UNIVERSITY OF NAPLES FEDERICO II

XXXI CYCLE OF PHD IN
CIVIL SYSTEMS ENGINEERING

DEPARTMENT OF CIVIL, ARCHITECTURAL AND ENVIRONMENTAL ENGINEERING

PHD THESIS

Investigating the potential of the combination of random utility models (CoRUM) for discrete choice modelling and travel demand analysis

Supervisor:
Prof. Andrea Papola

PhD Program Coordinator:
Prof. Andrea Papola

PhD candidate:
Fiore Tinessa

DECEMBER 2018

Alla mia famiglia,

perché tutto il tempo

che non le ho potuto dedicare

è stato impiegato per questo.

Summary

Summary **3**
List of figures **7**
List of tables **9**
Chapter 1 : Introduction **11**
1.1 Overview of choice modelling issues 11
1.2 The CoRUM basic idea 13
1.3 Objectives and contributions of the thesis 14
1.4 Thesis canvas 15
Chapter 2 : State of the Art **17**
2.1 Random utility theory – general assumptions 17
2.1.1 Random utility *17*
2.1.2 Additive and multiplicative approaches *18*
2.1.3 Maximizing benefit behaviour *18*
2.1.4 Random utility maximization models (RUMs) *19*
2.2 State of the art of RUMs 21
2.2.1 Error structure *21*
2.2.1.1 GEV models 22
Multinomial Logit 22
Nested Logit 25
Cross Nested Logit and further generalizations 26
FinMix 28
2.2.1.2 Multinomial Probit 29
2.2.1.3 Mixed models 32
Error Component Logit 32
Normal alternative specific Error Component GEV 34
2.2.2 Taste heterogeneity *34*
2.2.2.1 Observable taste heterogeneity 35
2.2.2.2 Parametric approach 36
Probit with random coefficient formulation 36

Random coefficient Logit ... 37
Posterior analysis and individual-level parameters on continuous distributions ... 38
2.2.2.3 Nonparametric approach ... 39
Latent class MNL ... 40
Discrete mixture of MNL ... 41
Variation on the theme ... 41
Posterior analysis with discrete distributions ... 42
2.2.2.4 Semi-nonparametric approach ... 42
Mixed Logit as mixture of parametric distributions ... 43
Latent class MNL with endogenous segmentation ... 43
Logit-Mixed Logit model ... 44
Latent class as a finite mixture of continuous distributions (Latent class Mixed Logit) ... 44
2.2.2.5 Variation on a theme – the scale heterogeneity ... 45
2.2.3 Unified framework ... 46
Probit and Mixed Logit with joint EC/RC specification ... 46
Mixed GEV ... 47
Mixed Probit ... 48
Latent class GEV ... 48
2.2.4 Going further ... 48
2.3 Application of RUMs to route choice ... 50
2.3.1 Route choice modelling: general issues ... 53
2.3.2 Logit based route choice models ... 54
2.3.2.1 Multinomial Logit - Dial's algorithm ... 55
2.3.2.2 C-Logit ... 57
2.3.2.3 Path-Size Logit ... 58
2.3.2.4 IAP Logit ... 59
2.3.3 GEV models for route choice ... 59
2.3.3.1 Link Nested Logit ... 60
2.3.3.2 Pair Combinatorial Logit ... 61
2.3.3.3 Link Based - Network GEV ... 62
2.3.3.4 Multinomial Weibit ... 62
2.3.4 Error component models ... 63
2.3.4.1 Multinomial Probit and Monte-Carlo algorithm ... 63
2.3.4.2 Mixed Logit ... 64
2.3.5 Going further ... 65

