Approximating Age-Period-Cohort Estimates by Averaging Multiple Models with Varying Window Restrictions

# Abstract

Key words: Computational Sociology, Methods, Life Course, Bayesian, Culture

# Introduction

# Background

Both classic and modern studies have proposed different effects across three social dimensions of time: age, period, and cohort (Ryder 1965). These dimensions of time operate on individuals in different ways as they flow through social space. Age effects are driven either by the common experiences associated with the biological process of aging (Jackson, Weale, and Weale 2003), or persistent age-structuring institutions, like high school or retirement (Leisering 2003; Moen 2014). Period effects are responses of everyone to contemporaneous social experiences, like recessions or wars (Lam, Fan, and Moen 2014). Finally, cohort effects are socialization effects. They are broad-based historical events that stick to the populations experiencing them, even after the event has long since ended (Vaisey and Lizardo 2016). In contemporary research, cohorts are virtually always defined in terms of birth year.

These three dimensions of time sit at the center of some recurrent debates. For example, cultural sociologists are divided over whether the main driver of culture is a period process (cultural fragmentation) or a cohort process (acquired dispositions) (Vaisey and Lizardo 2016). Clinical reasearchers and biodemographers are looking for biological age—indicators of biological deterioration—but argue about confounding from cohort changes (Jackson et al. 2003). And yet others are concerned about separating long-term and short-term impacts of important events, like the Great Recession (Burgard and Kalousova 2015).

Unfortunately, it is not trivial to estimate the unique effects of age, period, and cohort to find evidence that may settle these questions and debates. The fundamental problem is one of statistical identification: these three dimensions are linearly dependent. Two of the dimensions define the third. If a researcher knows an individual’s age, and the year of the survey, cohort is also defined. Because these variables are exactly colinear, they are not estimable using classic statistical techniques. A number of solutions have been proposed to address this conundrum, these include traditional methods like block/window constraints (Anon. 2005), and new methods, including statistical transformation (the “intrinsic estimator”), and random effects models (Yang and Land 2013).

The past several years have seen a resurgence in debates surrounding these issues (Bell and Jones 2017; Luo 2013; Luo and Hodges 2016). Perhaps the most common argument is that the “constraints”, *i.e.* assumptions used to break the APC identity, produce unknown biases into the models. In reality, this presents an extreme case of multicollinearity, a well-known, though difficult problem in statistical theory (Wooldridge 2009, at p. 95-98), with numerous cautions in applied texts that separating highly colinear effects is difficult (Camm 2016, at p. 328). From a Bayesian perspective, collinearity and multicollinearity significantly reduce the ability of the data to supply information to produce the estimates, potentially introducing sensitivity to the prior (Gelman 2014, at p. 305-306), including assumtpions of linearity. To put it another way, the critique of APC models in general, and window constraints in particular (Luo and Hodges 2016), is that different sets of assumptions lead to different (and inconsistent) results. But, what if Instead of focusing on finding a *best* fitting model, we use a simple nonlinear approach (block modelling), and .

These models make a set of simple and straightforward assumptions that, if true, moot the identification problem. In particular, they assume that *some* of the unique ages, periods, and cohorts have identical effects. This equality blocks the age, period, or cohort together, and breaks the exact APC identity. The major criticism of these models is that a constraint in one of the variables induces unkown (and difficult to test constraints) in the other estimators []. Instead of relying on an arbitrrary set of constraints, our averages over thousands of constraints, weighted by an approximation to the probability of the model terminology using Bayesian Model Averaging (BMA) algorithms.

# Description of the Method

revise here

## Blocking Windows: Piecewise Constant Functions

intro on piecewise constant functions

In terms of window constraints, the target model, which is inestimable, is built by estimating each unique value of age, period, and cohort, as a dummy variable:

where is an estimated effect, , and index unique values for age, period, and cohort, and , , and stand for matricies of dummy variable series for age, period, and cohort. This model is unidentified, because, as with the continuous case, the dummy variables in any two of the matricies above condition the third matrix. In other words, the indicator variable in is a function of the indicators in and (in mathematical terms,the probability that any given cohort dummy variable is one or zero is exctly dependent on the values of A and P, so that ) is either exaclty zero, or exactly one).

This dependency is broken, however, by combining one or more of the unique values together so that they share the same dummy variable series. By way of example, we can construct an age dummy variable series where is sliced into two groups based on some cut-point, say , so that …

work here

This would identify a binary dummy variable with an older and a younger group. We can generalize this expression to an arbitrary vector of cut-points, with subscript so that

If the vector and have the following properties: , , , and , then can describe any posible sets of window restrictions for age. Generalizing cross all dimensions of APC, permuting three similar vectors (say ) will describe any model for any possible window constraints detailed in equation 1. Accordingly, all permutations of G, as defined above, constitute the model space of window constraints (). For any given set of data, is finite, but it can be quite large. For example, 10 unique ages, periods, and cohorts present \_\_ possible models. describe how Of these\_\_ models, only 1 is not estimable because of perfect colinearity. This target model is (theoretially) the least biased, although it is not the most parsimonious model. The question is how to best use infromation from some subset fo possible models in to estimate unbiased APC effects. Bayesian Model Averaging (BMA) provides a straightforward way to combine models. The theoretical backdrop applies to this curcumstance quite well. In principal, there is no one true (or best) model; instead estimates are conditional on models from the modeling space (), and have a posterior distribution, which is calculated as a weighted average of all models (Raftery 1995:144–45). It has been applied in diverse areas from weather forecasting to biology to social science (Fragoso and Neto 2015) .

