

Modelling Ordinal Outcome Variables: If it's Worth Doing, it's Worth Doing Well*

Jack Bailey[†]& Matthew Barnfield[‡]

This version July 07, 2021.

Word count: X,XXX

Abstract

Ordinal data are an invaluable source of knowledge in political science. However, their enormous potential is not fully realised in most research, because of limitations in how they are analysed. In this paper, we hope to rectify this by providing a primer on how better to model ordinal outcome variables by using Bayesian methods to measure latent and observed properties of ordinal measures. Rather than repeating warnings raised elsewhere about the potential for inferential errors in common analytical approaches, we make a positive case for why modelling ordinal variables better is worthwhile, explaining visually and in plain language the logic and implementation of Bayesian ordinal regression models. We then apply our approach to use ordinal data to study political polarisation, through both a simulated example and a replication of a published empirical analysis. Our arguments and findings show that political scientists should continue to use ordinal variables, but learn more from them.

*Thank you to X, Y, and Z for their helpful comments.

[†]Research Associate, Department of Politics, University of Manchester, UK. If you have any comments or questions, feel free to contact me either by email (jack.bailey@manchester.ac.uk) or on Twitter (@PoliSciJack).

[‡]Postdoctoral Research Fellow, College of Social Sciences and International Studies, University of Exeter, UK

Introduction

Ordinal data are ubiquitous in political science. They serve an important purpose: to impose order on attitudes and beliefs that lack inherent scales. Where we hope to explain variation in political opinion, this is invaluable. Indeed, it is why we use ordinal scales to measure so many of our most important concepts. Consider, for example, voters' party identification (1 = "*Strong Republican*", 7 = "*Strong Democrat*"), their macro-economic perceptions (1 = "*Got a lot worse*", 5 = "*Got a lot better*"), and how often they pay attention to politics and government (1 = "*Always*", 4 = "*Some of the time*"). These, and other similar data, have been instrumental in improving our understanding of political behaviour in the decades since *The American Voter* (Campbell et al., 1960). Without this type of data we would likely be considerably worse-off.

Notwithstanding this widespread usage and rich contribution, ordinal data remain considerably under-exploited. Much applied research simply treats them as continuous. This is understandable insofar as this research deals with questions about how the expected value of some outcome of interest, such as government approval, changes on average. Some have pointed out, however, that this gives rise to type I and II errors, estimates with incorrect signs, and predicted outcomes that lie outside of the measurement scale (Lidell and Kruschke, 2018). Moreover, while many researchers use ordered probit or logit regression to check for such errors, typical implementations of these models face similar issues (Bürkner and Vuorre, 2019). With these important contributions in mind, we make a distinctly *positive* case for doing more with ordinal outcome data than simply treating them as continuous, or even using standard ordered logit/probit alternatives. We demonstrate that there is more to be learned from ordinal data than these analyses allow, and explain how to exploit this using a Bayesian approach.

In light of the recent "causal revolution" (Pearl and Mackenzie, 2018), this issue merits serious consideration. Causal claims come with stringent requirements. Given their potential to alter both how we understand politics and how we make real-world political decisions, this is appropriate. Yet, little advice exists when it comes to causal inference and ordinal outcomes. This is unfortunate, as these data are ubiquitous where research questions concern political

attitudes and beliefs. As a result, much causal research in political science takes great pains when it comes to causal identification, only to analyse the resulting data using methods that increase the chance of drawing false conclusions.

We begin by providing a brief overview of how ordinal outcome data are used in existing research and why this might be problematic. Then we make our more positive case for how Bayesian ordinal regression models are worth using instead, explaining their logic and implementation in the process. Finally, we present a clear case in which our approach could serve useful: political polarisation. After explaining how our approach better exploits ordinal data for questions about polarisation, we apply it to a simulated and a real replicated analysis on the effect of empathic concern on partisan polarisation (Simas, Clifford and Kirkland, 2019). The results of our replication demonstrate that ordinal regression models make it easier to distinguish between polarisation and other processes, and suggest that empathic concern does not appear to drive polarisation in quite the way the authors suggest.

References