2.4 The Combination of random utility models (CoRUM) as a unified framework 67

Chapter 3 : CoRUM for taste heterogeneity 70

3.1 Background and motivation: the Mixed GEV as a practical solution to joint EC/RC Mixed Logit model 70

3.2 Mixed RUMs: notation and formulation 73

3.3 The Combination of RUMs (CoRUM): a generalization for taste heterogeneity 75

3.4 Application on real data 78

3.4.1 Stated preference survey 78

3.4.2 Utilities specification 78

3.4.3 Models error structure 79

3.4.4 Estimation results 82

3.5 Conclusions and future research 83

Appendix 3/A: Estimation results on real data 85

Chapter 4 : Random utility models: regression vs forecasting 103

4.1 Closed-form R.U.M. and forecasting 103

4.2 Model's elasticities 104

4.3 Experimental analysis 105

4.3.1 Experiment setting 105

4.3.2 Specification of the structure of the tested models 106

4.3.3 Experimental results 106

4.4 Conclusions and future steps 114

Chapter 5 : The CoNL route choice model 115

5.1 Background and motivation 115

5.2 The CoRUM model and its CoNL particularization 117

5.3 A CoNL specification for route choice modelling 118

5.3.1 CoNL route choice: model structure 120

5.3.2 CoNL route choice: model parameters 122

5.4 The CoNL route choice model with implicit probability statement 124

5.4.1 Specification of mixing components and model structure 124

5.4.1.1 Double-step algorithm 125

5.4.1.2 Single-step algorithm 128

5.4.2 Calculation of model parameters 129

5.5 Experimental analysis 130

5.5.1 Four links-three routes network 130

5.5.2 Braess' network 132

5.5.3 Mesh network with long bypass 134

5.5.4 Grid network ... *137*
5.5.5 Sioux-Falls network ... *137*
5.5.6 Summary of experimental analysis ... *140*
5.6 Conclusions and research prospects ... 140
Chapter 6 : Some advance on CoNL route choice model ... 142
6.1 Route choice modelling issues – a brief recap ... 142
6.1.1 The computational problem in route choice: implicit enumeration algorithms ... *142*
6.1.2 The overlapping problem in route choice ... *144*
6.2 An implicit enumeration algorithm for CoNL route choice ... 146
6.2.1 CoNL recursive equations ... *146*
6.2.2 Specification of mixing components algorithm without explicit enumeration ... *149*
6.2.3 SNL CoNL algorithm ... *150*
6.2.4 Simplified SNL CoNL algorithm ... *152*
6.2.5 Experimental results ... *153*
6.3 An in depth analysis of CoNL route choice and Daganzo and Sheffi correlations ... 159
6.4 Experimental results on real data with unrestricted acyclic choice set ... 168
6.5 Conclusions and future steps ... 170
Chapter 7 : Conclusions of the thesis ... 172
Contributi a questo lavoro ... 176
Ringraziamenti personali ... 178
References ... 179

List of figures

Figure 3.1: Model estimation on SP survey for six alternatives mode choice Naples-Milan - Nested Logit specifications for error structure. 80
Figure 3.2: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Cross Nested Logit specifications for error structure. 81
Figure 4.1: Model's specification for three alternatives-context........ 108
Figure 4.2: Model's specification for four alternatives-context. 108
Figure 4.3: 0.95-0-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3)........ 109
Figure 4.4. 0.5-0.5-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3)........ 109
Figure 4.5: 0.7-0.7-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3)........ 110
Figure 4.6: 0.3-0.3-0.1 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3). 110
Figure 4.7: 0.5-03-0.3 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3). 111
Figure 4.8: 0.6-0.3-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3)........ 111
Figure 4.9: 0.8-0.4-0.2 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3). 112
Figure 4.10: 0.9-0.7-0.33 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3). 112
Figure 4.11: 0.95-0-0-0.3-0-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3). 113
Figure 4.12: 0.9-0.4-0.5-0.3-0.2-0.7 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3). 113
Figure 5.1: Topology of a toy network to showcase throughout the chapter the definitions of the proposed CoNL route choice model........ 120
Figure 5.2: Example of CoNL structure for the o-d pair **1-4** in the network in Figure 5.1........ 121
Figure 5.3: Efficient choice sets with respect to the o-d pair **1-4** in the example of Figure 5.1. 125
Figure 5.4: CoNL specification provided by the double-step algorithm for the o-d pair **1-4** in the network in Figure 5.1........ 128
Figure 5.5: Daganzo and Sheffi (1977) test network: topology and link cost structure. 130
Figure 5.6: Daganzo and Sheffi (1977) network: CoNL network levels for o-d pair **1-3** (top) and CoNL structure for destinations 2 and 3, given origin 1 (bottom). 131
Figure 5.7: Daganzo and Sheffi (1977) network: network levels and mixing components resulting from the application of the single-step illustrated in Section 5.4.1.2........ 131
Figure 5.8: Braess' network: link costs........ 132

Figure 5.9: Braess' network: CoNL structure for o=1 and d=2,3,4 respectively. 133
Figure 5.10: Mesh network: topology and link costs. 135
Figure 5.11: CoNL mixing components and corresponding nesting structure (elemental alternatives not illustrated for the sake of brevity) for the network of Figure 5.10, o-d pair **1-12** 135
Figure 5.12: Sioux-Falls network, South Dakota. 138
Figure 6.1: Regione Campania zoning (TransCad screenshot). 168
Figure 6.2: Grid network (Figure 5.1): CoNL specification for o-d pair **1-4** with the exhaustive acyclic choice-set. 169
Figure 6.3: Screenshot of the SNL software developed in Matlab R2018a. 170