## Bayesian Model Averaging (BMA) and the MC3 Algorithm

Markov Chain methods for Bayesian Model Averaging (BMA) provides a sensible way to sample over a subset of continually better-fitting models and combine their estimates to produce an approximation of APC effects. The next two sections describe the MC3 agorithm developed for BMA, and outline the unique implementation of MC3 for this particular set of models, drawing from the Dirichelet distribution.

## Using the Dirichlet Distribution to Sample Window Groups (G)

We simulate using two sets of nuisance parameters. For each dimension () of APC, the window breaks, of equaiton 2 are decomposed into (1) a cumulative sum from a Simplex for each dimension, with the same length of unique elements in (), and (2) a scalar integer, . is simply the product of times . We sample from the uniform distibution, as follows:

Where is the index number for the APC effects (the , , or of equaiton 1) is the continuous vector of values (the ,, or ). We sample the weights using a cumulative sum from a dirichelet distribution:

The for the Dirichelet distribution are for each of the vunique values in the dimension () for APC. The Dirchelet distribution draws a random projection on the standard simplex, *i.e.* a vector of weights weights between 0 and 1 which sum to 1. Multiplying the vector of weights () by a scalar provides a set of numbers which sum to the scalar value . We transform and into matrix by selecting the unique floor rounded values of the product of the cumulative sums of :

Where is a vector of the type described in equation 2, and indexes and is a subset of the unique values in .

# A Simulation

# An Empirical Example from the GSS

# Discussion

# Conclusion

Anon. 2005. “Strategies for Estimating Age, Period, and Cohort Effects.” Pp. 12–35 in *Cohort analysis*. 2455 Teller Road,?Thousand Oaks?California?91320?United States of America? SAGE Publications, Inc.

Bell, Andrew and Kelvyn Jones. 2017. “The Hierarchical Age-Period-Cohort Model: Why Does It Find the Results That It Finds?” *Quality & Quantity*.

Burgard, Sarah A. and Lucie Kalousova. 2015. “Effects of the Great Recession: Health and Well-Being.” *Annual Review of Sociology* 41(1):181–201.

Camm, Jeffrey D. 2016. *Essentials of Business Analytics*. Mason, OH: Cengage South western.

Fragoso, Tiago M. and Francisco Louzada Neto. 2015. “Bayesian Model Averaging: A Systematic Review and Conceptual Classification.” *arXiv preprint arXiv:1509.08864*.

Gelman, Andrew. 2014. *Bayesian Data Analysis*. Third edition. Boca Raton: CRC Press.

Jackson, Stephen HD, Martin R. Weale, and Robert A. Weale. 2003. “Biological Age—what Is It and Can It Be Measured?” *Archives of gerontology and geriatrics* 36(2):103–15.

Lam, Jack, Wen Fan, and Phyllis Moen. 2014. “Is Insecurity Worse for Well-Being in Turbulent Times? Mental Health in Context.” *Society and Mental Health* 4(1):55–73.

Leisering, Lutz. 2003. “Government and the Life Course.” Pp. 205–25 in *Handbook of the life course*, edited by Jeylan T. Mortimer and Michael J. Shanahan. New York: Kluwer Academic/Plenum Publishers,

Luo, Liying. 2013. “Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem.” *Demography* 50(6):1945–67.

Luo, Liying and James S. Hodges. 2016. “Block Constraints in Age–Period–Cohort Models with Unequal-Width Intervals.” *Sociological Methods & Research* 45(4):700–726.

Moen, Phyllis. 2014. “Constrained Choices: The Shifting Institutional Contexts of Aging and the Life Course.” Pp. 175–216 in *New directions in the sociology of aging*, edited by Linda J. Waite, Thomas J. Plewes, and National Research Council (U.S.). Washington, D.C: National Academies Press.

Raftery, Adrian E. 1995. “Bayesian Model Selection in Social Research.” *Sociological Methodology* 25:111–63.

Ryder, Norman B. 1965. “The Cohort as a Concept in the Study of Social Change.” *American Sociological Review* 30(6):843–61.

Vaisey, Stephen and Omar Lizardo. 2016. “Cultural Fragmentation or Acquired Dispositions? A New Approach to Accounting for Patterns of Cultural Change.” *Socius: Sociological Research for a Dynamic World* 2:2378023116669726.

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach*. 4th ed. Mason, OH: South Western, Cengage Learning.

Yang, Yang and Kenneth C. Land. 2013. *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*. Boca Raton, FL: CRC Press.