Should age-period-cohort analysts accept innovation without scrutiny? A response to Reither, Masters, Yang, Powers, Zheng and Land

Andrew Bell a b and Kelvyn Jones a b

a School of Geographical Sciences
University of Bristol
University Road
Bristol
BS8 1SS
UK

b Centre for Multilevel Modelling
University of Bristol
2 Priory Road
Bristol
BS8 1TX
UK

Corresponding Author:
Andrew Bell
School of Geographical Sciences
University of Bristol
University Road
Bristol
BS8 1SS
UK
Email: andrew.bell@bristol.ac.uk

Acknowledgements: Thanks to Ron Johnston for his helpful advice.
Abstract

This commentary clarifies our original commentary (Bell & Jones, 2014c) and illustrates some concerns we have regarding the response article in this issue (Reither et al., 2015). In particular, we argue that (a) linear effects do not have to be produced by exact linear mathematical functions to behave as if they were linear, (b) linear effects by this wider definition are extremely common in real life social processes, and (c) in the presence of these effects, the Hierarchical Age Period Cohort (HAPC) model will often not work. Although Reither et al. do not define what a ‘non-linear monotonic trend’ is (instead, only stating that it isn’t a linear effect) we show that the model often doesn’t work in the presence of such effects, by using data generated as a ‘non-linear monotonic trend’ by Reither et al. themselves. We then question their discussion of fixed and random effects before finishing with a discussion of how we argue that theory should be used, in the context of the obesity epidemic.

Research Highlights

- We clarify the nature of the identification problem in all APC analysis
- The Hierarchical APC model will sometimes work, but sometimes is not enough
- Simulations using plausible data structures show the model often does not work
- Relying in theory is problematic, but this is often all researchers can do

Keywords

Age-period-cohort models, obesity, collinearity, model identification, cohort effects, multilevel modelling
We thank the *Social Science & Medicine* editors for allowing us to respond to the above article and allowing the debate regarding age-period-cohort (APC) identification to be furthered. In their article, Reither et al. (2015, henceforth RMYPZL) argue the following:

- Only when period and cohort effects are exactly linear does the Hierarchical APC (HAPC) model give fallacious results.
- In the real world, period and cohort effects are never exactly linear.
- Thus, in the real world, the HAPC model will work.
- The HAPC model should only be used when goodness-of-fit statistics (such as AIC and BIC) suggest a simpler model (including only one or two APC dimensions) would be insufficient.

We address each of these arguments in turn below, showing that each of them is flawed, and that our original critique of the HAPC model (Bell & Jones, 2014c) remains justified.

What is a linear trend, and what is a non-linear monotonic trend?

RMYPZL argue that the HAPC model will only fail to produce accurate results in the presence of linear effects, and we agree with this. But what is a linear effect? One answer, and that suggested by RMYPZL, is that it is a process produced by an exact linear algebraic association: $y=mx+c$. However, in the real world data are generated by social processes, not equations, meaning RMYPZL are right to claim that such effects never occur exactly in real life. However, our definition of a linear trend is wider than this: we argue that a linear trend exists when, if an algebraically linear expression is removed from the data at hand, this would have the effect of flattening that trend. It would be difficult to argue that social processes never produce data that fills these criteria. Furthermore, there need only be a linear component to the data generating process to fulfil this definition. Other effects (stochastic, quadratic, or whatever else) can also be present, so long as there is also a linear component, defined as above. Whilst the HAPC model will sometimes work under these circumstances, as shown by the simulations in RMYPZL, we argue that it will often not work. We also argue that a model that only
works some of the time is not a particularly useful model to social scientists, and at least needs to be applied with care and awareness of its limitations.

