Global & Local Economic Review

Volume 13 No. 2 2009

TRACCE EDITION

Global & Local Economic Review

Aut. Trib. PE n. 7 del 14.7.1999, No. 2/2009
Editor
Nicola Mattoscio

ISSN (*print*) 1722-4241   ISSN (*online*) 1974-5125

«Global & Local Economic Review» is included in JEL on CD, e-JEL and Econlit, the electronic indexing and abstracting service of the American Economic Association

Editorial Offices: Corso Umberto I, n. 83 - 65122 Pescara
Tel. +39 0854219109 – Fax +39 0854219380
Website: www.gler.it
E-mail: gler@fondazionepescarabruzzo.it

G. & L. E. R., Vol. 13 No. 2, 2009

Antonio Lucadamo*, Luigi D'Ambra, Francesca Pagliara

MULTINOMIAL LOGIT MODELS: MULTICOLLINEARITY AND ITS CONSEQUENCES ON PARAMETER ESTIMATION AND ON INDEPENDENCE FROM IRRELEVANT ALTERNATIVES TESTS

Abstract

The Multinomial Logit Model is one of the most used Discrete Choice Models and it has been widely used to study the transportation system. This model has a closed form and so it is computationally cheap, but it needs some restrictive assumptions. The Independence from Irrelevant Alternatives (IIA) property for example, is not always respected and the tests used to verify it can lead to opposed results almost when there is multicollinearity between explicative variables. Furthermore, the presence of multicollinearity can generate problems in the parameter estimation. In this paper we study the effects of the multicollinearity on the results of the IIA tests and we propose a new method to estimate the parameters of the multinomial logit model. The data used for the analysis concern the choice of the residence in the Zurich area.

JEL CLASSIFICATION: C25; C51

KEYWORDS: DISCRETE CHOICE; MULTINOMIAL LOGIT; PRINCIPAL COMPONENT ANALYSIS; MULTICOLLINEARITY

1. Introduction

Urban mobility is a vital component of any city, often influencing its physical shape as well as its level of economic and social development. However the governance of urban mobility is very complex since it is

[1] TEDASS - University of Sannio, Italy - alucadam@unina.it.

the result of the interactions of several elements, i.e. of the functioning of a socio-technical complex system (Cascetta *et al.*, 2008).

More specifically, a *transportation system* can be defined as the combination of elements and their interactions, which produce the demand for travel within a given area and the supply of transportation services to satisfy this demand. This definition is general and flexible enough to be applied to different contexts. The specific structure of the system is defined by the problem itself (or class of problems) for whose solution it is system employed.

Almost all of the components of a social and economic in a given geographical area interact with different levels of intensity. However, it is practically impossible to take into account every interacting element to solve a transportation engineering problem. The typical system engineering approach is to isolate those elements, which are most relevant to the problem. These elements, and the relationships among them, make up the *analysis system* (Cascetta, 2009). The remaining elements belong to the *external environment* and are taken into account only in terms of their interactions with the analysis system. In general, the analysis system includes the elements and the interactions that are expected to be significantly affected by the projects under consideration. It follows that there is a strict *interdependence between the identification of the analysis system and the problem to be solved*. The transportation system of a given area can also be seen as a sub-system of a wider territorial system with which it strongly interacts. The extent to which these interactions are included in the analysis system, or else in the external environment, depends on the specific problem (Manheim, 1979).

The *transportation system* can be split into two main components: demand and supply.

The distribution of households and activities in the area is the determinant of *transportation demand* deriving from the need to use different urban functions in different places. Household members are the users of the transportation supply system and make "mobility choices" (holding a driving license, number of cars, etc.) and "travel choices" (trip frequency, time, destination, mode, path, etc.) in order to undertake activities (work, study, shopping, etc.) in different locations. The result of these choices is the transportation demand; i.e. the number of trips made among the different zones of the city, for different purposes, in different periods, by means of the different available transportation modes. Similarly, economic activities transport goods that are consumed by the households or by other economic activities. Goods movements make up the freight transportation demand.

Both mobility and travel choices are influenced by some characteristics of the transportation services offered by the different travel modes (individual car, transit, walking). These characteristics are known as level-of-service or performance attributes and include travel times, monetary costs, service reliability, riding comfort, etc. Thus, the choice of destination may be influenced by the travel time and cost needed to reach each destination. The choice of departure time depends on the travel time to the destination. The choice of transportation mode is influenced by times, costs, reliability of the available modes.

The characteristics of transportation services depend on the *transportation supply*, i.e., the set of facilities (roads, parking spaces, railway lines, etc.), services (transit lines and timetables), regulations (road circulation and parking regulations), and prices (transit fares, parking prices, road tolls, etc.) producing travel opportunities. The physical elements of the transportation supply system have a finite capacity; i.e. a maximum number of users that can be served in a given time interval.

