addressed one method here, other methods have also been proposed in recent years (Tu et al., 2011, Yang et al., 2008), and we would encourage anyone considering using them to run simulations of the sort used here first, before claiming a Panacea has been found.

Unsurprisingly, given the results of our simulation, we have found papers whose authors have been misled by the results that their HAPC model has produced. A recent example using the HAPC method (Dassonneville, 2012) looked at voter turnout volatility over time in the Netherlands. Much like the results from scenarios 2, 3, and 5 of our simulation study, they find a strong trend in their period-level residuals, which they argue goes against the prevailing view that societal change occurs by cohort replacement.

> The result that cohort effects are much less important in explaining volatility than period effects are, raises a number of questions with regard to previous findings. It is remarkable that research on time effects explaining the decrease in turnout in Western Europe does find generations to be crucial.
>
> Dassonneville (2012:9-10)

We suspect that, much like our simulation results, the period trend that they find is erroneous. It could easily have been produced by a combination of a cohort and an age effect.

This paper is not alone in findings of trends in periods and cohorts that are potentially problematic. Reither et al. (2009:1443) find that there is a very significant period trend in obesity; that "the pattern of predicted probabilities for U.S. adults shows a monotonic increase over time, with no sign of abatement in recent periods of observation".

Conversely, Schwadel (2010:13) argues that his model shows evidence for "a large across-cohort decline in [religious] service attendance when control variables are included in the model". Whilst Piontek et al. (2012) do not find a significant overall period or cohort variation in cannabis use, their results are suggestive of a period trend in some of their models, and they express surprise at the lack of a cohort trend. All of these results are questionable, with the APC effects found potentially being formed by a combination of other APC trends in the DGP.

Having said this, we do think that the conceptual structure that underlies the HAPC model may be valuable. It is necessary to understand the difference between a linear (or, indeed, a non-linear) trend and random variation. If there are linear (or non-linear) *trends* in the DGP, no model will be able to tell them apart, even if they are treated as random variation as they are in the HAPC model. However, the HAPC model is able to assess random variation in periods and cohorts, so long as any trends are absorbed in the fixed part of the model. These may well be of substantive interest, telling us, for example, that baby boomers have a higher level of literacy (a cohort effect), or that voter turnout in America was particularly high in the 1960 election between Kennedy and Nixon (a period effect); both of these results can be interpretations of the examples used by Yang et al. (2011)[12]. In both of these cases, it seems that the age terms in the fixed part of the model accounted for all APC trends[13]. However where this is not the case, a cohort term should be included in the fixed part of the model as in scenario 6 in this paper. This of course makes the

---

[12] The former example is also used in Yang and Land's other papers on HAPC models (Yang, 2006, Yang and Land, 2006, 2008). In addition, Yang's (2008a) paper on social inequalities in happiness also found minimal trends in the period and cohort residuals, meaning the results of that paper are probably justifiable.

[13] The authors do not recognise this, however; they do not distinguish between linear trends and random variation. In the case of the literacy example, they even argue that their results suggest "that there has been an intercohort decline in vocabulary knowledge" (Yang and Land, 2006:93).

assumption that there is no period trend, and this assumption, whilst often reasonable, must be made explicit.

Finally, it is worth pointing out that there are other challenges associated with using the HAPC model. We have not addressed issues of MCMC starting values and how sensitive the model is to these. The model also makes the assumption that periods and cohorts are independent of each other; this assumption is likely to be broken as near cohorts and periods will usually be more related than far apart ones. However, methods to overcome this have been suggested elsewhere (Stegmueller, 2012) so we do not address them here.

References

BELL, A. & JONES, K. 2012. Explaining Fixed Effects: Random effects modelling of time-series cross-sectional and panel data. Available at: http://polmeth.wustl.edu/mediaDetail.php?docId=1324 [Accessed 22nd August 2012].

BOX, G. E. & DRAPER, N. R. 1987. *Empirical model-building and response surfaces.* USA, Wiley.

BROOKS, S. P. & GELMAN, A. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics,* 7**,** 434-455.

BROWNE, W. J. 2009. *MCMC estimation in MLwiN, Version 2.25.* University of Bristol, Centre for Multilevel Modelling.