List of tables

Table 2.1: GEV generating functions for well-known GEV models. 23
Table 3.1: Model estimation on SP survey for six alternatives mode choice Naples-Milan - Attributes for the six alternatives choice scenario. 79
Table 3.2: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Multinomial Logit and Nested Logit estimation results. 86
Table 3.3: Model estimation on SP survey for six alternatives mode choice Naples-Milan – CoNL estimation results. 88
Table 3.4: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Cross Nested Logit estimation results. 92
Table 3.5: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Mixed Logit estimation results. 95
Table 3.6: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Mixed CoNL estimation results. 97
Table 3.7: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Mixed CNL estimation results. 102
Table 4.1: Experimental setting. 107
Table 5.1:– Illustration of the iterations of the algorithm described in Section 5.1.1 for the o-d pair **1-4** and the efficient routes depicted in Figure 5.3. 128
Table 5.2: Daganzo and Sheffi (1977) network: route choice probabilities for the o-d pair **1-2**, under the link cost configuration $c \rightarrow 10$, $h \rightarrow 0$ and $k=1$ 132
Table 5.3: Braess' network: route choice probabilities for the o-d pair **1-4**. 134
Table 5.4: Route choice probabilities for the o-d pair **1-12**, on mesh network with long bypass in Figure 5.10. 136
Table 5.5: Route choice probabilities for the network in Figure 5.1, o-d pair **1-4**, with the link cost configuration $c_{1-2}=c_{8-4}=3$, $c_{7-3}=2$ and 1 for all remaining links. 137
Table 5.6: Route choice probabilities for the Sioux-Falls network of Figure 5.12, o-d pair **1-15.** 139
Table 6.1: SNL CoNL algorithm without explicit enumeration of routes. 152
Table 6.2: Simplified SNL CoNL algorithm without explicit enumeration of routes. 153
Table 6.3: Braess network, o-d pair **1-4** – Comparison of link flows. 155
Table 6.4: Mesh grid network with bypass, o-d pair **1-12** – Comparison of link flows. 156
Table 6.5: Sioux-Falls network, o-d pair **1-15** – Comparison of link flows. 158
Table 6.6: Sioux-Falls network – Computation time. 159
Table 6.7: Braess network – Target correlation matrix and reproduced correlation values for the compared route choice models. 162
Table 6.8: Mesh network with bypass, o-d pair **1-12**: ID of the routes. 163
Table 6.9: Sioux-Falls network, o-d pair **1-15**: ID of the routes. 163

Table 6.10: Mesh network with bypass: target Daganzo and Sheffi correlation matrix for routes connecting the o-d pair **1-12**. 164
Table 6.11: Sioux-Falls network: target Daganzo and Sheffi correlation matrix for routes connecting the o-d pair **1-15**. 165
Table 6.12: Mesh network with bypass, o-d pair 1-12 –MSE indicator for reproduced correlation values. 166
Table 6.13: Sioux-Falls network, o-d pair 1-15 - MSE indicator for reproduced correlation values. 167
Table 6.14: Log-likelihood computation for the compared models. 170

Chapter 1: Introduction

1.1 Overview of choice modelling issues

People make choices every day. Many choices have a strong impact on the quality or their life. Each day a person wakes up and chooses which action he[1] wants to do before, what to have for breakfast, what to wear, what time to go outside, how to manage his/her day by virtue of the budget and time constraints, which place to move to and how, which activity to do, which one to do before or after and so forth. There are choices that are not made every day, but they have a strong impact on the decision maker's well-being. In fact, sooner or later, a person will decide his household location, whether to own a car, the typology and the vehicle model, whether to own a pet, which breed or size, how many children to have, in which school register them and much more.

Some of the above-mentioned choice examples involve mobility. Thus, it is easy to recognize why these kind of choices form the basis for the planning and policy actions in the transport field. What is called, at aggregate level, *congestion* or *traffic*, represents the sum of individual choices that everyone makes at different levels: do I move? What time do I move? Where I want to go? Which transport mode do I want to use? What itinerary do I travel?

This kind of choices, that can be termed *transport choices,* relating the so-called *travel behaviour*, are characterized by a significant modelling complexity. If, for example, an analyst wants to model the choice of the place to go (*destination choice*), this means that he should, in order: identify the set of all the feasible destinations, understand which sub-set a decision maker actually considers and consider the factors influencing the choices of a place rather than another one. Modelling the route choice means: representing all the elemental phases to make a trip, taking into account the significant quantity of alternatives depending on them, taking into account the similarities between the alternatives and how the perception of the choice set is influenced by these similarities, understanding which routes are more relevant (fast routes, highways, no lights, sightseeing roads), just to name a few aspects.