Individual trips can be aggregated into users flows, i.e. the number of users on the physical elements of the supply system in a given time interval. Examples are automobile and truck flows on road sections, passenger flows on transit lines, and so on.

When flow approaches capacity, the interactions among users increase and *congestion* effects are triggered. Congestion can significantly deteriorate the performances of transportation services for the users, e.g. travel times, service delays, fuel consumptions all increase with congestion. Congestion can also have other "external" negative effects (such as noise, air pollution and visual impacts in the case of road traffic). Congestion can have cross-modal effects; e.g. road congestion can influence the performances of surface transit services.

2. Modeling transportation systems

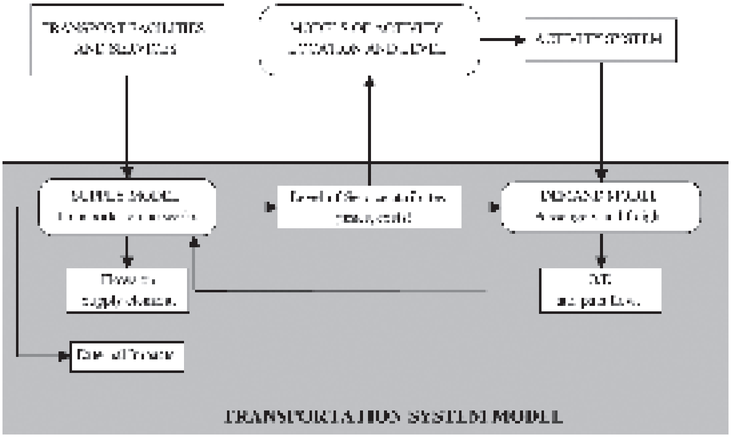
The relevant interactions among the various elements of a transportation system can be simulated with mathematical models; the models and their relationships are described in Fig. 1 (Cascetta, 2009).

Supply models simulate the transportation services available among the different zones with flow network models. More specifically, supply models simulate the performance of transportation infrastructures and services for the users, as well as the main external effects of transport (pollution, energy consumption, accidents). The level-of-service attributes, such as travel time and cost, will be input variables for the demand models. To simulate the performance of single elements (facilities) and the effects of congestion, especially for road systems, supply models use the results of traffic flow theory (Sheffi, 1985). For the transit system the hyper path approach is generally considered.

Demand models simulate the relevant aspects of travel demand as a function of the activity system and of the supply performances. Typically, the characteristics of travel demand simulated include the number of trips in the reference period (demand level) and their distribution among the different zones, the different transport modes, and the different paths. Other components of travel demand are simulated in specific applications such as the distribution between different time intervals within the reference period. Travel demand models are usually derived from random utility theory.

Assignment models (or network demand-supply interaction models) simulate how O-D demand and path flows load the various elements of the supply system. Assignment models allow the calculation of link flows, i.e. the number of users loading each link of the network representing the transportation supply in the reference period (Ortuzar and Willumsen, 2001). Furthermore, link flows may affect the transportation supply performances through congestion and therefore may affect the input to demand models. The mutual interdependencies of demand, flows and costs are simulated by assignment models.

Fig. 1 - Structure of transportation systems simulation models ***Source: Cascetta (2009).***



In this paper the focus is on demand models.

2.1 The random utility theory for modeling transportation users' behaviour

Random utility theory is based on the hypothesis that every individual is a *rational decision-maker*, maximizing utility relative to his/her choices. Specifically, the theory is based on the following assumptions:

a. the generic decision-maker i, in making a choice, considers m_i mutually exclusive alternatives that constitute his/her *choice set* C. The choice set may differ according to the decision-maker (for example, in the choice of transport mode, the choice set of an individual without a driver's license or car obviously does not include the alternative "car as a driver");
b. decision-maker i assigns to each alternative c in his/her choice set a perceived utility or "attractiveness" U_{ic} and selects the alternative that maximizes this utility;
c. the utility assigned to each choice alternative depends on a number of measurable characteristics, or *attributes*, of the alternative itself and of

the decision-maker: $U_{ic} = U_{ic}(\boldsymbol{X}_{ic})$, where $\boldsymbol{X}_{ic}$ is the vector of attributes relative to alternative *c* and to decision-maker *i*;

d. because of various factors that will be described later, the utility assigned by decision-maker *i* to alternative *c* is not known with certainty by an external observer (analyst) wishing to model the decision-maker's choice behavior, so U_{ic} must be represented in general by a random variable.

