



Age–Period–Cohort Models for Electoral Survey Data: An Application to British General Elections, 1964–2010

Contributors: Gabriel Katz

Pub. Date: 2019

Access Date: January 4, 2019

Academic Level: Postgraduate

Publishing Company: SAGE Publications Ltd

City: London

Online ISBN: 9781526465818

DOI: <http://dx.doi.org/10.4135/9781526465818>

© 2019 SAGE Publications Ltd All Rights Reserved.

This PDF has been generated from SAGE Research Methods Cases.

Abstract

This case study underscores the importance of using age–period–cohort models to disentangle life cycle, generational, and election-specific effects when examining the determinants of long-term political change. I illustrate how these models can be used to study the impact on turnout of the decline in the competitiveness of British elections over the last 50 years, while controlling for other factors that may mask the relationship of interest. A key issue with age–period–cohort models is that standard statistical techniques and software packages are unable to handle relatively large data sets such as the one used in this study, with roughly 40,000 observations from 13 U.K. general elections. During the course of the project, I also had to make hard methodological choices to tackle some of the common challenges that arise when working with electoral surveys, such as turnout over-reporting, high proportion of missing values, and the difficulties in operationalizing relevant—dependent and independent—variables. A fundamental lesson from this case study is that careful consideration of alternative measurement strategies and estimation methods is crucial for the application of age–period–cohort models to repeated cross-sectional electoral surveys, as decisions in these areas can radically affect the substantive conclusions drawn from the data. At a more practical level, the study highlights that researchers may not always be able to rely on canned estimation routines to fit their models. Being able to tailor the estimation approach to the data at hand may thus be critical for the project's success.

Learning Outcomes

By the end of this case, students should be able to

- Comprehend the importance of using age–period–cohort models for repeated cross-sectional survey research
 - Understand the identification problem behind age–period–cohort models
 - Appreciate the benefits of using Bayesian inferential techniques for fitting age–period–cohort models and their advantages over alternative—for example, frequentist—estimation methods
 - Realize that canned routines available in commonly used statistical software packages are not always appropriate for applying age–period–cohort models to “big” political science data, and that sometimes coding their own estimators may be the only way to fit these models
 - Recognize the importance of accounting for turnout over-reporting and missing data when working with electoral surveys
-

Using Age–Period–Cohort Analysis to Understand the Process of Political Change

2019 SAGE Publications, Ltd

Researchers are sometimes interested in tracking the long-term dynamics of political behaviors or attitudes and in uncovering the driving forces behind their evolution. To do so, political scientists often take advantage of the increasing wealth of publicly available electoral surveys, sometimes—in industrialized democracies at least—spanning several decades.

An important methodological challenge faced by scholars attempting to identify the determinants of long-term political transformations from survey data is how to disentangle the differential impact of three time-related variables, namely,

- The changes in political behaviors or attitudes associated with differences in survey respondents' *age*—related to variations in maturity, psychological development, political experience, income, or status/position in life between individuals belonging to different age groups.
- The changes attributable to differences in the *period* in which the data were gathered, reflecting, for instance, exogenous shifts in political, economic, or social conditions over time.
- The changes brought about by transformations in the political, social, or cultural views adopted by different generations or *cohorts* of citizens. Individuals in the same cohort move through life together, encountering the same historical, political, and social events at similar ages. These events and formative experiences, which may shape the values and worldviews of these individuals, are different from those faced by people belonging to a different cohort. In the context of voter turnout, for instance, all British citizens become eligible to vote at the same age. However, the political and social conditions witnessed by the cohort of voters who came of age in the 1960s, which forged their political attitudes and behaviors, were quite different from those met by individuals who started voting in the 2000s.

The difficulty in separating these three temporal effects becomes apparent if we consider the situation of a researcher examining a single cross-sectional survey conducted at a given point in time. Age and cohort effects would be completely confounded in this data set: Every 20-year-old respondent in the sample belongs to the same cohort. And, of course, all the participants in this survey are observed at the same time. Hence, it would be impossible to discriminate the effects of age, period, and cohort. Furthermore, various other relevant macro-level or contextual factors could potentially influence the behaviors and attitudes of all the participants in this survey, but their effects would also be very difficult to distinguish from those of the time-related variables.

