

Multiple choice models: why not the same answer? A comparison among LIMDEP, R, SAS and Stata

Giuseppe Bruno ¹

1. Bank of Italy Economic Research and International relations. giuseppe.bruno@bancaditalia.it

Keywords: Multinomial models, Probit, Gibbs sampler, MCMC.

The recent past has seen a great deal of research into categorical response models especially in the econometrics and behavioral literature. The most popular categorical response models are the Multinomial Logit (**MNL**) and the Multinomial Probit (**MNP**). The choice space for these models can be dichotomous (yes or no) or polytomous (e.g. points on a Likert scale). When dealing with more than two choices, a trade-off comes out between numerical tractability and model flexibility. In particular the more simple mathematical features of the **MNL** model entail the validity of the Independence of Irrelevant Alternatives (**IIA**) assumption. On the other hand, the adoption of the **MNP** model provides the maximum modeling flexibility at a much higher computational price. In this paper we investigate the availability and the accuracy of canned estimation algorithms for the multinomial models in the most spread packages employed by the researchers at the Bank of Italy. Here we compare the classical Maximum Likelihood method along with the Markov Chain Monte Carlo (**MCMC**) method available in the R package named **MNP**. A thorough comparison of the algorithms available in the packages **LIMDEP**, **R**, **SAS** and **Stata** is provided along with accuracy and performances indications. Preliminary results are quite insightful. As it is foreseeable, the availability of a closed-form expression for the choice probabilities in the multinomial logit model, shown in the following equation: $Pr(y_i = j|X_i) = \frac{\exp(X_i \cdot \beta_j + C_{ij} \lambda)}{\sum_{k=1}^J \exp(X_i \cdot \beta_k + C_{ik} \lambda)}$, allows the employment of stable and replicable algorithms. On the other hand the estimation of a multinomial probit with Maximum Likelihood involves the integration of a multivariate normal distribution. This task is not numerically straightforward. Different packages provides different answers and sometimes they don't even provide an answer. R provides the **MNP** package for fitting a Bayesian multinomial probit model with Markov chain Monte Carlo methods (see Imai and van Dyk 2005) . **SAS**, **STATA** and **LIMDEP** fit the multinomial probit model via Maximum Likelihood. Some empirical applications taken from the health insurance and the travel mode choices are presented. It is shown how the employment of the estimation algorithms for these models would benefit from the comparison with the implementation of the same algorithm in different packages and how helpful the comparison with already available numerical benchmarks would be (see Econometric Benchmarks 2010).

References

Econometric Benchmarks (2010).

<http://www.stanford.edu/~clint/bench/#logit>.

G. Glasgow and R.M. Alvarez (2008). Discrete Choice Methods,

http://www.polsci.ucsb.edu/faculty/glasgow/glasgow_alvarez_final.pdf.

- Melvyin Weeks (2003). The Multinomial Probit Model Revisited: A Discussion of Parameter Estimability, Identification and specification Testing. *Journal of Economic Surveys*, 11, 297–319.
- K. Imai and D. van Dyk (2005). MNP: R package for fitting the Multinomial Probit Model. *Journal of Statistical Software* Vol. 14 issue 3.
- H.M. Park (2009). Regression Models for Ordinal and Nominal Dependent Variables Using SAS, Stata, LIMDEP, SPSS.
www.indiana.edu/~statmath/stat/all/cdvm/cdvm_nominal.pdf.