BROWNE, W. J. & DRAPER, D. 2006. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis,* 1**,** 473-513.

DASSONNEVILLE, R. 2012. Questioning generational replacement: an age, period and cohort analysis of electoral volatility in the Netherlands, 1971-2010. *Electoral Studies,* in press.

FIREBAUGH, G. 1989. Methids for estimating cohort replacement effects. *Sociological Methodology,* 19**,** 243-262.

FIREBAUGH, G. 2008. *Seven Rules for Social Research.* Princeton, NJ, Princeton University Press.

FREITAS, D. & JONES, K. 2012. Cohort change and individual development of aerobic performance during childhood: results from the Madeira Growth Study.

GELMAN, A. & RUBIN, D. B. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science,* 7**,** 457-472.

GLENN, N. D. 1976. Cohort Analysts Futile Quest - Statistical Attempts to Separate Age, Period and Cohort Effects. *American Sociological Review,* 41**,** 900-904.

GLENN, N. D. 2005. *Cohort Analysis,* 2nd edition*.* London, Sage.

GOLDSTEIN, H. 1979. Age, period and cohort effects - a confounded confusion. *Journal of Applied Statistics,* 6**,** 19-24.

GOLDSTEIN, H., BROWNE, W. & RASBASH, J. 2002. Multilevel modelling of medical data. *Statistics in Medicine,* 21**,** 3291-3315.

HAUSMAN, J. A. 1978. Specification tests in econometrics. *Econometrica,* 46**,** 1251-1271.

JONES, K. & SUBRAMANIAN, S. V. 2013. Developing multilevel models for analysing contextuality, heterogeneity and change using MLwiN 2.2, volume 2. Available at:

http://www.mendeley.com/download/public/18570951/48914753 21/f22217256ecd2c871df1f07ab058ac78bfe3c74c/dl.docx [Accessed 19th February 2013].

LECKIE, G. & CHARLTON, C. 2013. runmlwin: A program to run the MLwiN multilevel modelling software from within Stata. *Journal of Statistical Software,* forthcoming.

MASON, K. O., MASON, W. M., WINSBORO, H. H. & POOLE, K. 1973. Some Methodological Issues in Cohort Analysis of Archival Data. *American Sociological Review,* 38**,** 242-258.

MUNDLAK, Y. 1978. Pooling of time-series and cross-section data. *Econometrica,* 46**,** 69-85.

OSMOND, C. & GARDNER, M. J. 1989. Age, Period, and Cohort Models - Non-Overlapping Cohorts Dont Resolve the Identification Problem. *American Journal of Epidemiology,* 129**,** 31-35.

PIONTEK, D., KRAUS, L., PABST, A. & LEGLEYE, S. 2012. An age-period-cohort analysis of cannabis use prevalence and frequency in Germany, 1990-2009. *Journal of Epidemiology and Community Health,* 66**,** 908-913.

RASBASH, J., CHARLTON, C., BROWNE, W. J., HEALY, M. & CAMERON, B. 2011. MLwiN version 2.25. Centre for Multilevel Modelling, University of Bristol.

REITHER, E. N., HAUSER, R. M. & YANG, Y. 2009. Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States. *Social Science & Medicine,* 69**,** 1439-1448.

ROBERTSON, C. & BOYLE, P. 1986. Age, Period and Cohort Models - the Use of Individual Records. *Statistics in Medicine,* 5**,** 527-538.

ROBERTSON, C. & BOYLE, P. 1998a. Age-period-cohort analysis of chronic disease rates. I: Modelling approach. *Statistics in Medicine,* 17**,** 1305-1323.

ROBERTSON, C. & BOYLE, P. 1998b. Age-period-cohort models of chronic disease rates. II: Graphical approaches. *Statistics in Medicine,* 17**,** 1325-1339.

RYDER, N. B. 1965. The Cohort as a Concept in the Study of Social-Change. *American Sociological Review,* 30**,** 843-861.

SASAKI, M. & SUZUKI, T. 1987. Changes in Religious Commitment in the United-States, Holland, and Japan. *American Journal of Sociology,* 92**,** 1055-1076.

