Electoral Studies 33 (2014) 28–40

Contents lists available at SciVerse ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

# Age-Period-Cohort analysis: A design-based approach

Elias Dinas [a,*], Laura Stoker [b]

[a] University of Nottingham, Nottingham, United Kingdom
[b] University of California, Berkeley, United Kingdom

ARTICLE INFO

*Article history:*
Received 16 May 2013
Accepted 24 June 2013

*Keywords:*
Age-Period-Cohort analysis
Identification problem
Cohort effects
Control group
Difference-in-differences

ABSTRACT

This paper develops a design-based approach to identifying cohort effects in APC analyses. Cohort effects arise when one cohort is treated by a unique set of formative socialization experiences, which causes it to differ from other cohorts in relevant outcomes. APC analyses typically compare treated and untreated cohorts from a single population. Our approach introduces a second group—a control group, in which no unit is treated but that is otherwise similar to the first—and adapts difference-in-differences estimation to the APC framework. The approach yields two identification strategies, each based on transparent and testable assumptions. We illustrate how the method works and what is to be gained through three examples.



## 1. Introduction

Studies in sociology, demography and political science have often used data from repeated cross-sections of individuals to estimate the effects of aging, period shocks, and early socialization experiences on attitudes or behaviors. The starting point for estimation is typically the Age–Period–Cohort (APC) model. In the APC model, outcomes can vary across individuals as they age (aging or life cycle effects), across time for all individuals (period effects), and across individuals depending upon the year of their birth (cohort effects). The well-documented problem with these models is that only two of these effects can be identified. Age (years since birth), period (year), and cohort (year of birth) are exact linear functions of each other: Age = Period–Cohort (introduction of this issue; Winship and Harding, 2008).

In order to estimate the relative contribution of age, period and cohort effects one must make one or more assumptions. The most common assumptions relate to the grouping of cohorts and the adoption of a polynomial function to model the effect of age or time. Analysts will group individuals born across adjacent years into cohorts (e.g., those born between 1965 and 1980 might be called Generation X) instead of working with annual birth cohorts. They will use continuous age and time variables, instead of dummy variables for each age or time point, and impose a functional form on the relationship between these variables and the outcome. Under such assumptions, the typical equation used to estimate these effects is of the following form:

$$Y_{it} = \alpha + b_1 Cohort_2 + b_2 Cohort_3 + \ldots + b_{k-1} Cohort_k + c_1 Time + c_2 Time^2 + d_1 Age + d_2 Age^2 + e_{it} \qquad (1)$$

In this setting, $Y_{it}$ represents the outcome for individual $i$ observed at time $t$ and $b_1$ through $b_{k-1}$ denote the average difference in the level of $Y_{it}$ between each new cohort and the oldest cohort, used as the reference category. This between-cohorts comparison is meaningful only if the cutpoints defining the cohort boundaries are defensible and the model accurately represents how period and aging effects operate.

This paper develops and illustrates an alternative approach to solving the APC identification problem. The approach introduces a control group to aid in the identification of cohort effects while also accounting for age

* Corresponding author.
*E-mail addresses:* elias.dinas@nottingham.ac.uk (E. Dinas), stoker@socrates.berkeley.edu (L. Stoker).

0261-3794/$ – see front matter 
http://dx.doi.org/10.1016/j.electstud.2013.06.006

and period effects. As such, it is primarily useful for readers interested in studying cohort effects within an APC framework. The approach yields two strategies for identifying cohort effects, each based on explicit and testable assumptions. Moreover, it identifies cohort effects without requiring that cohorts be grouped across adjacent birth years and without imposing assumptions about the functional forms of the age and period effects. In what follows, we first present an example to illustrate the logic of the approach. We then develop the core elements of the method. Our arguments build on prior work in sociology (Firebaugh and Chen, 1995), economics (Card and Krueger, 1994; Nielsen and Nielsen, 2012; Pischke, 2007), and statistics (Rosenbaum, 1987). Finally, we illustrate our arguments through three empirical examples.

