How Much Can Be Inferred From Almost Nothing? A Maximum Entropy Approach to Fundamental Indeterminacy in Ecological Inference With an Application to District-Level Prediction of Split-Ticket Voting

Martin Elff*, Thomas Gschwend†, and Ron Johnston‡

Ecological inference aims at reconstructing data and data-generating processes at the individual level based on aggregate data. A typical application is the estimation of the probability of split-ticket voting on the individual level based on voting results on the level of electoral districts. Therefore, ecological inference has to rely on crucial assumptions about data generating processes at the level of the individual that cannot be tested by classical means of statistical hypothesis testing. This leads to a fundamental modelling indeterminacy: It is impossible to make sure that the statistical model of the data-generating process on which an ecological inference procedure relies is appropriate.

Focussing on the estimation of general $R \times C$ voting matrices of individual electoral districts, we address this problem of modelling indeterminacy. Our approach at solving this problem builds on the Principle of Maximum Entropy. We show how this principle can be used to determine models that exploit all information about the available aggregate data but are structurally independent from unavailable individual-level data. Further, we consider the apparent paradox that such entropy maximizing models imply very restrictive assumptions about the data generating process and explain how to reconcile the need of such restrictive assumptions with the abovementioned fundamental modelling uncertainty.

By way of extensive simulation studies conducted with R, we show that entropy maximizing models lead to unbiased predictions of district-level voting matrices. With these simulation studies we show further, how uncertainty about these predictions, to which fundamental modelling indeterminacy leads, can be assessed based on the Principle of Maximum Entropy. Finally, we apply this approach to the prediction of split-ticket voting at the level of electoral districts during the 1992 General Election of New Zealand.

*Dept. of Social Sciences, University of Mannheim (corresponding author)
†MZES, University of Mannheim
‡School of Geographical Sciences, University of Bristol