DOCTORAL DISSERTATION:

Modelos de elección discreta en transportes con coeficientes aleatorios

(English translation: “Discrete choice models with random coefficients in transport”)

AUTHOR: Alfonso Orro. University of A Coruña (Spain)

THESIS SUPERVISOR: Francisco G. Benitez. University of Sevilla (Spain)

EXTENDED ABSTRACT

To represent the behaviour of travellers when they are deciding how they are going to get to their destination, discrete choice models, based on the random utility theory, have become one of the most widely used tools. The field in which these models were developed was halfway between econometrics and transport engineering, although the latter now constitutes one of their principal areas of application. In the transport field, they have mainly been applied to mode choice, but also to the choice of destination, route, and other important decisions such as the vehicle ownership.

In usual practice, the most frequently employed discrete choice models implement a fixed coefficient utility function that is linear in the parameters. The principal aim of the present thesis was to investigate the viability of specifying utility functions with random coefficients and that are nonlinear in the parameters in applications of discrete choice models to modeling transport.

Nonlinear specifications in the parameters were present in discrete choice theory at its outset, although they have seldom been used in practice. The specification of random coefficients, however, began with the probit and the hedonic models in the 1970s, and, after a period of apparent little practical interest, has burgeoned into a field of intense activity in recent years with the new generation of mixed logit models.

The specific objectives of this present work are:

- A contribution in some depth to the knowledge of models with random coefficients, studying the problems that can arise in their specification and estimation.
- Analysis of the application of the random coefficient discrete choice models to the specific problems of transport engineering, especially to mode choice modeling.
- Study of the justification of nonlinear utility functions, and development of a model that allows random coefficients and nonlinearities to be estimated together.
• Study of the possibility of imposing constraints on the maximization of the random utility in that model, and implementing the estimate with different types of bounds and constraints.

• Analysis of the consequences of using inadequate models that do not match reality with respect to the linearity of the utility function and the random variations in the coefficients.

The body of the document is structured into three blocks, and then ends with the conclusions and suggested lines of future research.

The first block describes the current situation of knowledge in this field of study and its application in this thesis. It gives a review and analysis of the literature on the topic, in many cases simultaneous with the development of the work. Chapter 2 sets out the bases of the models of behaviour of travellers when they are making transport choices. The random utility theory is presented, and the different models that have been developed in the framework of that theory are briefly described. Chapter 3 focuses on the description, formulation, and specification of the mixed logit model on which this thesis is based, and presents the base estimation that was implemented. An analysis is given of the characteristics of a broad sample of models that have been estimated with this specification in recent years. Chapter 4 discusses the significance and consequences of considering random coefficients in transport mode choice models.

The second block deals with the conceptual developments of the thesis. Chapter 5 presents and justifies the use of nonlinear utility functions in transport models. The Box-Cox Mixed Logit model, which is original in this thesis, is defined and formulated. Chapter 6 represents a contribution to the knowledge of the problem of identification in mixed logit models with random coefficients. The results of the thesis of Joan Walker are adapted to the specification of random coefficients, and new theoretical and experimental results are obtained, with conceptual explanations put forward for the phenomena discovered. Chapter 7 describes the estimation of the Box-Cox mixed logit model with a priori information when constraints are imposed on the maximization of the utility. The constraints are formulated, and the implementation of the estimation is described.

The third block constitutes the experimental and application part of the thesis, recalling part of the experimental results already presented in Chapter 6. Chapter 8 discusses the studies conducted on synthetic data in order to analyze the consequences of using inadequate models that do not match reality. To this end, a series of "fictitious realities" is constructed on which to apply different specifications, and forecasts are made for a set of policies changing the attributes. An analysis is given of the results, the usefulness of the different statistical tests applied to determine the best specification, and the risks of using inadequate specifications. Chapter 9 applies the models that have been analyzed to a real case. The results of the experimental part show that there is an interaction between the random coefficients and the non-
linearity. Chapter 10 is centred on the analysis of this phenomenon, its conceptual explanation, and its repercussions.

Chapter 11 presents the conclusions, and the lines of future research that have been opened. The work ends with a—necessarily extensive because of the nature of the thesis—list of the literature consulted.

The thesis combines a compilation and synthesis of the existing literature with new theoretical and experimental developments. The principal contributions that it includes will be briefly synthesized in the following, indicating the chapter of the document in which they are to be found:

- The characteristics were analyzed of a major fraction of the mixed logit estimations that have been made public. These results are presented at the end of Chapter 3.

- The characteristics of random coefficient transport models were unified, analyzing the justification of the presence of the different types of attributes in the utility function. Formulae were derived for point elasticity in mixed logit models. An analysis was made of the consequences in the field of transport of the choice of a particular type of random distribution for the coefficients. These results, which are essentially of compilation and synthesis, are presented throughout Chapter 4.

- Computer code was developed and implemented for a new discrete choice model that includes the combined estimation of random coefficients and non-linearity of the variables by means of a Box-Cox transformation. Simultaneously with the development of this thesis, and independently, Matthieu de Lapparent has described theoretically a similar model, but without coming to its estimation. Also, the last versions of one of the estimation computer programs (BIOGEME) have the capacity to estimate a model like the present, but there is no evidence that these developments were carried out prior to the present work. Chapter 5 presented the formulation for the model's specification and estimation, and Chapter 9 applied it to a real case.

- The issue of identification for the mixed logit with random coefficients specification was analyzed, which Joan Walker had extrapolated from the basis of other specifications of the model. A conceptual interpretation was given of the constraint on the valid normalizations for the variances of the specific constants of the choices, which had been previously deduced analytically. It was shown empirically that obtaining random coefficients for continuous attributes of the choices will depend on the relative importance of the independently and identically distributed errors with respect to the differences between the choices. It was confirmed experimentally that the conclusions drawn by Ben-Akiva, Walker and Bolduc (2001) for a single attribute of the mode choices can be extrapolated to several attributes. The random component identification
rules were deduced analytically and verified experimentally for the case of random coefficients applied to the categorical characteristics of the decision-maker. These had not been defined previously in the literature that is known to the author. These developments relating to the problem of identification are presented in Chapter 6.

- The estimation implemented included bounds, constraints that are linear in the parameters, and constraints on the elasticity of the choice rates (pre-defined prognosis criteria) in the Box-Cox mixed logit model. These constraints had been implemented by Francisco García Benítez and Javier Vázquez for other discrete choice models. The interpretation of pre-defined prognosis criteria as elasticity constraints is original in this thesis. These formulations were presented in Chapter 7.

- An experimental study was made, by means of various series of synthetic data, of the consequences of utility function specifications different from those that individuals use in their choices, both with respect to the randomness of the coefficients and to the linearity of the parameters. The capacity of standard statistical tests to correctly identify the effects present was analyzed. Analyses were also made of the consequences of incorrect specifications on the forecasts of the models relative to changes in the attributes, and of the differences in the forecasts using different specifications with real data. Chapter 8 presented the results with synthetic data, and Chapter 9 with real data.

- In the case of the absence of linearity in an attribute, it was shown experimentally, and a conceptual explanation put forward, that the conventional specification of the mixed logit model with random coefficients may well pass the standard statistical tests, but yield erroneous forecasts and mistaken interpretations. This problem is overcome with the use of the present Box-Cox mixed logit model or with additional statistical tests. This risk of erroneous results, set out in Chapter 10, was not reflected in the known literature, and is possibly the contribution of most importance of this thesis.