A Challenge from Christopher Winship (Harvard) and Ethan Fosse (Toronto):

Age-Period-Cohort Analysis: A Way Forward

Since its beginnings nearly a century ago, age-period-cohort (APC) analysis has been stymied by the lack of identification of parameter estimates arising from the linear dependence among age, period, and cohort (age = period – cohort). Researchers have dealt with this in two main ways. First, researchers have imposed an identifying constraint either by setting some set of parameters values to be equal or requiring some other mathematical property to hold. In practice, such constraints have been difficult if not impossible to defend. Second, more recently and less commonly, analysts have focused only on the nonlinear effects, which are identified and are often of substantive interest.

In a series of articles, we have developed a set of methods that allows the analysis of APC to move forward despite the identification problem. Specifically:

1. We show that the data **without any assumptions** rule out certain theoretically plausible combinations of linear effects.

2. Because the nonlinear effects are identified, parameterizing one's model so that the linear effects are in some way orthogonal to the nonlinear effects reduces the number of unidentified parameters to just three.

3. The claim sometimes made that only two dimensions of the three APC dimensions are needed to model a set of data rests on the assumption that an absence of nonlinear effects (which can be tested) implies an absence of the corresponding linear effect. This assumption needs to be theoretically justified.

4. Monotonicity assumptions (e.g., that the risk of death increases with age after 35) imply bounds on the APC effects. Often such assumptions are theoretically plausible. In our experience these bounds can be quite tight, giving something close to point estimates. Of course, any assumptions need to be theoretically justified.

5. Observed data on the mechanisms associated with each of the three APC variables can help in either providing full identification or narrowing bounds. In the former case, the needed assumptions are at least partially testable.

6. We provide a rigorous definition of what it means for age, period, or cohort to have "causal effects" consistent with the potential outcome (or counterfactual) framework.

Taken together we believe that the above work provides a solid methodological foundation for APC analysis, one that has not existed previously. By making this statement, we mean that we have developed a set of methods that can produce substantively important results in which the assumptions involved are both explicit and likely to be plausible.

After nearly a century of effort this is a big claim. How might we test it? In mathematics, if someone claims to have proved a theorem, the proof is not considered valid until others have rigorously analyzed it. Our request and hope with our claim to have provided a solid methodological foundation for APC analysis is that others will interrogate our claim with similar rigor.

To be clear, we are not claiming to have solved all the issues involved in APC analysis. We most certainly have not solved the APC identification problem. It is unsolvable. Methods for examining just the nonlinear effects have a great deal of potential. There is the question of how deal with interactions among the three APC variables, a situation in which identification issues can be even more daunting. There are also many other issues that APC analysis shares with standard regression analysis, such as how to deal with unobserved/unmodeled causal heterogeneity.

What we are claiming is that we have developed a set of methods that allows researchers to get around and go beyond the classic APC identification problem. Of course, in the end the real test is whether the methods we propose are useful for doing actual empirical analyses. As described below we have a beta versions of R programs for analyzing APC data.

We currently have five published papers and two others that are forthcoming. They can be found at https://scholar.harvard.edu/cwinship/age-period-cohort Our *Annual Review* paper per provides a full introduction to these methods. The *Sociological Science* paper examines critical technical issues. Our *Demography* paper shows how bounding analysis can be done. Chris' paper with Lying and others shows how Intrinsic Estimator estimates are a function of the coding scheme used. Chris' older paper with David Harding illustrates the potential importance of explicitly modeling the mechanisms involved in an APC model. The paper by Fosse, Winship and Daoud provides a more general introduction to our methods. Finally, Ethan's paper shows how the identification problem can be managed using Bayesian methods. Here, there are close links to the bounding method described in depth in our *Demography* paper.

Methods without the software to employ them are not very useful. As already noted, we have very early beta versions of R programs. We are most eager for people to use the programs both to try these methods out and to test the usability of the scripts themselves. The programs can be found at the bottom of the Home Page.

To run the R-scripts, you must have R installed. Copy the zip folder labeled "APC-R-Software" on to your computer and open it. Open the README file for instructions on how to run the software.

Papers

Fosse, Ethan, and Christopher Winship. "Analyzing Age-Period-Cohort Data: A Review and Critique." *Annual Review of Sociology* 45 (2019).

Fosse, Ethan, and Christopher Winship. "Bounding analyses of age-period-cohort effects." *Demography* 56.5 (2019): 1975-2004.

Fosse, Ethan, and Christopher Winship. "Moore–Penrose Estimators of Age–Period–Cohort Effects: Their Interrelationship and Properties." *Sociological Science* 5 (2018): 304.

Luo, Liying, James Hodges, Christopher Winship, and Daniel Powers. "The sensitivity of the intrinsic estimator to coding schemes: comment on Yang, Schulhofer-Wohl, Fu, and Land." *American Journal of Sociology* 122, no. 3 (2016): 930-961.

Winship, Christopher, and David J. Harding. "A mechanism-based approach to the identification of age–period–cohort models." *Sociological Methods & Research* 36.3 (2008): 362-401.

Fosse, Ethan, Christopher Winship and Adel Daoud. Forthcoming. "Learning from Age-Period-Cohort Data: Bounds, Mechanisms, and 2D-APC Graphs." In Andrew Bell, editor, *Age, Period and cohort effects: the identification problem, and what to do about it.* Taylor Francis.

Fosse, Ethan. forthcoming. "Bayesian Age-Period-Cohort Models." In Andrew Bell, editor, *Age, Period and cohort effects: the identification problem, and what to do about it.* Taylor Francis.