Age–period–cohort analysis is a statistical technique especially suited for the study of repeated cross-sectional surveys in which individual-level data from a representative sample of the population are observed over a number of years. By comparing data from age-specific observations recorded at different points in time and from different cohorts of participants (typically grouped into multi-year periods), this empirical approach breaks the perfect linear relationship between the three time-related variables and is thus able to identify the distinct influence of each of them. Moreover, by identifying and pinning down age, period, and cohort effects, this method also allows estimating the impact of other macro-determinants of political change, net of the influence of these temporal variables.

2019 SAGE Publications, Ltd

In the present case study, I review the main difficulties I faced in the application of age–period–cohort models to the analysis of British electoral survey data, detail the methodological choices I had to make to overcome these difficulties, and summarize the most important lessons learned during this process. In doing so, I also consider other, more general empirical dilemmas commonly faced by scholars analyzing electoral surveys and discuss the way in which I decided to deal with such problems in practice.

Project Overview: Studying the Long-Standing Effect of Electoral Competitiveness on Turnout in the United Kingdom

From a substantive perspective, the main goal of the project featured in this case study, co-authored with Professors Jack Vowles and Dan Stevens from the University of Exeter, was to evaluate whether and to what extent secular changes in the competitiveness of U.K. general elections affected voter turnout over the long run.

Political scientists have long noticed a downward trend in turnout in the United Kingdom and other advanced industrial democracies. In Britain, turnout rates have dropped from around 80% in the 1950s to slightly above 60% on average in the 21st century. The reasons for such decline are unclear and have been the subject of much academic debate, with different scholars proposing alternative explanations to account for this decay in political participation. Oliver Heath (2007), for instance, asserts that the drop in turnout in recent U.K. elections goes hand in hand with the erosion of party identification among younger cohorts of eligible voters, whereas Andre Blais and Daniel Rubenson (2013) emphasize the role played by the weakening sense of civic duty—that is, the belief that it is every citizen’s duty to vote—among younger generations.

A core point of the research described in this case study is that, along with these cultural and generational changes, there are also more “behavioral” mechanisms behind the drop in participation rates in the United Kingdom. Specifically, the key substantive argument of the project was that the protracted decline in the competitiveness of British elections over the last 50 years undermined citizens’ propensity to show up at the polls. This idea is not especially novel. In a comprehensive study of turnout in Western democracies, Mark Franklin (2004) had already made a similar argument: that citizens are more likely to participate in elections when they believe that elections “matter”—that is, when races are close, results are uncertain, and individuals have the feeling that their vote can actually affect the outcome.

Nevertheless, Franklin (2004) focused mostly on the short-term influence of electoral competitiveness on turnout. By contrast, Vowles, Stevens, and I contended that the competitive environment at the beginning of citizens’ political life had an enduring effect on turnout, because individuals acquire the habit of voting (or abstaining) during the first few elections in which they have the opportunity to participate. Therefore, people who entered the electorate in an environment of low competition and learned that voting “does not matter” during these formative years are more likely to abstain in subsequent elections than citizens who started voting around closely fought races. If the fall in competitiveness lasts long enough to socialize increasing

2019 SAGE Publications, Ltd

proportions of the electorate into low turnout habits, then it could have a cumulative negative effect on political participation.

Nonetheless, the empirical evidence for these claims has been largely inconclusive, with most research showing weak or negligible effects of competitiveness on turnout in U.K. elections. There are reasons to believe that these null findings are spurious, though, stemming from two basic flaws of prior work: (a) the use of incorrect measures of electoral closeness, which masked the downward trend in the competitiveness of British general elections over the past half century, and (b) the reliance on inappropriate inferential methods that failed to discriminate between age–period–cohort effects, on one hand, and the impact of protracted declines in competitiveness, on the other. Hence, designing better measures of electoral closeness and implementing better statistical approaches to estimate the effect of changes in the competitive environment on turnout were the primary methodological objectives of the research project.

