

# Two-part Regression Models for longitudinal zero-inflated count data

Marco Alfò

Dipartimento di Statistica, Probabilità e Statistiche Applicate  
“Sapienza” Università di Roma

P.le Aldo Moro, 5 00185 Rome - ITALY

Tel: +390649910672. e-mail: [marco.alfò@uniroma1.it](mailto:marco.alfò@uniroma1.it)

Antonello Maruotti

Dipartimento di Istituzioni Pubbliche, Economia e Società  
Università di Roma Tre

## Abstract

Two-part models are quite well established in the economic literature, since they resemble accurately a principal-agent type model, where homogeneous, observable, counted outcomes are subject to a (prior, exogenous) selection. The first decision can be represented by a binary choice model, modeled using a probit or logit link; the second can be analyzed through a truncated discrete distribution such as truncated Poisson, negative binomial and so on. Only recently, a particular attention has been devoted to the extension of two-part models to handle longitudinal data. We discuss a semi-parametric estimation method for dynamic two-part models and propose a comparison with other, well established, parametric and semiparametric alternatives. Heterogeneity sources that influence the first level decision process, i.e. the decision to take a certain decision, are assumed to influence also the (truncated) distribution of the *positive* outcomes. Estimation is carried out through an EM algorithm without parametric assumptions. Furthermore, we investigate the extension of the finite mixture representation to allow for transition between the two parts of the model (an observable transition) as well as between two components in each of these parts (i.e. an unobservable transition). The proposed models are discussed using empirical as well as simulated data.

*Keywords* Overdispersed Counts, Hurdle Models, Nonparametric Maximum Likelihood, Random Effects, Markov models.