Bayesian Variable Selection in Cost–Effectiveness Analysis

M.A. Negrín[1], F.J. Vázquez–Polo[1], E. Moreno[2] and F.J. Girón[3]

[1] Department of Quantitative Methods. University of Las Palmas de G.C., Spain.
[2] Department of Statistics and Operation Research. University of Granada, Spain.
[3] Department of Statistics and Operation Research. University of Málaga, Spain.

Acknowledgements: This research has been partially support by the grants SEJ–02814 (Junta de Andalucía, Spain), SEJ2007–65200 and SEJ2006–12685 (Ministerio de Educación y Ciencia, Spain).

Correspondence to: Miguel A. Negrín, Department of Quantitative Methods, University of Las Palmas de Gran Canaria, Faculty of Economics, E–35017 Las Palmas de Gran Canaria, Spain.
E-mail: mnegrin@dmc.ulpgc.es

Bayesian Variable Selection in Cost–Effectiveness Analysis

Abstract

Recently, linear regression models are being used to explain the cost and effectiveness of medical treatment. The covariates may include sociodemographic variables, such as age, gender or race; clinical variables, such as initial health status, years of treatment or the existence of concomitant illnesses; and a binary variable indicating the treatment received. However, most studies estimate only one model, called the full model, that includes all the covariates. This procedure ignores the uncertainty in model selection. In this paper we show four alternative Bayesian variable selection methods proposed in the literature. This analysis allows for the estimation of the probability of inclusion of each covariate in the real model followed by the data. Besides, variable selection can be useful for the estimation of the incremental effectiveness and incremental cost through Bayesian model averaging and for the sub–group analysis.

Key Words: *Variable selection, Bayesian analysis, cost–effectiveness, BIC, Intrinsic Bayes Factor, Fractional Bayes Factor, sub–group analysis.*

1 Introduction

Modelling risks, resource use, and outcomes of alternative medical treatments, in the econometric literature, are strongly based on the use of covariates in regression models applied to microdata (Kathleen *et al.*, 2002; Willan and O'Brien, 2001; O'Neill *et al.*, 2000; Healey *et al.*, 2000; Kronborg *et al.*, 2000; Russell and Sisk, 2000; among others). Recently, several papers have proposed the use of covariates for the comparison of technologies through a cost–effectiveness analysis (CEA). Hoch *et al.* (2002) were pioneers in this research demonstrating how the use of regression analysis could allow more precise estimates of treatment cost–effectiveness by adjusting for covariates. This paper uses net benefit data as the dependent variable. Willan *et al.* (2004) directly considered costs and effects jointly, assuming a bivariate normal distribution. Vázquez-Polo *et al.* (2005b) used an asymmetric framework where the costs are explained by the effects but the effects are not affected by cost. In a posterior work, Vázquez-Polo *et al.* (2005a) proposed a general framework where the effectiveness can be measured through a quantitative variable or a binary variable. Costs are also analyzed taking into account the presence of a high degree of skewness in the distribution. Nixon and Thompson (2005) developed Bayesian methods where the costs and the effects are considered jointly, and allow for the typically skewed distribution of cost data using Gamma distributions. Manca *et al.* (2005) also included covariates in a multilevel framework for multicentre studies.

One of the aims of regression models in CEA is to infer causal relationships between a dependent variable (cost or effectiveness) and the variable of interest (treatment). Other variables, known as control variables, are included to minimize the bias and uncertainty of the estimation when there are differences in the basal characteristics of the treatment groups, as usually occurs in observational studies. If the model is correctly specified, estimates of these coefficients are unbiased. However, the majority of the works in this area omit the analysis of selecting the variables having