Research Design: Data and Measurement Issues

The Benefits and Challenges of Working With Electoral Surveys

Data from cross-sectional surveys conducted by the British Election Study (BES) between 1964 and 2010 were used to gauge the relationship between electoral competitiveness and turnout. Although a simpler approach to study the link between competitiveness and turnout would have been to look at aggregate turnout rates, using survey data allows controlling for a rich set of covariates that might also impinge on political participation. For instance, more educated, older, and wealthier citizens are generally more likely to show up at the polls than other segments of the population. Similarly, as noted before, the political or cultural values (e.g., level of party identification, sense of civic duty) characteristic of different cohorts of citizens have also been posited to influence turnout. Failing to account for these variables would have rendered it more difficult to identify and isolate the effect of electoral closeness.

At the same time, working with survey data—rather than with official turnout figures—poses some methodological challenges. To begin with, a decision had to be made regarding the definition of the dependent variable. In every BES survey since 1964, participants have been asked to report whether they voted or not in the corresponding election. Would it be wise to take these responses at face value? Some participants in the sample may not remember whether they voted or not, while others might be embarrassed to admit that they failed to comply with what is widely believed to be a citizen's duty.

Because of these reasons, researchers administering the BES routinely check the responses to the turnout question against official electoral records. Participation rates computed from this validated vote are almost 10 percentage points lower on average than those emerging from the self-reported vote. The former are also much more closely aligned with the official turnout figures. Nevertheless, vote validation has only been conducted since 1987, so limiting the analysis to validated BES surveys would amount to removing the previous seven U.K. general elections from the sample. Moreover, this verification is only possible when

2019 SAGE Publications, Ltd

survey participants give their consent to having their information matched to the electoral registers.

Consequently, when using survey data to study voting—arguably one of the most essential political acts—researchers often face a trade-off. Should they rely on the self-reported vote, or should they instead examine only the validated vote? Although the latter measure is in all likelihood more reliable, it is typically only available for a subset of the survey participants, and thus has a more limited coverage than the former. As the researcher in charge of the empirical analysis for this project, I opted for a “compromise solution,” inspecting both the self-reported and validated vote and assessing the robustness of the results obtained from the two dependent variables.

A second issue to consider when working with electoral survey data has to do with the treatment of missing values (in the dependent and/or independent variables). The statistical literature has long established that simply discarding observations with missing values—a procedure called “list-wise deletion”—is bad research practice. However, it leads to the loss of potentially large amounts of information: In this particular case study, resorting to list-wise deletion would have led me to drop more than 10,000 observations from the sample. In addition, restricting attention to survey participants whose outcome and covariate values are completely observed can bias the results, as these individuals might differ systematically from other voters (e.g., they may be more likely to vote, more interested in politics).

The prevailing practice among political scientists relying on survey data is to impute the missing values, employing one of the many software packages available to this end (see the “Web Resources” section for an example). It is worth noting, however, that virtually all these programs assign values to the missing observations by assuming that the propensity for a data point to be missing is not related to the missing data itself. This is the “missing at random” (MAR) assumption. What does this mean in lay terms? Suppose that a BES participant failed to indicate whether she voted or not. When imputing her answer using standard software tools, we would be ruling out the possibility that the participant might have abstained and refused to respond to the turnout question precisely because she did not vote—for example, because she was embarrassed to admit that she did not show up at the polls. This is a strong assumption, and the researcher must conscientiously determine whether it is reasonable to impose it or not. For the purpose of this study, I decided to create two data sets, one resorting to list-wise deletion and another using multiple imputation for the outcome and explanatory variables, and conduct separate analyses on each data set. I could then assess the sensitivity of the findings to the MAR assumption by comparing the results from the two analyses.

Measuring Electoral Competitiveness: Alternatives and Justifications

Although BES surveys provided the dependent variable (turnout) and relevant individual controls (e.g., age, education, income) for the analysis, they did not offer a satisfactory measure for the key independent variable: electoral competitiveness. A crucial decision that had to be taken before proceeding to the quantitative estimation, then, was how to operationalize this fundamental concept.

