A Comparative Empirical Study of Discrete Choice Models in Operational Contexts

Demand estimation is a fundamental task in retail operations and revenue management, providing the necessary input data for inventory control and price optimization models. The task is particularly difficult in operational contexts when product availability varies over time and customers may substitute. In fact, empirical studies of different industries show that stockout-based substitution is a common occurrence. For airline passengers, recapture rates are acknowledged to be in the range of 15%–55% (e.g., Ja et al. 2001), while Gruen et al. (2002) report recapture rates of 45% across 8 categories at retailers worldwide.

When the first choice of a customer is not available, demand could be lost or recaptured (i.e., shifted to another product). Unfortunately, this phenomenon is not directly observable from sales transactions, and various statistical techniques have been proposed to correct for them. Moreover, the instantaneous availability of data from online platforms reflecting consumers’ preferences in real time also imply the need for efficient computational procedures to frequently update the estimates of customer preferences.

In the last decade or so, several proposals have emerged both in the modeling and computational sides to capture choice-based demand in operational contexts. In addition to the classical Multinomial Logit (MNL) model (Luce, 1965) and its variants (e.g., Nested Logit, latent class MNL), new demand models have been proposed (e.g., the Markov Chain model (Blanchet et al., 2016) or revisited (e.g., the rank-based and exponomial models). At the same time, new computational approaches were developed to ease the estimation function.
In general, choice-based demand proposals in the recent and prolific OM-related literature are benchmarked versus one or two alternatives, typically the MNL and the independent demand models (e.g., Alptekinoglu and Semple (2016)), Simsek and Topaloglu (2016)).

In this paper, we conduct a systematic, empirical study of different demand models, including the single class MNL, latent class MNL, exponomial, Markov chain, independent demand, rank-based, and discrete uniform. Through an exhaustive set of numerical experiments on both synthetic and real data, we provide comparative statistics of the quality of the different choice models and estimation methods, and characterize operational environments more suitable for particular model/estimation implementations according to the characteristics of the input data. Our results show that the relative out-of-sample prediction performance obtained by the different discrete choice models strongly depend on the in-sample data size, the allocated time available for parameter estimation, and on the complexity of the true consumer preferences.

This work also evaluates alternative state-of-the-art algorithms spanning traditional nonlinear optimization methods to maximize the likelihood functions, expectation-maximization (EM) proposals, and column generation approaches.

References