As for what a ‘non-linear monotonic trend’ is, the answer is less clear. RMYPZL state what it is not, but do not say what it is. This is convenient for their argument: it means that there is no possibility of questioning the model with simulations because any simulated DGP which the HAPC model fails to replicate can be dismissed as being unrealistically linear. All that we have to go on are the six trends (period and cohort trends from each of equations 2-4 in RMYPZL) which we must therefore assume are examples that are consistent with ‘real life’ situations.

Some counter-simulations

However, RMYPZL’s argument does not even stand up for these six trends. We generated data where the period trend was the same as that in RMYPZL’s equation 2, the age trend was the same as that in RMYPZL’s equation 4, and the cohort trend was based on the period trend in figure 4 (we took every other period effect so the numbers matched the numbers necessary for 7-year cohort groupings). Thus, the DGP used was as follows:

\[
\text{Logit}[\text{Pr}(Y=1)] = -1.988 + (0.059\times\text{Age-gm}) + (-0.001\times(\text{Age-gm})^2) + (-0.474\times\text{C1}) + (-0.423\times\text{C2}) + (-0.362\times\text{C3}) + (-0.314\times\text{C4}) + (-0.214\times\text{C5}) + (-0.135\times\text{C6}) + (-0.054\times\text{C7}) + (0.029\times\text{C8}) + (0.172\times\text{C9}) + (0.256\times\text{C10}) + (0.410\times\text{C11}) + (0.529\times\text{C12}) + (0.605\times\text{C13}) + (-0.02\times\text{P1}) + (0.03\times\text{P2}) + (0.04\times\text{P3}) + (0.04\times\text{P4}) + (0.02\times\text{P5}) + (-0.03\times\text{P6}) + (-0.03\times\text{P7}) + (-0.05\times\text{P8}) + (-0.05\times\text{P9}) + (-0.05\times\text{P10}) + (-0.05\times\text{P11}) + (-0.05\times\text{P12}) + (-0.06\times\text{P13}) + (-0.05\times\text{P14}) + (-0.05\times\text{P15}) + (-0.02\times\text{P16}) + (0\times\text{P17}) + (-0.02\times\text{P18}) + (-0.02\times\text{P19}) + (-0.02\times\text{P20}) + (0\times\text{P21}) + (-0.02\times\text{P22}) + (0.02\times\text{P23}) + (0.05\times\text{P24}) + (0.08\times\text{P25}) + (0.1\times\text{P26}) + (0.1\times\text{P27}) + \text{U}_c + \text{U}_p
\]

\[
\text{U}_c \sim \text{N}(0, .01), \text{U}_p \sim \text{N}(0, .01)
\]

We fitted these data to the HAPC model 100 times, specifying 5-year groupings in the model. If RMYPZL are correct, unbiased results should be produced because the DGP contains, by their own definition, no linear effects and only ‘non-linear monotonic trends’ for periods and cohorts.
The results are shown in figure 1. As can be seen, the model fails to pick up the cohort trend, finds a period trend where there is none, and underestimates the strength of the age trend; in sum, the HAPC gets it radically wrong in not identifying the true patterns.

[Figure 1 about here]

The use of fit statistics

RMYPZL argue that fit statistics should be used to check that all of the elements of APC are required. Whilst some of the authors have stated this in the past with regard to the Intrinsic Estimator (e.g. Yang & Land, 2013:126), in neither their articles nor their book (as far as we are aware) have they stated that this is necessary before using the HAPC model. Indeed they have regularly claimed that the HAPC approach “completely avoids the identification problem” (Yang & Land, 2013:70) without any such ifs or buts. Thus, many researchers will (and have) used the HAPC model without taking this step. Consequently, we welcome this important clarification.

