1 Abstract

The main aim of this work is to propose a multilevel approach for repeated cross-sectional data. Multilevel data (Goldstein, 2010; Skrondal and Rabe-Hesketh, 2004) consist of units of analysis of different type that are hierarchically nested. At the lowest level such units can be described by some variables, and they are also grouped into larger units, which, in turn, could be described by other variables. The general specification of multilevel models allows a large variety of applications. In particular, repeated measures data can be seen as a specific case of multilevel data with occasions $i$ at level-1 and units $j$ at level 2 (Maas and Snijders, 2003). The dependence among level-1 errors that characterizes panel data can be handled by including correlation structures at level-1 (Goldstein, 2010). Moreover, it is possible to allow heteroscedastic within-group errors through variance functions (Davidian and Giltinan, 1995). This flexibility in the specification of covariance structures represents an important feature of mixed-effect models for longitudinal data.

Differently from longitudinal data, repeated cross-sectional data consist of observations on individual survey respondents drawn from the same context (e.g. the same country) at many different time-points, and can therefore be treated as clustered within time-points (Firebaugh, 1997). Being a new sample each time point, this collection of data does not allow to follow specific individuals over time but allow to catch social changes. DiPrete and Grusky (1990) were the first to adopt a multilevel framework for repeated cross-sectional data. The key difference of this model with the traditional multilevel framework is the possibility of modelling the time effects by allowing for serial correlation among level-2 units (time-points). The authors took into account this case by deriving generalized least-square estimators. A similar idea has been considered by Browne and Goldstein (2010) in a context of spatial correlation in a Bayesian framework. The independence assumption among level-2 disturbances is relaxed and the spatial correlation between pairs of clusters is modelled through an explicit function of the distance between them. However, to our knowledge, the analysis of repeated cross sectional data in a multilevel framework is not well established both from the theoretical and the practical point of view. In fact, most specifications are ad hoc solutions with no available software so that such models result poorly developed and rarely applied.
In this work we propose a multilevel framework similar to that by DiPrete and Grusky (1990) with individuals as level-1 units and time-points as level-2 units. The time dependence is modelled at the second level by means of an autoregressive structure between random effects. Rather than GLS estimators, we derive and implement full maximum likelihood estimators with known desirable properties. We obtain model estimates through an EM iterative algorithm and derive robust standard errors by means of a bootstrap scheme.

The work has been motivated by the analysis of the first world database of Tribal art prices. Such database, built by a team of researchers of the University of Bologna (Rimini campus), in conjunction with other institutions, contains information on over 20000 artwork items sold by the most important auction houses from 1998 to 2011. Auction data do not constitute a proper panel so that the classical multilevel approach for longitudinal data cannot be applied. Indeed, they have a structure like that of repeated cross-sectional surveys in that different artwork items are sold in different time-points. Hence, we apply the proposed multilevel model on Tribal art prices. This new specification turns out to be a natural and more convenient choice over the traditional hedonic regression model (Rosen, 1974) for modelling artwork prices. The overall result is a parsimonious yet powerful specification that can also reveal a useful tool to forecast the future values of prices.

References