From the above assumptions, it is not usually possible to predict with certainty the alternative that the generic decision-maker will select. However, it is possible to express the probability that the decision-maker will select alternative *c* conditional on his/her choice set *C*; this is the probability that the perceived utility of alternative *c* is greater than that of all the other available alternatives:

$$p_i\,[\,c/C\,] = \boldsymbol{Pr}\,[\,U_{ic} > U_{ib} \;\; \forall\; b \neq c,\, b \in C\,] \tag{1}$$

The perceived utility U_{ic} can be expressed as the sum of two terms: a systematic utility and a random residual. The *systematic utility* V_{ic} represents the mean (expected value) utility perceived by all decision-makers having the same choice context (alternatives and attributes) as decision-maker *i*. The *random residual* ε_{ic} is the (unknown) deviation of the utility perceived by user *i* from this mean value; it captures the combined effects of the various factors that introduce uncertainty into choice modeling:

$$U_{ic} = V_{ic} + \xi_{\mathrm{ic}} \;\; \forall\; c \in C \tag{2}$$

The probability of choosing an alternative depends on the systematic utilities of all competing (available) alternatives, and on the joint probability law of the random residuals ε_{ic}. It follows:

$$p_i(c \mid C) = \boldsymbol{Pr}\,[V_{ic} + \varepsilon_{ic} > V_{ib} + \varepsilon_{ib} \quad \forall b \neq c,\, b \in C] \tag{3}$$

Systematic utility is expressed as a function $V_{ic}(X_{ic})$ of attributes X_{ic} relative to the alternatives and the decision-maker. Although in principle the function $V_{ic}(X_{ic})$ may be of any type, it is usually assumed for analytical and statistical convenience that the systematic utility V_{ic} is a linear function, with coefficients β, of the attributes X_{ic} or of functional transformations of them:

$$V_{ic}(\boldsymbol{X}_{ic}) = \beta^{\mathrm{T}} \boldsymbol{X}_{ic} \tag{4}$$

3. The Multinomial Logit Model

As shown in the previous section, the probability that any element *c* in C is chosen by the decision maker *i* can be expressed according to the equation (3).

Any particular multinomial choice model can be derived using the previous equation given specific assumptions on the joint distribution of the disturbances. It can be shown, as in Domencich and McFadden (1975), that if the error terms are:

- Independently distributed;
- Identically distributed;
- Gumbel distributed with a location parameter η and a scale parameter $\mu \geq 0$;

then the probability that alternative *c* will be chosen is

$$p_i(c \mid C) = \frac{\exp(\mu V_{ic})}{\sum_{b \in C} \exp(\mu V_{ib})} \tag{5}$$

If we consider that the utility is linear in parameters and we fix the scale parameter to 1, we can write:

$$p_i(c \mid C) = \frac{\exp(x_{ij} \beta_{jc})}{\sum_{b \in C} \exp(x_{ij} \beta_{jb})} \tag{6}$$

where C is the choice set of alternatives, x_{ij} are the values that the individual *i* assumes for the variable *j* and β are the parameters to be estimated (Ben-Akiva & Lerman, 1985).

The logit model has some special properties that under certain circumstances greatly simplify estimation of the parameters. Most of these theories can be addressed to McFadden (1974). The technique generally used to estimate a logit model is the Maximum Likelihood and for parameter estimation it is necessary to use some iterative algorithm. One of the most used is the Newton-Raphson algorithm that considers the first and second derivatives of the previous equation.

The first derivative is also called gradient vector and it is given by

$$g = \frac{\partial \log L(\beta)}{\partial \beta_c} = \sum_i^n x_i^{'} [y_{ic} - P_{ic}] \tag{7}$$

The second derivative, or Hessian matrix is:

$$H = \frac{\partial^2 \log L(\beta)}{\partial \beta_c \partial \beta_q^{'}} = -\sum_i^n P_{iq} [\delta_{cq} - P_{iq}] \boldsymbol{x}_i{'} \boldsymbol{x}_i \tag{8}$$

or in matrix form:

$$H = -X^{'} Z_{cq} X \tag{9}$$

Where Z_{cq} is a diagonal matrix with generic term $P_{iq} [\delta_{cq} - P_{ic}]$, $\delta_{cq} = 1$ if q=c and $\delta_{cq} = 0$ otherwise.

The parameter estimate according to Newton-Raphson is then:

$$\beta^{t+1} = \beta^t - (H)^{-1} g \tag{10}$$

It's important to consider the Hessian matrix, because its inversion could lead some problems in the estimation of the Multinomial Logit Model (D'Ambra, 2008).

4. Logit properties

One of the most discussed aspects of the multinomial logit is the

Independece from Irrelevant Alternatives property (IIA). This property states that for any two alternatives c and *d* the ratio of the logit probabilities

$$\frac{P_i(c)}{P_i(d)} = \frac{\exp(V_{ic})}{\exp(V_{id})} = \exp(V_{ic} - V_{id}) \tag{11}$$

does not depend on any alternatives other than c and *d.* Therefore the relative odds of choosing *c* over *d* are the same no matter what other alternatives are available or what the attributes of the other alternatives are. Since the ratio is independent from alternative other than *c* and *d,* it is said to be Independent from Irrelevant Alternatives. This assumption is realistic in some choice situation, but sometimes it can be clearly inappropriate. One of the classical example is the famous red-bus blue-bus problem, but there are many other situation in which the IIA assumption is not respected.