However, there is a problem with this. In each of the models presented in table 1 of RMYPZL, there are age, period and cohort effects present in the DGP s, even if those effects are only random variation (generated by the U_c and U_p coefficients). As such, in all four of the simulated cases, the model fit statistics find the incorrect answer. Moreover, in two cases, different model fit statistics give different answers. This is unsurprising given our previous arguments about the APC identification problem (Bell & Jones, 2013, 2014a): model fit statistics will never be able to solve the identification problem, because they cannot tell the difference between DGPs with different linear (by our definition) APC effects. Model fit statistics showing the full APC model is preferable only suggests that there is significant non-linear variation present in each of the three dimensions – it does not make it possible to assign linear (by our definition) trends correctly.
Fixed and Random effects

A theme that runs through RMYPZL’s article is the question of whether one should use fixed (FE) or random (RE) effects models. We want to emphasise that this is a separate issue to the identification problem and RMYPZL conflate the two in their article, which distracts from the issue at hand. We agree with the conceptual treatment of cohorts and periods as random effects. But an appropriate conceptual treatment does not mean the model works in practice. RMYPZL claim that we “assume” the classical APC accounting/linear regression model where “the effects of all three temporal dimensions are fixed [effects]” (RMYPZL:6). This is not the case; indeed this claim does not really make sense. Statistical models are not assumed; they are used to appropriately represent the social processes that produced the data at hand. Our argument is simply that in many situations the model used by RMYPZL fails to accurately represent these processes. Should we use FE or RE? Whilst in other scenarios we have actually argued strongly for the latter (Bell & Jones, 2015b), if we are looking for an automatic, general solution to the identification problem, the answer is that we should use neither.

The problem is that RMYPZL are completely wrong when they say that the identification problem “is not data specific, but model specific” (RMYPZL:4) – if linear trends (by our definition) exist in the dataset, you will encounter problems regardless of what model you use, whether a RE HAPC, a FE accounting model, or whatever else.

Solid Theory

Finally, we want to thank RMYPZL for their discussion of solid theory which, in general, we agree with. In our article we did not give an opinion one way or another regarding which of period or cohort effects are more likely to be responsible for the obesity epidemic. Our point was simply that this question cannot be answered using the data and methods of Reither et al. (2009). We therefore welcome RMYPZL’s theoretically informed review of previous studies of obesity. Indeed that is exactly what we called for in our original commentary (Bell & Jones, 2014c).
In particular, RMYPZL cite two papers (Flegal et al., 2002; Lee et al., 2011) which find a non-linear trend in periods: obesity/BMI appears relatively flat until about 1980 and then increases from there. It is unlikely that such a nonlinearity could be the result of cohort effects. We do not dispute this logic except to make two points. First, another study, albeit in a different cultural context, have used a similar analysis of non-linearities, and found that cohorts fit best (Olsen et al., 2006). Second, the non-linearities were not, and could not, be found by Reither et al’s study, because their data only went as far back as 1976 and so no such non-linearities were present.

Our point about strong theory was not that it is a solution to the APC identification problem. We agree that “compelling speculation can never replace evidence in any field of scientific inquiry” (RMYPZL:22). However this doesn’t mean that any evidence will do, and where the choice is between compelling speculation and misleading evidence dressed up as science, we would always choose the former.

Conclusions

RMYPZL ask the question “Should APC studies return to the methodologies of the 1970s?” The answer to this rather loaded question is clearly no. However that previous methods don’t work does not mean that new innovations do, and this discussion should guide readers in making their own minds up about whether the HAPC model is appropriate for their purposes. We have argued before that there are some situations where the HAPC model could be used (e.g. when periods and cohorts have no continuous trends) and it can easily be adapted to incorporate theory where appropriate (Bell, 2014; Bell & Jones, 2014b, 2015a). However the model does not work as a general purpose APC model; no model does. Our concerns mirror those of others regarding another the Intrinsic Estimator (Luo, 2013; Pelzer et al., 2014) and we agree with Fienberg (2013:1983) that “the search for methodological solutions to the APC identity is an endless and fruitless quest. It is surely time to move onto substantively focused considerations of the meaning of the three components in settings of interest.”
References


Figure 1: The DGP (black) and results (grey) from fitting the HAPC model to 100 datasets generated as in equation 1.