2019 SAGE Publications, Ltd

Scholars seeking to estimate the connection between competitiveness and turnout drastically differ in their opinions about the way in which electoral closeness should be operationalized: whether based on actual results or pre-electoral polls, on seats or vote shares, on national- or constituency-level indicators, or even whether it should be measured in terms of objective results or subjective perceptions.

In planning the measurement strategy for the project, I carefully weighed each of these possible alternatives before settling on three indicators. At the national level, competitiveness was proxied by the difference in public support for the two leading parties in the last pre-electoral polls conducted during the campaign and by the seat-share gap between the two major parties in the previous general election. At the constituency level, electoral closeness was measured through the winning candidate's vote-share margin over the runner-up at the previous race. The reasons for these choices were the following:

- First, I drew upon the evidence from a vast literature indicating that British voters care about the election results in their constituency (which determine who will represent them in Parliament) as well as about the national electoral outcome (which determines the composition of the government).
- Second, voters cannot perfectly forecast the electoral outcome before ballots are cast. Hence, they must rely either on previous election results or on pre-electoral polls to try to anticipate whether their vote is going to “matter” or not. This seems almost a trivial point, but scholars claiming that the closeness of the race is a central criterion citizens take into account in deciding whether to show up at the polls or not often operationalize competitiveness in terms of the actual results of the election in question.
- Even when the last public opinion polls conducted during the campaign usually provide a reliable—yet imperfect—indication of the competitiveness of the upcoming general election, electoral results in Britain are predominantly reported in terms of parliamentary seats, rather than votes. This is because winning an election in the United Kingdom does not require obtaining the largest vote share but capturing a majority of the seats in the House of Commons. Therefore, the two measures—the difference in public support for the two leading parties in the last pre-electoral polls and the seat margin from the previous election—provide voters with valuable information about the expected closeness of the race. Consequently, rather than choosing between these two variables, I decided to include both of them as indicators of electoral closeness at the national level.
- At the constituency level, seat margins are obviously irrelevant. Nonetheless, both the local vote-share margins from the previous general election and the results anticipated based on constituency-specific pre-electoral polls seemed in principle valid measures of competitiveness. Unfortunately, pre-electoral polls were not available for all British constituencies over the 50 years covered in the study. This was not a major setback after all, because in those constituencies where the two sources of information are available, local vote returns from previous races have been shown to be more powerful predictors of participation than polls.

The rationale followed to operationalize electoral closeness illustrates that, when conducting empirical analyses, several alternative measures for the underlying constructs of interest are usually available. The

2019 SAGE Publications, Ltd

process of selecting the appropriate or best possible indicators must be guided by theoretical insights, by prior research, and—quite critically—by data constraints.

From the empirical perspective, the three chosen measures revealed that the average closeness of British elections dropped between 1964 and 2010, even when this decline was not constant. These indicators demonstrated that, when properly defined and operationalized, declining competitiveness was a persistent phenomenon in U.K. general elections over the sample period and not just a short-term factor.

Would other measures of electoral closeness have led to different conclusions? The answer is probably yes, but the same could be said about pretty much any indicator used in any quantitative political science analysis. As long as the researcher is transparent about the decisions taken to operationalize the relevant variables and is able to clearly justify why her chosen measures are preferable to other alternatives, this should not be a problem. Debates and disagreements about measurement issues are an intrinsic part of academic practice. It is ultimately up to the scholarly community—or its more immediate gate-keepers, such as journal editors and reviewers—to judge the relative merits, value, and contribution of different operationalization strategies.

Age–Period–Cohort Models in Action: Challenges and “Solutions”

Having decided on the operationalization of the independent variables of interest, it was time to move to the empirical analysis. As suggested in the introductory section, age–period–cohort models are particularly well suited for estimating the association between the closeness of elections and turnout without confounding the effect of competitiveness with

- The effect of age on turnout (lower on average for younger than for older respondents);
- The effect of cultural transformations and changes in values related to generational replacement;
- The effect of election-specific characteristics—for example, the quality of the competing candidates, the attractiveness of their policy platforms, the relative salience of the issues, and the economic environment—that could also help explain why some races exhibit higher participation rates than others.

