APPLYING A DATA DUPLICATION TECHNIQUE IN LINEAR REGRESSION ANALYSIS OF WAITING TIME TO PREGNANCY

HALIMAH AWANG

Faculty of Economics and Administration, University of Malaya, Kuala Lumpur, Malaysia

Summary. This analysis demonstrates the application of a data duplication technique in linear regression with censored observations of the waiting time to third pregnancy ending in two outcome types, using data from Malaysia. The linear model not only confirmed the results obtained by the Cox proportional hazards model, but also identified two additional significant factors. The method provides a useful alternative when Cox proportionality assumption of the hazards is violated.

Introduction

Studies on the determinants of birth intervals and their effect on fertility, infant and child mortality and health have been well documented (Trussell et al., 1985; DaVanzo & Starbird, 1991; Miller et al., 1992; Curtis et al., 1993; Kuate Defo, 1997; Awang, 2003). Previous research on pregnancy interval lengths has concentrated on the study of the determinants of the time to pregnancy ending in only a live birth: that is, the analysis is of inter-live birth intervals. However, it is possible that a pregnancy could end in a stillbirth, miscarriage or an induced abortion. With multiple outcomes the analysis should take these competing risks into consideration.

Separate analyses could be carried out for each specific outcome type by treating other outcome types as censored data. This approach does not treat the different outcome types jointly, complicating the comparison of the effects of different outcome types. Another approach is to treat the multiple outcome types simultaneously in the same model, thus facilitating comparisons of their effects on the waiting time. Previous research in competing risk analysis used models that are not only complex in nature, but which also require advanced computational programmes and software that are not readily available (Larson & Dinse, 1985; Pandey & Suchindran, 1988; Kuk, 1992; Narendrathan & Stewart, 1993). In this paper, the different types of outcomes are accommodated in the analysis using a data duplication technique, as