## 2. Using a control group: an example

Over-time data on the fraction of children who repeat second grade, adapted from Pischke (2007), are depicted in Fig. 1. Until the 1960s, most German children started school in the spring. This changed in 1967, when the beginning of the school term was moved to the fall. This transition required two short school years, where time in school was compressed from 37 to 24 weeks. The upper curve of Fig. 1 presents the average level of grade repetition for all second grade cohorts from 1962 to 1973. There seems to be an upward spike in the average level of grade repetition among those cohorts who were affected by this policy change, those whose school year ended in 1967 or 1968. The 1969 cohort also seems to have slightly higher repetition rates than the preceding and following ones, even though they had not been directly affected by the policy. These differences, however, are more modest than those that differentiate the 1967 and 1968 cohorts from the posterior cohorts. The pattern is relatively flat from 1971 to 1973 (Pischke, 2007; Angrist and Pischke, 2009).

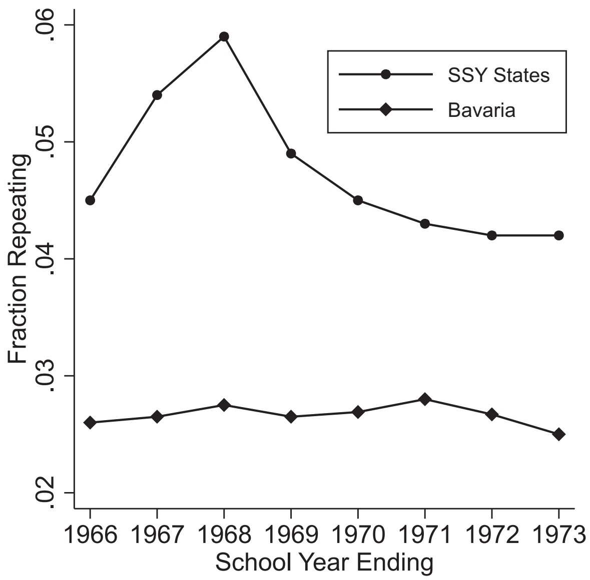


**Fig. 1.** Use of a Control Group in Pischke's (2007) Study of Retention Rates. *Note*: SSY stands for Short School Years.

In this case, we know that the treatment of a shortened school year was assigned to only two of the examined cohorts so there is no ambiguity about which cohorts should be distinctive if the treatment had an effect. Moreover, age is held constant by design. Still, however, estimates of the between-cohort differences rest on assumptions about how period effects are operating. Are they operating across younger cohorts and older cohorts to the same extent? Would we observe this gap even without this policy change?

To answer this question, Pischke (2007), leverages the fact that not all German schools shifted their school terms at the same time. Specifically, Bavarian schools started in the fall throughout the period. In the analysis, Pischke (2007) uses Bavaria as a control group, estimating the impact of the policy change through a within-cohort comparison. Rather than comparing the affected cohorts to those who entered the same schools earlier or later, he compares the affected cohorts with their same-school-year counterparts in schools where no policy change occurred. If the policy change caused an increase in grade repetition rates, this gap should be larger than the comparable gaps formed for cohorts entering school earlier or later. Crucially, period effects are controlled in the analysis to the extent that they operate on children in each set of schools to the same extent.

Although our examples come from a different theoretical background, the logic is very similar. Much is to be gained, we will argue, by adding a control group in the APC model. Doing so helps to test and justify the identification assumptions made in the analysis and yields new avenues for testing the robustness of the evidence for cohort effects.

## 3. Adding a control group in APC models: an augmented difference-in-differences approach

Most APC analyses look only at between-cohort differences within a targeted sample—a sample that contains a treated cohort as well as those born earlier and/or later. We add a control sample, which helps with the identification of cohort effects through a combination of both between- and within-cohorts comparison. The approach requires the identification of untreated and treated subjects within a given cohort. It supplements the traditional between-cohort comparison with a within-cohort comparison.

To develop the idea, it is useful to distinguish between a cohort, a generation, and a generation unit.[1] The defining property of a cohort is the year (or interval of years) of birth. For a cohort to be called a generation, important attributes must be common to its members and distinguish

[1] See Alwin and McCammon (2003) for an extended discussion of these and related concepts. Our abbreviated discussion here is designed to show how the use of a control group can be related to these ideas.