SCHWADEL, P. 2010. Age, Period, and Cohort Effects on US Religious Service Attendance: The Declining Impact of Sex, Southern Residence, and Catholic Affiliation. *Sociology of Religion,* 71**,** 2-24.

STEGMUELLER, D. 2012. Bayesian hierarchical age-period-cohort models with time-structured effects: an application to religious voting in the

US, 1972-2008. Available at: http://daniel-stegmueller.com/files/HierarchicalAPC_Draft_Stegmueller.pdf [Accessed 8th November 2012].

STEGMUELLER, D. 2013. How many countries to you need for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science,* Online.

SUZUKI, E. 2012. Time changes, so do people. *Social Science & Medicine,* 75**,** 452-456.

TU, Y. K., SMITH, G. D. & GILTHORPE, M. S. 2011. A New Approach to Age-Period-Cohort Analysis Using Partial Least Squares Regression: The Trend in Blood Pressure in the Glasgow Alumni Cohort. *PloS ONE,* 6.

YANG, Y. 2006. Bayesian inference for hierarchical age-period-cohort models of repeated cross-section survey data. *Sociological Methodology,* 36**,** 39-74.

YANG, Y. 2008a. Socail inequalities in happiness in the United States, 1972 to 2004: An age-period-cohort analysis. *American Sociological Review,* 73**,** 204-226.

YANG, Y. 2008b. Trends in US adult chronic disease mortality, 1960-1999: Age, period, and cohort variations. *Demography,* 45**,** 387-416.

YANG, Y., FRENK, S. M. & LAND, K. C. 2011. Assessing the significance of cohort and period effects in hierarchical age-period-cohort models. Available at: http://www.ssc.wisc.edu/cde/demsem/07-08-11TestingStatSigInHAPCModel_Manuscript.PDF [Accessed 22nd August 2012].

YANG, Y., FU, W. J. 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.

YANG, Y. & LAND, K. C. 2006. A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological Methodology,* 36**,** 75-97.

YANG, Y. & LAND, K. C. 2008. Age-period-cohort analysis of repeated cross-section surveys - Fixed or random effects? *Sociological Methods & Research,* 36**,** 297-326.

YANG, Y., SCHULHOFER-WOHL, S., FU, W. J. J. & LAND, K. C. 2008. The intrinsic estimator for age-period-cohort analysis: What it is and how to use it. *American Journal of Sociology,* 113**,** 1697-1736.

Table 1: DGPs and fitted models for ten scenarios for simulation. If the model works, the estimate of the cohort variance (σ_{u1}^2, that is the variance of the residuals u_{j_1}) in the fitted model should include the combined variance of the random cohort term (u_c) and the linear cohort trend ($0.1 * Cohort$) when they are both present in the DGP. The bracketed superscripts on the cohort terms in both the DGP and the fitted model refer to the cohort groupings. So a grouping of 5 on the cohort terms in the DGP means that cohorts were grouped into 5-year intervals and these groups used as the units making up the cohort normal distribution (u_c) and/or linear effect ($Cohort$) found in the DGP. In the fitted model, a grouping of 5 on the cohort residual term means that the fitted model uses 5 year groups to define the cohort's random effects units (u_{j_1}). A value of 1 in either column means there is no grouping. All variables in both the DGP and the fitted model were grand-mean-centered.

