Industrial location and interpretation of zero counts

Daniel Liviano · Josep-Maria Arauzo-Carod

1 Motivation

A common characteristic of data on new firm locations is the existence of a high percentage of zeros, that is, a high number of potential sites where no firms decide to locate. This phenomenon has been, in some ways, solved by the most popular...
econometric framework used in recent location analysis (Count Data models, henceforth CD models), but how important is the existence of a high percentage of zeros in such models? How can it be dealt with, from both a methodological and an economic point of view? In theory, CD models, such as Poisson (P) or negative binomial models, allow the existence of zero counts. However, once the proportion of zero counts becomes larger, these models cease to be efficient (Cameron and Trivedi 1998). A common solution has been to use zero-inflated models, such as the zero-inflated Poisson and the zero-inflated negative binomial models (Manjón-Antolín and Arauzo-Carod 2011; Arauzo-Carod 2008; Tadesse and Ryan 2004; Gabe 2003; Tomlin 2000) and the zero-altered Poisson models (Basile 2004); but another solution for dealing with an excess of zeros is to estimate hurdle models, which assume that the set of covariates and their coefficients might differ for two different events, these being the presence of entries or not and the number of entries, provided that this number is positive.

Specifically, industrial location data suffer from two main problems. On the one hand, there can be overdispersion, which occurs whenever the variance in the data exceeds the variance assumed by the model and which may result from neglected or unobserved heterogeneity that is inadequately captured by the covariates in the conditional mean function. The consequence of this is a loss of efficiency and the invalidation of posterior inference. On the other hand, although a model such as the Poisson model can deal with situations in which there is a high number of zero value observations, certain problems arise when this number is excessive and exceeds the number of zero counts expected by the model. This may be caused by (a) the existence of unobserved heterogeneity and (b) a selectivity problem, which occurs when the observed outcomes are produced by two latent processes; that is, a count data process and a selection process that are generally independent from each other. From a methodological point of view, the excess of zeros problem has been dealt with by estimating mixture models such as the negative binomial (NB), zero-inflated and hurdle models. However, are all these models equally appropriate whenever there is a high number of zero counts? Previous methodological problems imply that contributions that not properly take them into account and not try to tackle them might suffer from important econometric inconsistencies. Therefore, we aim to contribute to existent literature of empirical industrial location by providing a technical solution in order to deal with a very common problem related to this type of data, which is the high number of areas that are not being chosen by entering firms. In this sense, while first CD models applied to location analysis were mainly the simplest versions (e.g. Poisson models), later scholars have introduced in a gradual way better tools like zero-inflated Poisson and the zero-inflated negative binomial models, but even if acknowledging that this is a step forward, the problem is still far away to be solved, so this is the reason for this paper, to provide alternative specifications that (potentially) fit better with typical data of location decisions of new plants.

Let us consider the case under study in this paper: the location of manufacturing plants in Catalonia.\(^1\) If one takes a look at Catalan municipalities, one will find highly heterogeneous types of geographical areas. Some of these are large urbanised

\(^1\) Catalonia is an autonomous region located in north-eastern Spain and has Barcelona as its capital. It has about 7 million inhabitants (15 % of the Spanish population) and contributes 19 % of Spanish GDP.