Drawing on the approach to age–period–cohort analysis proposed by Yang and Land (2006), the age of each participant in the BES was regarded as an individual-level attribute (i.e., a trait or characteristic of each survey respondent), and each individual-level observation was taken as concomitantly belonging to two hierarchical or higher level contexts: election- (or survey-) year, and cohort, defined according to the first election in which a BES respondent became eligible to vote. This avoided treating the three temporal variables as additive individual-level covariates, creating instead a multi-level data structure in which respondents of different ages were simultaneously clustered within cohorts and election- (survey-) year. Doing so contributed—along with the use of multiple cross-sectional surveys and the definition of different time intervals for age, period, and cohort—to circumvent the identification problem behind age–period–cohort models.

The multi-level structure of the data set is illustrated in [Table 1](#), where each row in the table represents an

2019 SAGE Publications, Ltd

election and each column represents a cohort. The numbers in each cell give the sample sizes for each combination of cohort and election.

Table 1. Multi-level structure of the data.

Election	Cohort							
	Pre-1955	1955–1964	1964–1970	1970–1979	1979–1987	1987–1997	1997–2001	Post-2001
1964	1,510	259						
1966	1,453	332	89					
1970	1,229	266	255	92				
1974 (February)	1,493	380	319	269				
1974 (October)	1,360	377	324	304				
1979	953	302	313	294	31			
1983	1,723	574	551	721	439			
1987	1,467	535	549	674	596	75		
1992	1,049	473	495	632	544	341		
1997	922	445	424	613	673	537	1	
2001	681	420	350	473	544	449	118	
2005	699	539	507	661	735	714	170	136
2010	339	405	388	481	471	576	158	257

Nevertheless, two methodological problems remained to be addressed in order to be able to fit the age–period–cohort model to this particular data structure.

Issue 1: Too Few (Higher Level) Units

As seen in [Table 1](#), the data comprised about 40,000 individuals, each of them cross-classified into one of 13 possible elections and one of eight possible cohorts. Standard statistical practice prescribed a random effects multi-level formulation to account for potential correlation in the turnout decisions of individuals belonging to the same cohort and/or interviewed after a specific election.

Nonetheless, according to leading authorities in multi-level analysis such as Maas and Hox (2005), at least 30 and ideally more than 50 higher level units (i.e., elections and cohorts) are required to estimate a hierarchical model. The number of rows and columns in [Table 1](#) falls substantially short of this requirement. However, Maas and Hox (2005)—and, in fact, virtually all authors insisting on the need for a “minimum number” of clusters or higher level units for multi-level analysis—use frequentist (e.g., maximum likelihood) estimation methods, which rely on asymptotic or large sample approximations to obtain unbiased model parameter estimates. Bayesian inference, on the contrary, does not rest on the assumption that the number of clusters is “large” and can yield accurate parameter estimates even with very few higher level units. Hence, I decided to take advantage of this aspect of Bayesian methods and estimate the model via Markov chain Monte Carlo (MCMC) simulations to avoid statistical problems—and reviewers’ concerns—regarding the small number of cohorts and elections under study.

Issue 2: Too Many Observations

This posed a second methodological challenge, though. Bayesian inferential methods are known to be computationally intensive, so much so that most of the widely used software packages specializing in MCMC simulations, such as BUGS or JAGS (see the “Web Resources” for links to these packages), can only handle small sample sizes in a computationally efficient manner (i.e., without crashing, freezing, or running at an extremely slow pace). It was doubtful that these programs would be able to handle the almost 40,000 observations in my data set. I still attempted to fit a preliminary model in BUGS, but the estimation was indeed extremely slow—to the extent that the MCMC algorithm failed to converge even after letting it run for 2 weeks. And this was just a single trial run. For the purposes of the project, the same age–period–cohort model had to be estimated four times: using self-reported and validated vote as dependent variables and—for each of these outcomes—resorting to list-wise deletion as well as multiple imputation to deal with missing values.

These computational demands were so daunting that they could jeopardize the project even before any actual analysis had been conducted. To side-step this problem, I decided to code the estimation algorithm myself in R and C++, instead of relying on canned procedures. This was arguably not only the most challenging aspect of the research process but also one of the most intellectually stimulating parts. Like the vast majority of social scientists, I usually rely on readily available routines to fit my models, plugging the data and variables into the “black box” of standard software programs. Directly coding the routine to fit the age–period–cohort model called for—and allowed me to acquire—a much deeper understanding of the statistical method and of the computational aspects of the estimation approach.

