ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 23 МАТЕМАТИКИ Выпуск 2
2016

Yu. S. Khokhlov, **O. I. Sidorova**, **I. V. Zakharova** (Moscow, Lomonosov Moscow State University; Tver, Tver State University). **New multivariate discrete choice model.**

In many choice cases, one can choose between two or more categories. For example, households usually can choose between many brands within a product category. In our report we deal with quantitative models for such discrete choices, where the number of choice options is more than two. The models assume that there is no ordering in these options, based on, say, perceived quality.

We have the alternatives $1, 2, \ldots, m$ and need to choose one of them. To do this we use the utility functions. Let $u_k = \mu_k + \varepsilon_k$ be the utility to choose the alternative k, where μ_k is nonstochastic function of explanatory variables and unknown parameters and ε_k is an unobservable random variable.

It is assumed that the individual chooses the alternative for which the associated utility is highest. McFadden proved that the polynomial logit model is derived from utility maximization if and only if ε_k are independent and identically distributed (i.i.d.) and have the distribution function $F(y) = e^{-e^{-y}}$, $y \in \mathbf{R}^1$ (Type I extreme-value distribution). If Z is the result of the choice then

$$\mathbf{P}\{Z = k\} = \frac{e^{\mu_k}}{e^{\mu_1} + e^{\mu_2} + \cdots + e^{\mu_m}}.$$

This model is very simple, good for estimation and interpretation, but the choice between two categories depends only on the characteristics of categories under consideration.Last property of the model is known as *independence of irrelevant alternatives* (IIA). It is not good from economic point of view (see example in [2, p. 86]). IIA is the consequence of the fact that the errors ε_k are i.i.d.

Let I be the set of multi-indices $\mathbf{i} = (i_1, i_2, \ldots, i_m)$, $i_k = 0 \vee 1$. For every $\mathbf{i}$ define random variables $u_{\mathbf{i}} = \mu_{\mathbf{i}} + \varepsilon_{\mathbf{i}}$, where $\mu_{\mathbf{i}}$ is nonstochastic and random variables $\varepsilon_{\mathbf{i}}$ are i. i. d. and have extreme-value distribution. Now define $\widetilde{u}_k = \max_{\mathbf{i}:\, i_k = 1} u_{\mathbf{i}}$. Again we choose the alternative for which the associated utility is highest.

The main result of our contribution is the following one.

Theorem. *Under the above conditions we have*

$$\mathbf{P}\{Z = k\} = \Bigl(\sum_{\mathbf{i}:\, i_k = 1} e^{\mu_i}\Bigr) \Big/ \Bigl(\sum_i e^{\mu_i}\Bigr).$$

R e m a r k 1. Our model is simple enough, good for estimation and interpretation and have no IIA property.

R e m a r k 2. If we consider some subset I_0 of indices we get new model. For example, for $I_0 = \{(1, 0, \ldots, 0), (0, 1, \ldots, 0), (0, 0, \ldots, 1)\}$ we come back to classical polynomial logit model.

R e m a r k 3. Multivariate distribution of the errors in our model is a special max-stable distribution with dependent marginals. If we take another max-stable distribution we get new model.

Acknowledgements. This research is supported by Russian Scientific Foundation, project 14-11-00364.

REFERENCES

1. *McFadden D.* Conditional Logit Analysis of Qualitative Choice Behavior. — In: Frontiers in Econometrics. / Ed. by P. Zaremba. N. Y.: Academic Press, 1974, p. 105–142.
2. *Frances Ph. Y.*, *Paap R.* Quantitative Models in Marketing Research. Cambridge: Cambridge Univ. Press, 2001, 206 p.