Improving the Estimation of Random Coefficient Logit Models of Demand

Marietta J Tretter
Dept. Information and Operations Management
Texas A&M University
College Station, TX 77843-4217

Abstract

Ordinary Logistic Regression is well used in Data Mining when the target variable is binary. However it is not possible to use such existing Data mining tools when one is interested in modeling demand over time from very large data bases with many different products, retailers, and prices. To mine such data it is necessary to fit random coefficient logit models of demand. Fitting these models is complex and often not very precise. Model fitting involves optimization and simulation. This research will present an approach using inexact arithmetic, also known as interval arithmetic, to make the estimation process more efficient and to produce interval error bound on the estimates. Applications of the technique will also be presented.

Introduction

Interest in the method of random coefficient logit models began with the desire to study demand in large data sets over time. In general, Data Mining with time series or time ordered data is still in the exploratory stage. There are practically non existent data mining software suites that can easily incorporate time dependent data and analyze it within the same suite of software. While, for example, the SAS Enterpise Miner Data Mining Suite allows a time component, it is necessary to go outside the Enterprise Miner Suite to actually deal with the time dependent data. Random coefficient logit models add another layer of complexity beyond the time problem in that they require complex and imprecise simulation and optimization methods to estimate the random coefficients of the Logit Model. The purpose of this research is to try to improve this estimation process using inexact arithmetic. Although the term inexact arithmetic seems to indicate even more imprecision, in reality it can provide precise bounds on the estimates.

This paper first presents the random coefficient logit model of demand and briefly outlines the relevant estimation issues. A brief theory of inexact arithmetic is then presented and then the combination of the two theories is indicated.

Estimation in the Random Coefficient Logit Model of Demand

An often referenced work in the process of estimating random coefficients in logit demand models is that by Aviv Nevo (2000, 2001) which provides techniques and program code for obtaining model estimates within an application context. Chintagunta et. al., 2002, has used the techniques of Nevo to produce important results using very large grocery store demand data sets available through the internet from the Kilts Center at the University of Chicago. Going back about 12 years, a simple form of the model has utility for a customer i on product j, (member of group g), represented as:

$U_{ij} = \delta_j + \zeta_{ig} + (1- \sigma)\ \varepsilon ij$, (Berry,1994).

More recent and complex models also include an index over time, t. The ζ_{ig} represent the random coefficients in the model.

The details of the estimation process become complicated and complex. The details of interest in this paper involve the following list of potential steps:

1) A Complex integral has to be computed by use of simulation for the full random coefficients model. This then can lead to several induced in accuracies in the result (Nevo, 2000).
2) In some circumstances a systemof non-linear equations must be solved. This can introduce several situations where the accuracy of the final solution is not controlled.
3) The success of the final stages of the estimation process depends on the accuracy of the first two steps above.

Through the use of inexact or interval arithmetic a higher level of guaranteed accuracy can be insured for the random-coefficients model. The next section discusses inexact arithmetic and its improvements to the estimation process.

Inexact Arithmetic

The concept of inexact arithmetic, originally known as interval arithmetic started with a Stanford University PhD dissertation written by R.E. Moore (1979). The technique involves representing each number precisely as an interval value. The smaller the interval, the more precise the number is. Arithmetic operations have rules comparable to those used in Fuzzy arithmetic, (Moore 1979). To use interval arithmetic correctly and produce bounded computations it was desirable to have compilers that supported this arithmetic type. Initially interval computations were very time consuming and compiler support was not easy. With the introduction of more flexible programming languages and systems such as C and Mathlab, and with the improvement in computer speed the draw backs to using interval computations have mostly disappeared. Interval arithmetic does still require the modification of algorithms to realize bounded computations, (Dinkel et. al. 1988, 1991, 1993, Moore 1988, Ratschek and Rokne 1988).

Moore, 1984 also devised a technique for using interval arithmetic to eliminate the need for large samples of points in simulation. When these techniques are applied to the above indicated stages of estimation in the random-coefficient logit problem above, some positive improvements have been obtained. Examples will be shown using the original data used by Nevo 2000, available on the web at:
http://rasmusen.org.g604/lectures/blp/frontpage.htm.

References

Berry, Steven T. 1994. Estimating Discrete-Choice Models of Product Differentiation. *RAND J. Econom.* 25(2) 242-262.

Chintagunta, P. K., Bonfrer, Andre, Song, Inseong. 2002. Investigating the effects of store-brand introduction on retailer demand and pricing behavior. *Management Science* 48(10), 1242-1267.

Dinkel, J.J. and Tretter, M.J. and Wong, D. 1988. Interval Newton Methods and Perturbed Problems, *Applied Mathematics and Computation*, 28(3), pp. 211-222.

_______, Wong, D. 1991. Some Implementation Issues Associated with Multidimensional Interval Newton Methods, *Computing*, 47, pp. 29-42.

Dinkel, J.J., Tretter, M.J. 1993. Mathematical Programs and Interval Analysis, *Mathematical Programming*, 61 (3), pp.377-384.

Moore, R. E. 1979 *Methods and Applications of Interval Analysis*. SIAM, Philadelphia, Pa.

_____. 1984. Risk analysis without Monte Carlo methods. *Freiburger Interval-Berichte. 1. 1-48.*

_____*.ED. 1988. Reliability in Computing: The Role of Interval Methods in Scientific Computing.* Academic Press. Boston.

Nevo, Aviv. 2000. A practitioner's guide to estimation of random-coefficients logit models of demand*. J Econom. Management Strategy* 9(4) 513-548.

_____. 2001. Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica 69(2), 307-342.*

Ratschek, H., Rokne, J. 1988. New Computer Methods for Global Optimization. John Wiley and Sons, New York.