2019 SAGE Publications, Ltd

To be absolutely frank, this was not a simple task—to say the least. It required deriving the mathematical formulae behind the algorithm, learning how to translate these formulae into computer code, having the patience to run the program, fix—multiple—errors in the code, and run it again (and again, and again). This rendered the estimation process difficult—and often very frustrating. Still, as intimidating as this may sound, writing the code, debugging it, and fitting all the necessary model specifications took less than 10 days in the end—so it definitely paid off!

The R program I developed and implemented in the published version of the paper—along with replication instructions—has been made publicly available and can be found following the link included in the “Web Resources” section; an analogous C++ routine is available upon request. Other valuable resources that can help interested readers go into the details of age–period–cohort analysis, become acquainted with Bayesian inferential methods for these kinds of models, and improve their statistical programming skills are listed in the “Further Readings” section.

Is Coding Your Own Estimator Really Necessary?

No, definitely not always. Even when my experience in this project has taught me that developing and coding estimation routines from scratch can help gain familiarity with the theoretical, statistical, and computational aspects of age–period–cohort models (and, I would argue, of nearly any other quantitative method), doing so takes considerable time and effort, as pointed out above. Such investment may not be worthwhile in view of the tight deadlines and competing pressures students and researchers are usually subject to.

Determining the most convenient estimation strategy for an empirical project will obviously depend on the characteristics of the data and of the statistical technique to be implemented. In the case of age–period–cohort analysis, a simple set of guideline rules I would recommend for choosing the appropriate estimation approach could be summarized as follows:

- When the number of cohorts and periods in the sample is large enough to justify the asymptotic approximations of frequentist or classic estimation methods, the researcher can safely resort to canned routines included in any commonly used statistical software package like SPSS[®] or Stata[®].
- In many—most—political science applications, however, scholars will have fewer than 50 (or 30, or 20) cross-sectional surveys at their disposal and will probably also care about a limited number of substantively meaningful cohorts. Under these conditions, the Bayesian paradigm offers a viable estimation approach to deal with the hierarchical structure of the data.
- If the sample size does not exceed a few thousand observations, specialized software tools for MCMC estimation (like BUGS or JAGS) provide a relatively efficient way of fitting age–period–cohort models.
- By contrast, in situations in which the researcher is studying only a few elections/cohorts but the sample size is relatively large, no readily available routine or Bayesian statistical package may be capable of fitting the model in a reasonable amount of time. In such circumstances, directly coding

2019 SAGE Publications, Ltd

the algorithm in more general programming languages like R, C++, or C is likely to be the best—even only—option.

A Summary of the Empirical Results

A detailed account of the findings reported in the published article emerging from the project is not the focus of this case study. That said, it is worth mentioning that the results confirmed the central argument that led to this investigation. The age–period–cohort model estimates revealed that the individuals who learned that elections “do not matter” in the first few occasions they could vote continued to be less inclined to show up at the polls throughout their adult life, and that the decline in the competitiveness of U.K. elections over the past half century had a durable and negative influence on turnout.

These conclusions held for the three measures of competitiveness employed and were robust to alternative definitions of the dependent variable and to different procedures used to deal with missing values. The inferences drawn from the empirical analysis therefore contradicted much previous work on the relationship between electoral closeness and turnout in Britain. This underscores how rigorous variable–operationalization and estimation strategies can challenge or even debunk conventional academic wisdom, contributing to enhance our substantive knowledge of political phenomena.

Conclusion

Age–period–cohort models enable researchers studying political phenomena from a long-term perspective to identify the determinants of observed transformations and trends beyond—or in addition to—those emanating from purely time-related factors. Although these models have seen fewer applications in political science than in other social sciences such as demography or sociology, the growing availability of long cross-sectional electoral surveys will probably render them increasingly popular in the near future. At the same time, such easy access to relatively large survey data sets poses complex measurement and estimation problems that must be taken into account in empirical work.

