Available online at www.sciencedirect.com
SCIENCE DIRECT®
Mathematical Biosciences 196 (2005) 226–237
Mathematical Biosciences
www.elsevier.com/locate/mbs

Modelling bivariate count series with excess zeros

Andy H. Lee [a,*], Kui Wang [a], Kelvin K.W. Yau [b], Philip J.W. Carrivick [a,c], Mark R. Stevenson [d]

[a] *Department of Epidemiology and Biostatistics, School of Public Health, Curtin University of Technology, GPO Box U 1987, Perth, WA 6845, Australia*
[b] *Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong*
[c] *Occupational Health Service, University of Oxford, Oxford, United Kingdom*
[d] *Injury Prevention and Trauma Care Division, The George Institute for International Health, The University of Sydney, Sydney, NSW, Australia*

Accepted 20 May 2005
Available online 18 July 2005

Abstract

Bivariate time series of counts with excess zeros relative to the Poisson process are common in many bioscience applications. Failure to account for the extra zeros in the analysis may result in biased parameter estimates and misleading inferences. A class of bivariate zero-inflated Poisson autoregression models is presented to accommodate the zero-inflation and the inherent serial dependency between successive observations. An autoregressive correlation structure is assumed in the random component of the compound regression model. Parameter estimation is achieved via an EM algorithm, by maximizing an appropriate log-likelihood function to obtain residual maximum likelihood estimates. The proposed method is applied to analyze a bivariate series from an occupational health study, in which the zero-inflated injury count events are classified as either musculoskeletal or non-musculoskeletal in nature. The approach enables the evaluation of the effectiveness of a participatory ergonomics intervention at the population level, in terms of reducing the overall incidence of lost-time injury and a simultaneous decline in the two mean injury rates.

* Corresponding author. Tel.: +61 8 9266 4180; fax: +61 8 9266 2958.
E-mail address: andy.lee@curtin.edu.au (A.H. Lee).

0025-5564/$ - see front matter © 2005 Published by Elsevier Inc.
doi:10.1016/j.mbs.2005.05.001

Keywords: Autoregression; Bivariate Poisson; EM algorithm; Random effects; Zero-inflated Poisson model; Zero-inflation

1. Introduction

Time series of count events containing many zeros are frequently encountered in the biosciences. A motivating example arises from occupational health, where population injury data aggregated over time are observed [1]. It is plausible that certain individuals, such as experienced workers, are at low risk of injury (contributing to the extra zeros), while the other sub-population consists of members who are susceptible and may incur injury more often. In such applications, the discrete time series of counts contain excess zeros relative to the Poisson distribution, which may lead to spurious associations and misleading inferences if the zero-inflation problem is ignored [2].

For count data of independence events, the extra-Poisson variation or overdispersion induced by the zero-inflation can be handled using a compound probability model such as the zero-inflated Poisson (ZIP), in which the response count variable is modeled as a mixture of a Poisson distribution and a degenerate component with point mass at zero [3]. Böhning [4] reviewed the related literature and provided a variety of examples from the biosciences. The ZIP model incorporates more zeros than those allowed by the Poisson. A plausible interpretation is in terms of its (unobserved) two-point heterogeneity, corresponding to the 'perfect' and 'imperfect' states of the underlying process [5].

In many situations, the zero-inflated outcomes are classified in a binary manner (Y_1, Y_2) according to a feature of interest. Analogous to the univariate situation, a bivariate Poisson distribution cannot account for the relatively few events and thus extra zeros in the data [6]. A bivariate zero-inflated Poisson (BZIP) model was recently proposed for analyzing two types of occupational injury that simultaneously occur but are rarely observed [7]. In this study, the BZIP methodology for individual data is generalized to model a temporal sequence of bivariate discrete counts. After briefly reviewing the BZIP approach in the next section, a class of BZIP autoregression models with random effects is presented in Section 3 to accommodate zero-inflation and the inherent serial correlation between successive observations on the dual outcomes. The technique is applied in Section 4 to evaluate the effectiveness of a participatory ergonomics intervention for reducing musculoskeletal and non-musculoskeletal workplace injuries sustained by a group of hospital cleaners over a seven-year period. Finally, some concluding remarks and extensions of the method are given in Section 5.

2. Bivariate zero-inflated Poisson model

For bivariate count data with excess zeros, a bivariate zero-inflated Poisson (BZIP) model can be formulated as a mixture of bivariate Poisson distribution and a point mass at $(0,0)$ [7]. Consider a BZIP distribution for bivariate events (Y_1, Y_2)