Latent Markov multinomial logit regression for discrete choice data: implications for willingness to pay for alternative urban freight policies

Valerio Gatta a, Francesco Lagona b, Edoardo Marcucci c

Extended abstract

Objectives and motivation
Discrete choice models are commonly used in behavioural freight modelling to estimate willingness to pay (WTP) for specific attributes such as number of loading and unloading bays, probability of finding them free, etc. (e.g. Gatta and Marcucci, 2014; Marcucci et al., 2015). While choice modelling develops, it also makes available a growing set of models with progressively higher degrees of behavioural richness. This can be testified and illustrated by the progression from multinomial logit, nested logit, cross-correlated nested logit and mixed logit. At the same time the request for more realism, greater behavioural relatedness, higher parameter accuracy and credibility is not over and analysts are exploring alternative methodological approaches to parameter estimates.

More precisely, multinomial logit regression provides a general framework for the statistical analysis of discrete choice data. Under this setting, the influence of the covariates on the outcomes is assumed constant across subjects and along time. The available covariates, however, often explain only a portion of outcome variability. As a result, subject-specific random effects are typically used to account for unobserved heterogeneity across individuals. Mixed-effects multinomial logit models, including both fixed and random effects, can account for this source of latent heterogeneity. However, under this setting, the influence of the covariates is only allowed to vary across subjects. A limitation of this approach is that both fixed and random effects are assumed time-constant. This assumption can be too restrictive in specific situations, especially when dealing with stated preference data when repeated choices are made by the same subject. A second limitation pertaining to mixed-effects multinomial logit models relates to the need to specify a priori the parametric distribution of the random effects. This last point might have strong implications for parameters estimates since a mis-specified random effects distribution can seriously bias the whole analysis.

General description
The paper adopts a hidden Markov approach to extend the multinomial logit specification so to overcome the limitations discussed above. Precisely, the regression coefficients of a multinomial logit model is assumed to vary across subjects and along time, according to a hidden Markov chain that emulates the dynamic effects of latent variables on the repeated choices behaviour of the subject. This allows for the estimation of the time-varying effects of the covariates without assuming a specific parametric distribution of the random effects. The paper also describes an efficient expectation-maximization algorithm to obtain maximum likelihood estimates and a computationally inexpensive routine to calculate bootstrap standard errors. Methods of model selection and outcome prediction are finally discussed.

Results and conclusions
The paper compares the results obtained under the proposed model with those found by a standard multinomial logit regression with random effects. This study demonstrates that the latent Markov approach flexibly accommodates for time-varying latent heterogeneity, avoiding possible specification errors due to erroneous distributional assumptions on the random effects. These results are particularly important for behavioural research in urban freight transportation. In these studies, questionnaires with a substantial number of choice exercises are popular, due to the lack of participation and high cost of each interview. It is common practice to administer a considerable number of choice exercises to each interviewee. Given this context it seems not only appropriate but also needed to test for time varying effects of latent variables that have a longer time for manifesting themselves. The paper will help understand the behavioural implications of the questionnaire administration procedures adopted and unveil the likely biased results ascribable to the commonly adopted data treatment while discussing their possible implications for the willingness to pay for alternative urban freight policies.

References

**Keywords:** dynamic behaviour; hidden Markov model; time-varying heterogeneity; urban freight distribution; stated preferences.