



Actuarial Studies Seminar
Macquarie University
4 November 2009

Dr. David Pitt
Centre for Actuarial Studies, The University of Melbourne

Model Selection and Claim Frequency for Workers' Compensation Insurance

We consider a set of workers' compensation insurance claim data where the aggregate number of losses (claims) reported to insurers are classified by year of occurrence of the event causing loss, the US state in which the loss event occurred and the occupation class of the insured workers to which the loss count relates. An exposure measure, equal to the total payroll of observed workers in each three-way classification, is also included in the dataset. Data are analysed across ten different states, 24 different occupation classes and seven separate observation years. A multiple linear regression model for these data, with only predictors for main effects, could be estimated in more than 17 billion ways! In addition, one might expect that the number of claims recorded in each year in the same state and relating to the same occupation class, to be positively correlated. Different modelling assumptions as to the nature of this correlation should also be considered. On the other hand, it may reasonably be assumed that the number of losses reported from different states and from different occupation classes are independent. Our data can therefore be modelled using the statistical techniques applicable to panel data and we work with generalised estimating equations (GEE) in the paper. For model selection, Pan (2001) suggested the use of an alternative to the AIC, namely the quasi-likelihood under independence model criterion (QIC), for model comparison. This paper develops and applies a Gibbs sampling algorithm for efficiently locating, out of the more than 17 billion possible models that could be considered for the analysis, that model with the optimal (least) QIC value. The technique is illustrated using both a simulation study and using workers' compensation insurance claim data.