In this context we must search for a better model specification:

- Finding alternatives with missing or mis-specified variables;
- Point toward an acceptable nested logit structure.

To verify if the IIA property holds many tests exist. We can divide them in two groups:

- Estimate a model with a subset of the choice set. Reject IIA if the parameter estimates differ from the full choice set estimates.

  1) Hausman and McFadden (1984)
  2) McFadden, Tye and Train (1976)
  3) Small-Hsiao (1985)

- Implement a Lagrange multiplier test of IIA with the full set of alternatives.

  1) McFadden test (1987)

4.1 Estimating model with choice subsets

If we suppose IIA holds, then:

$$p_i(c \mid C) = \frac{\exp(x_{ij}\beta_{jc})}{\sum_{b \in C} \exp(x_{ij}\beta_{jb})} \tag{12}$$

and

$$p_i(c \mid \tilde{C} \subseteq C) = \frac{\exp(x_{ic}\beta_{jc})}{\sum_{b \in \tilde{C}} \exp(x_{ib}\beta_{jb})} \tag{13}$$

where $\tilde{C}$ is a subset of the full set of alternatives. It should provide similar estimates, since under IIA, exclusion of alternatives does not affect the consistency of estimators. As it was said, it is possible to use:

- Hausman-McFadden test (HM).

We have to build the following statistic:

$$\left(\hat{\beta}_{\tilde{C}} - \hat{\beta}_C\right)' \left(\Sigma_{\hat{\beta}_{\tilde{C}}} - \Sigma_{\hat{\beta}_C}\right)^{-1} \left(\hat{\beta}_{\tilde{C}} - \hat{\beta}_C\right) \tag{14}$$

That is asymptotically χ^2 distributed with $\tilde{C}$ degrees of freedom, where C is the number of elements in the subvector of coefficients that is identifiable from the restricted choice set model. So the null hypothesis that IIA holds is rejected if the value that comes from the equation 14 is bigger than the tabulated value of χ^2.

- McFadden, Tye and Train test (MTT).

In this case it's possible to build an approximate likelihood ratio test statistic with C degrees of freedom:

$-2\left[l_{\tilde{C}}\left(\hat{\beta}_C\right) - l_{\tilde{C}}\left(\hat{\beta}_{\tilde{C}}\right)\right]$ where the two log likelihood values are calculated on the estimation sample for the restricted choice set model. This statistic is not a proper likelihood ratio test because β_C is not a vector of constants. For this reason we can consider the following correction:

• Small and Hsiao (SH). To remove the bias they proposed to use:

$$\frac{1}{1-C_1(\alpha C)}\left\{-2\left[l_{\tilde{C}}\left(\hat{\beta}_C\right)-l_{\tilde{C}}\left(\hat{\beta}_{\tilde{C}}\right)\right]\right\} \tag{15}$$

Where C is the number of observations in the unrestricted choice set estimation, C_1 is the number of observations in the restricted choice set estimation ($C_1 < C$) since those observations with chosen alternatives not in the restricted choice set are omitted, and $\alpha \geq 1$ is a scalar. Asymptotically this corrected likelihood ratio statistic actually is χ^2 distributed with C degrees of freedom. Sometimes the assumption made by this correction that a scalar difference between the covariance matrices exists is not defensible, so it was proposed an exact test for the IIA assumption. To perform the test Small and Hsiao randomly divided the full estimation data set into two parts, denoted A and B. On sample A, using the restricted choice sets, they estimated $\hat{\beta}^A_C$, the subvector of coefficients corresponding to the parameters that are identifiable when the restricted set of alternatives are used; next, on sample B, using the restricted choice set, they estimated $\hat{\beta}^B_C$ and the corresponding log likelihood, $l^B_{\tilde{C}}\left(\hat{\beta}^B_C\right)$, finally again on sample B, but now based on the unrestricted choice sets, they obtained $\hat{\beta}^B_C$. They showed that if we form the following convex combination:

$$\hat{\beta}^{AB}_{\tilde{C}} = \left(1/\sqrt{2}\right)\hat{\beta}^A_C + \left(1\text{-}1/\sqrt{2}\right)\hat{\beta}^B_C \tag{16}$$

And use it to evaluate the log likelihood of the sample B with the restricted choice set, denoted as $l^B_C\left(\hat{\beta}^{B}_{AB}\right)$, then the statistic $-2\left[l^B_{\tilde{C}}\left(\hat{\beta}^{AB}_C\right)-l^B_{\tilde{C}}\left(\hat{\beta}^B_C\right)\right]$ is asymptotically χ^2 distributed with $\tilde{C}$ degrees of freedom, C being the common dimension of the $\hat{\beta}^A_C, \hat{\beta}^B_C, \hat{\beta}^B_{\tilde{C}}, \hat{\beta}^B_{AB}$ parameter vectors. This test is more computationally and time-consuming. Generally it's better if the simpler corrected approximate likelihood ratio test is carried out first, and then, only if its underlying assumption is violated, should the exact test procedure be used.

