

A BAYESIAN APPROACH to the PROBLEM of ANTICIPATORY COVARIATES in THE ANALYSIS of RETROSPECTIVE DEMOGRAPHIC DATA *

**Gebrenergus Ghilagaber⁺, PhD, Associate Professor
Department of Statistics, Stockholm University**

Abstract

Some of the most important explanatory variables in retrospective demographic data, such as educational level and social class, are measured at the interview time but are used in investigations of behaviour in life segments that occur before the interview. An interesting question would then be as to what extent can the changes in the behaviour under study be attributed to changes (misclassification) in the distribution of individuals across the various levels of, say education; and to what extent do they reflect real differences in the behaviour of interest. The present study uses Bayesian approaches in order to utilize available information in defining appropriate priors and estimating posterior patterns of relative intensities of the event of interest across the levels of the anticipatory variable. The issues addressed are illustrated by estimating effects of education and social class on entry into and exit from family unions among Swedish men born in 1936-64.

Introduction

In retrospective demographic surveys, some important explanatory variables such as educational level and social class are measured at the interview time but are used in investigations of behaviour in life segments that occur before the interview.

For instance, educational level achieved by 2005 may be used to explain differentials in the intensities of entry into marital or non-marital unions that took place long time before the date of interview.

* Paper submitted for presentation at the 2006 Annual Meeting of the Population Association of America, Los Angeles, March 30 – April 1, 2006.

⁺ Address for correspondence: Department of Statistics, Stockholm University, SE-106 91 Stockholm – Sweden. E-mail: Gebre@stat.su.se

In other words, the explanatory variables mentioned above are anticipatory and show in many (possibly most) studies simply because the data-collection procedure focuses on, say birth or fertility histories, but contain no history on explanatory variables such educational progress and/or social class mobility.

Ignoring the anticipatory nature of such variables potentially produces biased estimates and, thus, incorrect conclusions with respect to the phenomena under study.

For instance, in the example above education-level was measured at the time of interview (2005) and was used as a fixed variable throughout the analysis despite the fact that most of the analysis is aimed at explaining behaviour which took place before the survey. Educational progress is likely to occur between the time of the event (say entry into marriage) and the survey date.

An interesting question would then be the following. To what extent can the changes in marriage-formation patterns across educational levels be attributed to changes in the distribution of young adults across the various levels of education; and to what extent do they reflect real differences in behaviour of men in different levels of education.

Unfortunately, apart from awareness of problem, our knowledge about the strength and direction of the biases in realistic situations is too scanty. The ambition of the current paper is to give this important, but ignored, problem the scientific attention it deserves.

We use Bayesian methods to define appropriate priors by assuming that some, if not all, of the respondents had a lower educational level at the time of family formation than what they had at the date of interview. An extreme case would then be to consider the worst scenario that the respondent had the lowest possible educational level at the time of family formation; re-estimate the model and compare the results. This is just an example of one possible configuration.

A clear picture of extent and direction of the bias under various combinations of configurations, is expected to lay the foundation for developing appropriate methods for correcting these biases.

The issues addressed are illustrated by analysis of family formation and dissolution among Swedish men born 1936-64. Results show that lack of proper account of anticipatory variables may seriously plague the results of the analyses.