However, scholars must decide how to handle two recurrent problems that plague electoral surveys: turnout over-reporting and missing—outcome and covariate—values. Whereas vote validation increases the reliability of responses to the turnout question, validated data are difficult and expensive to collect, and it is typically only feasible to do so for a limited proportion of the sample. Hence, students of political participation have to acknowledge the trade-offs involved in relying on self-reported or validated vote, and attempt to maximize the accuracy of the survey responses while retaining as much—especially historical—data as possible. In the same line, the number of missing data points tends to increase with the overall sample size. Researchers should meticulously examine the validity of the assumptions underlying the most widely used procedures to deal with missing values and decide on their relative benefits and costs.

2019 SAGE Publications, Ltd

From the inferential standpoint, this plethora of survey data determines that off-the-shelf routines and software packages do not always provide researchers with the flexibility or computational efficiency they need. Hence, the estimation of age–period–cohort models may have to be tailored to the specific data set at hand. In particular, issues related to the number of cross-sectional surveys available for analysis, the number of substantively interesting cohorts, and the overall sample size must be thoroughly considered at the moment of selecting an estimation procedure.

This case study has attempted to highlight these problems and to review some—necessarily limited or partial—“solutions” I implemented to overcome them in practice. Hopefully, quantitatively oriented students and early career researchers working on political participation can take these suggestions as a useful starting point when thinking about their own work.

Exercises and Discussion Questions

1. This case study argues that controlling for the impact of life cycle, generational, and election-year effects is important when attempting to estimate the long-term influence of electoral closeness on turnout. What is the potential risk of failing to do so? What would have been the concerns of drawing conclusions from standard (pooled) turnout regression models that ignore age–period–cohort effects?
2. Would age–period–cohort models be necessary to disentangle the effects of the three time-related variables if I had used longitudinal (panel) electoral data instead of cross-sectional surveys?
3. In your opinion, is the MAR assumption reasonable for imputing missing individual-level turnout values? Is this assumption suitable when dealing with missing values in control variables like age, education, or income?
4. In the research described in this case study, I repeated the empirical analysis using list-wise deletion and multiple imputation to assess the sensitivity of the findings to the MAR assumption. Can you think of another way of going around the missing data problem without having to rely on this assumption?

Further Reading

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall/CRC Press.

Lynch, S. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York, NY: Springer.

Yang, Y., & Land, K. (2008). Age–period–cohort analysis of repeated cross-section surveys: Fixed or random

2019 SAGE Publications, Ltd

effects? *Sociological Methods & Research*, 36, 297–326. doi:<http://dx.doi.org/10.1177/0049124106292360>

Yang, Y., & Land, K. (2013). *Age-period-cohort analysis: New models, methods and empirical applications*. Boca Raton, FL: Chapman & Hall/CRC Press.

Web Resources

The BUGS Project homepage: <https://www.mrc-bsu.cam.ac.uk/software/bugs/>

JAGS homepage: <http://mcmc-jags.sourceforge.net/>

Amelia II: A Program for Missing Data: <https://gking.harvard.edu/amelia>

Vowles, J., Katz, G., & Stevens, D. (2016). Replication data for “Electoral competitiveness and turnout in British elections, 1964-2010.” *Harvard Dataverse*, V1. Retrieved from <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FAWCGP>

References

Blais, A., & Rubenson, D. (2013). The source of turnout decline: New values or new contexts? *Comparative Political Studies*, 46, 95–117. doi:<http://dx.doi.org/10.1177/0010414012453032>

Franklin, M. (2004). *Voter turnout and the dynamics of electoral competition in established democracies since 1945*. Cambridge, MA: Cambridge University Press.

Heath, O. (2007). Explaining turnout decline in Britain, 1964-2005: Party identification and the political context. *Political Behavior*, 29, 493–516. doi:<http://dx.doi.org/10.1007/s11109-007-9039-4>

Maas, C., & Hox, J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1, 86–92. doi:<http://dx.doi.org/10.1027/1614-1881.1.3.86>

Yang, Y., & Land, K. (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. doi:<http://dx.doi.org/10.1111/j.1467-9531.2006.00175.x>