4.2 Lagrange multiplier test of IIA

This test checks if cross-alternative variables enter the model. If so, IIA assumption is violated. The test is composed of 3 steps:

- **Step 1**: We estimate the systematic utilities $\hat{V}_{ic}$ and fitted choice probabilities $\hat{p}_i(c \mid C)$ using all N observations: $\hat{V}_{ic} = \beta' x_{ic} \forall c \in C$;

$$\hat{P}_i(c \mid C) = \frac{\exp(x_{ij} \hat{\beta}_{jc})}{\sum_{b \in C} \exp(x_{ij} \hat{\beta}_{jb})}.$$

- **Step 2**: For a given $A \subset C$ we calculate auxiliary variables in the following way:

$$\hat{V}_{iA} = \frac{\sum_{b \in A} \hat{V}_{ib}\, \hat{p}_i(b \mid C)}{\sum_{b \in A} \hat{p}_i(b \mid C)} \qquad i = 1, \ldots, I \tag{17}$$

$$Z_{icA} = \begin{cases} V_{ic} - \hat{V}_{iA} & f \quad c \in A \\ 0 & otherwise \end{cases} \qquad i = 1, \ldots, I \tag{18}$$

Since Z is non zero only for the alternatives in the set A, it contains information regarding the other alternatives in A. The spirit of the proposed test is to verify the presence of cross alternative variables.

- **Step 3:** We can now estimate:

$$\hat{p}_i(c \mid C) = \frac{\exp\left(x_{ij} \hat{\beta}_{jc} + \upsilon^A Z_{icA} \right)}{\sum_{b \in C} \exp\left(x_{ij} \hat{\beta}_{jb} + \upsilon^A Z_{ibA} \right)} \tag{19}$$

The hypotheses are the following:

$$H_0 : \upsilon^A = 0$$
$$H_1 : \upsilon^A \neq 0$$

υ^A is distributed as a χ^2.

If we reject H_0 we reject the IIA assumption. If H_0 is not rejected nest A is considered to satisfy IIA.

5. Multicollinearity

Both in linear regression model and in logistic regression model a very desirable condition is that there is no multicollinearity among the regressors. The term multicollinearity is due to Ragnar Frisch (1934). Originally it meant the existence of a exact linear relationship among some or all explanatory variables of a regression model. For p-variables regression involving explanatory variables $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ (where $\mathbf{x}_0 = 1$ for all observations to allow for the intercept term), an exact linear relationship is said to exist if the following condition is satisfied:

$$a_0\mathbf{x}_0 + a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \ldots + a_p\mathbf{x}_p = 0 \tag{20}$$

where a_0, a_1, a_2, ..., a_p are constants such that not all of them are zero simultaneously. The chances of one's obtaining a sample of values where the regressors are related in this fashion are indeed very small, in practise except when the number of observations is smaller than the number of regressors, so today the term multicollinearity is used in a broader sense to include the case of perfect multicollinearity, as shown by the previous formula as well as the case where the $\mathbf{x}$ variables are intercorrelated but not perfectly so, as follows:

$$a_0\mathbf{x}_0 + a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \ldots + a_p\mathbf{x}_p + \boldsymbol{v}_i = 0 \tag{21}$$

where v_i is a stochastic error term.

5.1 Multicollinearity and its links with test to identify the IIA property

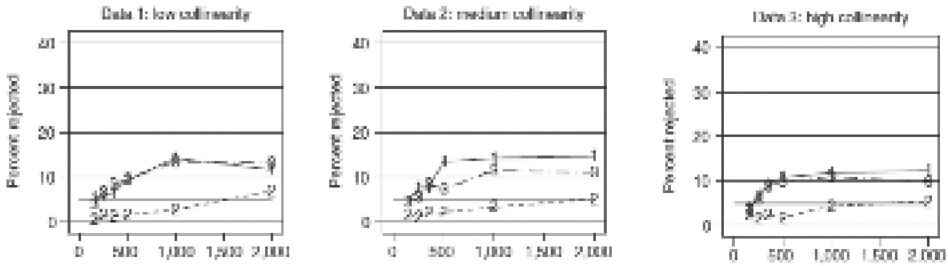
The presence of multicollinearity in the regressor matrix can lead to different results if we compute the tests to detect if the IIA property is respected or not. Cheng and Long (2007) showed in fact how the degree of collinearity, the choice of the category that is deleted in the restricted estimation and the sample sizes can change the properties of the tests. They run simulations for sample sizes of 150, 250, 350, 500, 1000 and 2000. The simulations involved the following steps:

- Draw a random sample size of N with replacement from the population.
- For this sample, estimate the Multinomial Logit Model for a data set with three predictors and the outcome variable with three categories.
- Using estimates from the previous step, compute three variations of the Hausman-McFadden test and of the Small-Hsiao test, excluding the first category for the restricted estimation, then the second and at least the third.