Scenario	DGP equation	Fitted model	Discrepancy between DGP and fitted model
1	$y = 1 + (0.1 * Age) - (0.005 * Age^2) + u_c^{(1)} + u_p + e_{L1}$	$y_{i(j_1 j_2)} = \beta_{0 j_1 j_2} + \beta_1 Age_{i(j_1 j_2)} + \beta_2 Age_{i(j_1 j_2)}^2 + e_{i(j_1 j_2)}$ $\beta_{0 j_1 j_2} = \beta_0 + u_{j_1}^{(1)} + u_{j_2}$	None
2	$y = 1 + (0.1 * Age) - (0.005 * Age^2) + \left(0.1 * Cohort^{(1)}\right) + u_c^{(1)} + u_p + e_{L1}$	As scenario 1	Cohorts in DGP not IID
3	As scenario 2	$y_{i(j_1 j_2)} = \beta_{0 j_1 j_2} + \beta_1 Age_{i(j_1 j_2)} + \beta_2 Age_{i(j_1 j_2)}^2 + e_{i(j_1 j_2)}$ $\beta_{0 j_1 j_2} = \beta_0 + u_{j_1}^{(5)} + u_{j_2}$	Cohorts in DGP not IID; different cohort groupings
4	$y = 1 + (0.1 * Age) - (0.005 * Age^2) + \left(0.1 * Cohort^{(5)}\right) + u_c^{(5)} + u_p + e_{L1}$	As scenario 3	Cohorts in DGP not IID
5	$y = 1 + (0.1 * Age) - (0.005 * Age^2) + \left(0.1 * Cohort^{(7)}\right) + u_c^{(7)} + u_p + e_{L1}$	As scenario 3	Cohorts in DGP not IID; different cohort groupings
6	As scenario 2	$y_{i(j_1 j_2)} = \beta_{0 j_1 j_2} + \beta_1 Age_{i(j_1 j_2)} + \beta_2 Age_{i(j_1 j_2)}^2 + e_{i(j_1 j_2)}$ $\beta_{0 j_1 j_2} = \beta_0 + \beta_3 Cohort_{j1}^{(1)} + u_{j_1}^{(1)} + u_{j_2}$	None
7	$y = 1 + (0.1 * Age) - (0.005 * Age^2) + (0.1 * Period) + u_c^{(1)} + u_p + e_{L1}$	As scenario 6	Periods in DGP not IID
	$u_c \sim N(0,1)$ for cohorts, $u_p \sim N(0,1)$ for periods, and $e_{L1} \sim N(0,4)$	$e_{i(j_1 j_2)} \sim N(0, \sigma_e^2)$, $\quad u_{j_1} \sim N(0, \sigma_{u1}^2)$, $\quad u_{j_2} \sim N(0, \sigma_{u2}^2)$	

Figure 1: Cross-classified structure of Yang and Land's HAPC model for repeated cross-sectional data.
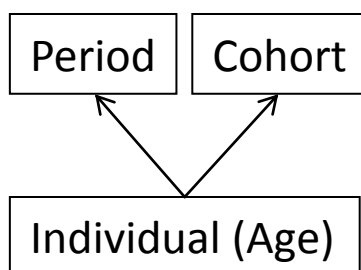

Figure 2: Potential extension to the structure of Yang and Land's HAPC model. Where panel (rather than repeated cross-sectional) data is available, there are multiple occasions (and thus multiple periods and ages) per individual. Individuals are still nested within birth cohorts. Those individuals can also be nested within geographical units (here, countries) where the data is present.
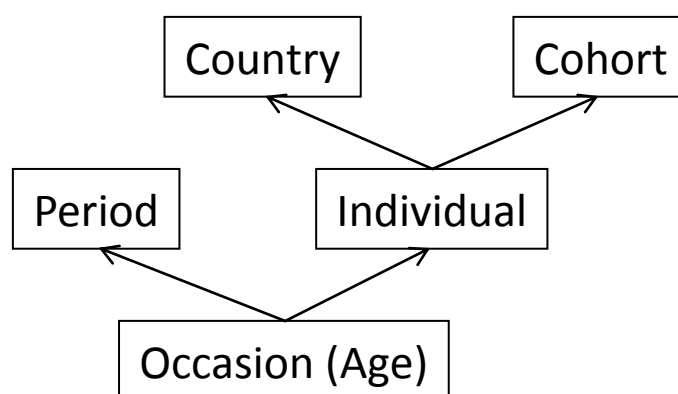
Figure 3: Results for each of the simulation scenarios explicated in table 1. *Row 1*: Relative bias (median, and 95% coverage intervals[14]) of each of the parameters in the model; corresponding to the fitted model equations in table 1, 0 is β_0 (coefficient associated with the constant), 1 is β_1 (coefficient associated with the linear age term), 2 is β_2 (coefficient associated with the quadratic age term), 3 is β_3 (coefficient associated with the linear cohort term, where present in the model), 4 is σ_{u1}^2 (cohort level variance), 5 is σ_{u2}^2 (period level variance), and 6 is σ_e^2 (level 1 variance). If the model is unbiased, all the medians should lie on the zero line. *Row 2*: Cohort-level shrunken residuals as estimated for each of the 1000 simulations. *Row 3:* Period-level shrunken residuals as estimated for each of the 1000 simulations.