These steps were repeated 500 times for each sample size in each data set. To determine the empirical size for each test, they computed the percentage of times that each test rejected the null hypothesis that IIA held in the population at the 0,05 level of significance.

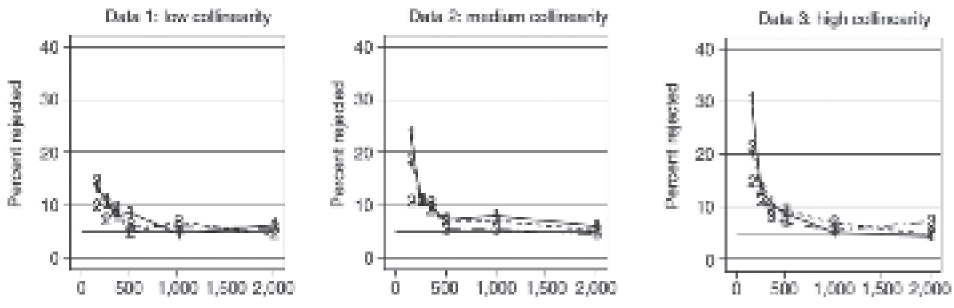
Fig. 2 shows the percentage of times the HM test rejected the null hypothesis of no violation of the IIA assumption using the 0,05 level. There are the graphs relative to three degrees of collinearity and with the elimination of one of the 3 categories for the outcome variables.

Fig. 2 - Size properties of the Hausman-McFadden test of the Independence of Irrelevant Alternatives



The results illustrate that the HM test does not reliably converge to its appropriate size even when the sample is 2000. Second, the properties of the test depend on which outcome category is deleted in the restricted estimation. In this case the degree of collinearity doesn't seem to have a particular influence on the results. But if we consider the Fig. 3 we can see that there are some differences.

Fig. 3 - Size properties of the Small-Hsiao test of the Independence of Irrelevant Alternatives



It is possible in fact to note that the SH test approximates its nominal size as the sample increases to 500 or 1000. The magnitude of departures from the nominal size and the sample size at which these distortions are largely removed depends on the degree of collinearity in the data. For example, with high collinearity, the size properties are quite poor with samples smaller than 500 and require a sample of at least 1000 before they are nearly eliminated.

The previous results show that tests of the IIA assumption that are based on the estimation of a restricted choice set are unsatisfactory for applied work. The Hausman-McFadden test shows substantial size distortion that is unaffected by sample size in the simulations. The Small-Hsiao test has reasonable size properties in some data sets but has severe size distortion even in large samples when there is collinearity in the regressors. For this reason it's important to be careful when a researcher wants to test the IIA assumption and it's important to consider not only the results of the tests, but the structure of the data too.

5.2 Multicollinearity and parameter estimation

As previously stated the presence of multicollinearity may indicate that some explanatory variables are linear combinations of the other ones. When this situation exists the logistic model estimations become unstable. To solve this problem in the binary case Marx (1992) introduced iteratively reweighted partial least squares algorithm, Bastien et al (2005) proposed partial least squares logit regression, Aguilera et al (2006) presented Principal Component Logistic Regression (PCLR), Vágó and Kemény (2006) developed the ridge logistic regression. A method to provide an accurate estimation of the model parameters in Multinomial Logit Model has instead been proposed by Camminatiello and Lucadamo (2008). They introduced in fact an extension of PCLR model, called Principal Component Multinomial Regression (PCMR). They proposed to use as covariates of the multinomial model a reduced number of Principal Components (PC) of the predictor variables.

PCMR creates at first step the PC's of the regressors as linear combination of the original variables $Z = XV$, where $X = \left[x_1, x_2, \ldots, x_p\right]$ is the set of p quantitative independent variables and $V = \left[v_1, v_2, \ldots, v_p\right]$ matrix., built by the eigenvectors of the correlation matrix R, whose elements are the correlation coefficients among the regressors. At second step the multinomial model is carried out on the subset of *p* PC's. The probability for the individual *i*, to choose the alternative *c* can be expressed in terms of all PC's as

$$p_i(c) = \frac{\exp\left\{\sum_{j=1}^{p}\sum_{k=1}^{p} z_{ik} v_{kj} \beta_{jc}\right\}}{\left\{\sum_{b=1}^{s}\exp\sum_{j=1}^{p}\sum_{k=1}^{p} z_{ik} v_{kj} \beta_{jb}\right\}} = \frac{\exp\left\{\sum_{k=1}^{p} z_{ik}\gamma_{kc}\right\}}{\left\{\sum_{b=1}^{s}\exp\sum_{k=1}^{p} z_{ik}\gamma_{kb}\right\}} \tag{22}$$

where $z_{ik}, (i = 1, \ldots, n; g = 1, \ldots, p)$ are the elements of the PC matrix, $v_{kj}, (k = 1, \ldots, p)$ are the elements of the transposed matrix $\mathbf{V}^T$, $\gamma_{kb} = \sum_{k=1}^{p} v_{kj}\beta_{jb}, (b = 1, \ldots, s)$ are the coefficients to be estimated, β_{jb} are the parameters expressed in function of original variables and *s* is the number of alternatives of the data set.

At third step only some of the components are chosen and the multinomial logit model is carried out only on this subset and finally the multinomial model parameters can be expressed in function of original variables: $\hat{\mathbf{a}}^{(a)} = \mathbf{V}^{(a)}\tilde{\mathbf{a}}^{(a)}$. The problem is to decide which components to retain in the analysis. One way is to choose only the components that have an eigenvalue bigger than one, that is one of the criteria used in Principal Component Analysis, but in this way we don't take in consideration the relation with the dependent variable. For this reason, the proposal is to use only the components that influence in statistical significant manner the response variable, without taking in account the eigenvalue.

The authors showed, via a bootstrap resampling, that this leads to lower variance estimates of model parameters comparing to classical multinomial model. The data they used for the analysis concern the choice of the residence in the Zurich area (Burgle, 2006) and the analysis are carried out by BIOGEME (Bierlaire, 2003). For this analysis they showed what's happen about the bootstrap estimate of variance of the estimated parameters. The results are reported in table 1 and they are relative only to the variables that are significant on the original dataset.

Tab. 1 – Variance of estimated parameters for different methods

Variables	Basic value	MNL	PCMR(6)	PCMR(2)
Distance from working place	0.795633	0.034862	0.01947	0.00264
Rent in CHF	0.101965	0.08315	0.00262	0.000545
Density of open space (500 m radius)	0.000354	0.016107	0.0026	0.00191
Density of young people (1 km radius)	4.390690	0.002317	0.01596	0.00296
Sun index	0.119155	0.026059	0.00974	0.00593
Driving time to city center	2.917186	0.456245	0.00444	0.00417

In the first column there are the names of the variables, in the second column there are the coefficient values, then, from the third to the fifth column, the variances of the estimated parameters, calculated for different datasets: the original matrix; the six PC's with eigenvalue bigger than one; the two components that are significant to explain the dependent variable. It is easy to observe that the variance of the estimated parameters computed on the original variables, is always bigger than the estimated variance calculated using the PCMR and that lower variance estimates of model parameters (last column) are obtained retaining in the model only the significant components.

6. Conclusions and further perspectives

Transport planning, infrastructure project evaluation and policy making, particularly at the urban level, continue to be important issues in the 21st century. Transport modelling requires mathematical techniques in order to make predictions, which can then be utilised in planning and design. This is the basis for improved decision-making and planning in the transport arena.

This paper has focused on some of the issues connecting with discrete choice models, which are the most commonly used tools to model users' behaviour. By far the model specification which is used most often is the Multinomial Logit Model which provides a convenient closed form for the underlying choice probabilities without any requirement of multivariate integration. Therefore, choice situations characterized by many alternatives can be treated in a computationally convenient manner. The ease of computation and the existence of a number of computer programs has led to many applications of the Logit model. Yet it is widely known that a potentially important drawback of the MNL model is the independence from irrelevant alternatives property. This property states that the ratio of the probability of choosing any two alternatives is independent of the attributes of any other alternative in the choice set. But we showed that, in many circumstances, this property is not respected and the application of the Multinomial Logit Model can lead to some problems. Furthermore, the different tests used to verify if the IIA property has been violated can give opposite results when there is multicollinearity between the explicative variables. In these circumstances, if at least one of the tests leads to a rejection of the IIA property, it is better to use another model, for example a Nested Logit model. In this case, we need to build the nests for the model, but there is not a standard way to construct the classes. Therefore, when we have many alternatives, there are different possibilities and it is not so easy to find the right combination. For this reason, we think that the use of a cluster analysis combined with a multivariate method can be useful to provide information about the construction of the nests.