---
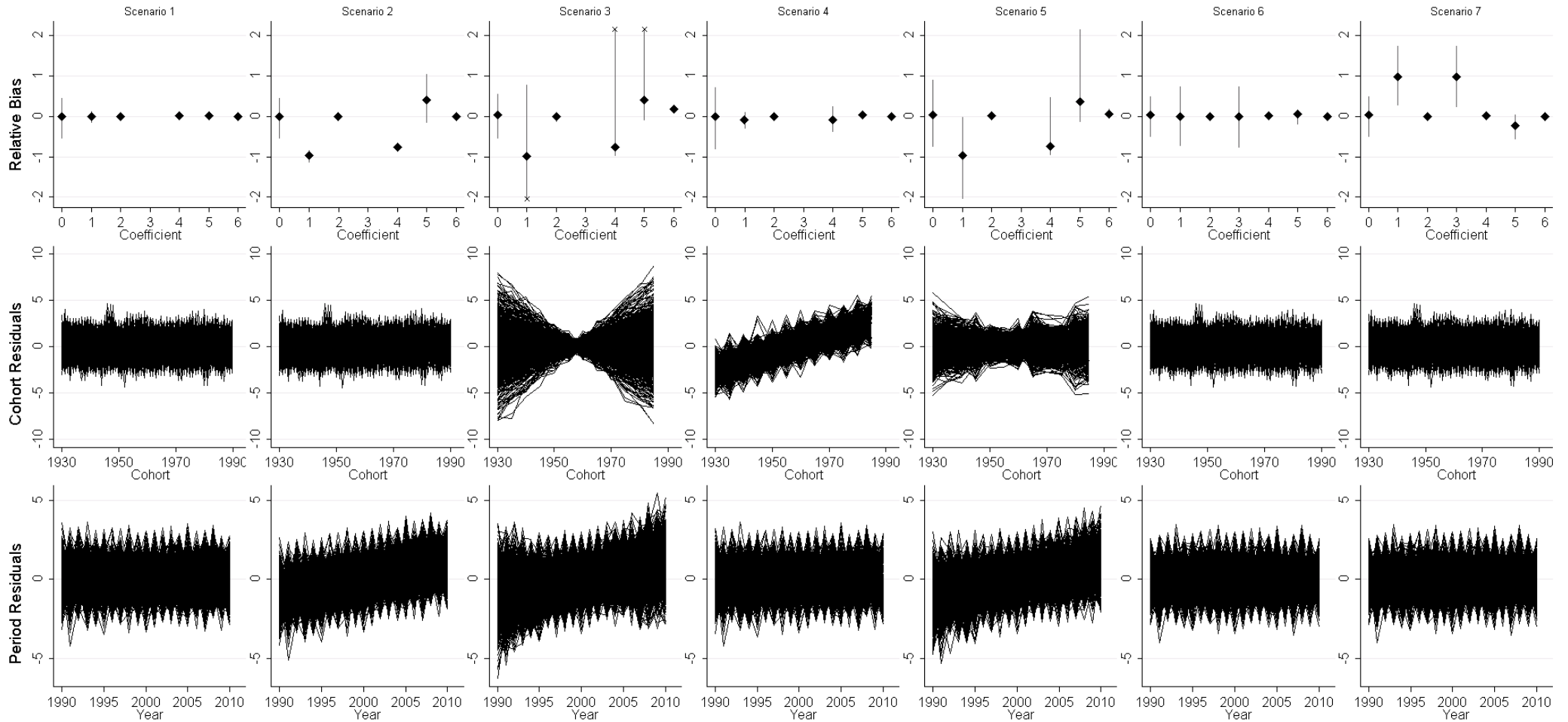
[14] Some of the parameter estimates for scenario 3 had relative biases with coverage intervals extending beyond the range of the graph; these have been cropped out, as indicated by 'x'.