Furthermore we showed that the use of PCMR can improve parameters estimation in Multinomial Logit Model. Obviously further research is necessary on the new proposed method as, for example, an extensive simulation study to verify the results and the method for selecting the PC's, but we think that it can be a first solution to the problem of multicollinearity. At the same time a generalization of other models proposed in literature for solving the problem of high-dimensional multicollinear data in the binary logit model could be interesting.

References

Aguilera A.M., Escabias M., Valderrama M.J. (2006), *Using principal components for estimating logistic regression with high-dimensional multicollinear data*, Computational Statistics & Data Analysis, 50, 1905-1924.

Bastien P., Esposito V., Tenenhaus M. (2005), *PLS generalised linear regression*, Computational Statistics & Data Analysis, 48 (1), 17-46.

Ben-Akiva M., Lerman S. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge, Ma.

Bierlaire M. (2003), *BIOGEME: A free package for the estimation of discrete choice models*, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Burgle M. (2006), *Residential location choice model for the greater Zurich area*, in '6th Swiss Transport Research Conference', Monte Verità, Ascona.

Camminatiello I., Lucadamo A. (2008), *Estimating multinomial logit model with multicollinear data,* Proceedings of conference "Metodi Modelli e Tecnologie dell'informazione a Supporto delle Decisioni" – Lecce, Italy.

Cascetta E. (2009), *Transportation Systems Analysis: Models and Application*, Springer New York.

Cascetta E., Pagliara F., Papola A. (2008), *Governance of urban mobility: complex systems and integrated policies*, Advances in Complex Systems, 10, 339-354.

Cheng S., Long J.S. (2007), *Testing for IIA in the multinomial logit model*. Sociological Methods & Research 35 (4), 583-600.

D'Ambra L. (2008), *Modelli di regressione a risposta discreta*. Appunti del corso "Modelli statistici per le decisioni economiche" A.A. 2008-2009.

Domencich T., McFadden D. (1975), *Urban Travel Demand – A behavioral Analysis*, North Holland, Amsterdam.

Hausman J., McFadden D. (1984), *Specification tests for the multinomial logit model*, Econometrica 46, 403–426.

Manheim M. (1979), *Fundamentals of transportation systems analysis*, MIT Press, Cambridge, Mass.

Marx B.D. (1992), *A continuum of principal component generalized linear regression*, Computational Statistics & Data Analysis, 13, 385-393.

McFadden D. (1974), *Conditional logit analysis of qualitative choice behavior*, in P. Zarembka, ed., 'Frontiers in Econometrics', New York: Academic Press, 105–142.

McFadden D., Tye W., Train K. (1976), *Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model*, Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper No. 7616, Aug.

McFadden D. (1987), *Regression-based specification tests for the multinomial logit model*. Journal of Econometrics. 34, 63-82.

Ortuzar J.de D., Willumsen L.G. (2001), *Modelling Transport*, John Wiley and Sons, New York, 3rd edition.

Sheffi Y. (1985). *Urban transportation networks*. Prentice Hall, Englewood Cliff, NJ.Wachs.

Small K., Hsiao C. (1985), *Multinomial Logit Specification Tests*, Technical report, depatment of Economics Princeton University, Princeton, N. J.

Vágó E., Kemény S. (2006), *Logistic ridge regression for clinical data analysis (a case study)*, Applied ecology and environmental research 4(2), 171-179.

Finito di stampare
nel mese di Dicembre 2009
dalla Litografia Brandolini - Sambuceto
per le Edizioni TRACCE
Via Eugenia Ravasco, 54
65123 PESCARA
Tel. 085/76658
www.tracce.org

In the latest centuries, generally speaking, history records the alternation of some important seasons which lend themselves to represent economic models, which are the bases of modern economic thought.
First of all, there is the age of *colonial economy* centered on the role of imperial states, together with the birth of monopolistic companies, in the management of trades with dominion areas.
Then, the age of *international economy* was lived, culminating in the second post war trade relation system. It was mainly founded on the functions of the national states and their authorities to support both national espansionary fiscal policy and exchange clearings, in their trade ratios with the rest of the world.
At last, in the latest years, *interglobal economy* took vehemently the lead through the modern electronic infrastructures of telematic and telecommunications.
As the former models, the *interglobal economy* too does not automatically assure either stable equilibrium or the overcoming of traditional crises.
It gives benefits and disadvantages too.
From the normative and positive points of view, one of the disadvantages which most drew the attention of researchers is the weakening and disappearance of national and subnational economic and monetary policy instruments.
Instead one of the benefits which most attracted interest might be located on the nature itself of the technological revolution in progress, foreboding new opportunities in the integration process of local economic systems, which might qualify themselves as network growth links (or growth poles?).
The Review has the aim to represent and to inquire the normative and positive profiles of the foundamentals which might characterize the thin and difficult frontier between globalization and economic localism.

FONDAZIONE
PESCARABRUZZO

Euro 12,91

ISSN 1722-424 1
9 771722 424009