Many typologies of choices can be involved. Some of them are discrete, i.e. made in a finite choice set (e.g., transport mode to be used), while some of them are continuous, i.e. made on an infinite choice-set (e.g., time of departure to go outside). Moreover, the strategies of the decision makers could be different when choosing an alternative: maximizing the benefit, minimizing the undesirable effects, minimizing the risks, choosing the first alternative that allows for a minimum degree of satisfaction, choosing by thinking of collective benefits or on the basis of other people choices and so forth.

The *random utility theory* represents the most widely used paradigm in modelling the behaviour of people who make choices. It starts form the *homo-economicus* assumption, which states that the decision makers want to maximize their benefit when they choose a certain alternative. Quoting Simon (1978):

[1] In the current work, the decision maker and the analyst will be indicated always with "he".

"The rational man of economics is a maximizer, who will settle for nothing less than the best."

The *random utility models* (*RUMs*[2]) are discrete econometric models. The term "econometric" refers to the fact that they are generally used downstream of having performed a regression on real data. The term "discrete", instead, indicates that their outcomes (the dependent variable) assume a finite number of values. Going back to Taussig (1912), as reported in McFadden (2001):

"An object can have no value unless it has utility. No one will give anything for an article unless it yield him satisfaction. Doubtless people are sometimes foolish, and buy things, as children do, to please a moment's fancy; but at least they think at the moment that there is a wish to be gratified."

Practically, the RUMs are mathematical models simulating the behaviour of a decision maker, whose choice set is finite, whose alternatives are distinct from each other, by means of an entity called *perceived utility*. This perceived utility, unknown to the analyst, can be represented as a random variable. This randomness, according to the econometric approach, is given by a lack of knowledge that an analyst has when modelling people's behaviour. According to the psychometric approach, instead, the randomness is inherent in the decision maker.
Generally, such perceived utility is defined as a function of observable quantities (*attribute/regressors*), and the choices are reproduced (with some model errors) as a non-linear function of such utilities. In other words, it is impossible to reproduce deterministically a value of 1 or 0, respectively, if the decision maker chooses or does not choose the given alternative. Thus, an intermediate value can be reproduced, which can be interpreted as the probability of choosing the given alternative. Such a choice probability operationally needs the density function of the perceived utilities, and therefore, in general, involves some multi-dimensional integral computation. Sometimes (as deepened in Chapter 2) the choice probabilities can be computed with an analytical closed-form, while sometimes not and, therefore, the integral form stays, requiring simulation.
The issues that an analyst must face when he models choice probabilities are multiple. Mainly, the choice criteria could be different among the individuals (*decision rule*). The choice set could be different (*choice set formation*). The relevant quantities playing a role or, in other words, the utility functions, could be different (*relevant attributes*). Each measure of the observable quantities can be affected by some errors (*error measures*). The tastes, i.e. the importance that the decision makers attach to a certain attribute, can be different (*inter-respondent taste heterogeneity*), just like for the same decision maker they could vary over time or across different choice scenarios (*intra-respondent taste heterogeneity).* Two or more alternatives can be perceived as very similar and thus not perfectly disjoint (*inter-correlation among alternatives* or *substitution pattern).* For some alternatives, the perception of the utility could be affected by more randomness (*heteroskedasticity*). There may be more random decision makers than others (*scale heterogeneity*). There can be factors that are latent but they influence the choices (*latent constructs*). The past (*history*), the personal inclinations (*attitudes*) and the habits to choose always a given alternative only because previously made (*inertia*) can strongly influence the current choices. Some unobservable components can influence the observable quantities (*endogeneity*). When an

[2] Sometimes the acronym RUM means Random Utility Maximization, and the relative models are indicated with RUM models.

individual states a choice on some hypothetical scenarios, he/she could make some errors due to the fact he/she has a biased perception of the scenario itself (*stated preference effect*).
For facing these modelling issues, the analyst needs models that are theoretically robust, i.e. capable to handle different problems in different situations. However, such models must be also operational. The theoretical background of a model is essentially represented by the stochastic assumptions made on the random part of it, that is on the perceived utilities. The fact of being operational, on the other hand, depends on the degree of simplicity, intuitiveness and computational cost necessary to work with it.
In this regard, obtaining a complex model as a mere combination of easier component models, may represent a good answer to these two questions. In fact, if the constituent bricks are chosen appropriately, combining models allows leading to a more general theoretical model. Moreover, it can help to better understand the complex model obtained and, consequently, to make it more easily operational.

1.2 The CoRUM basic idea

The conceptual-vehicle of the *Combination of random utility models* (*CoRUM*; Papola, 2016) is just the one described in the premise, i.e. obtaining a complex model by combining easier models. The idea is to fill the gaps of the existent models, creating a simple way for unifying their advantages. Combining partially general models could allow obtaining a model that is more general. This is the idea behind the CoNL model, a particular specification of the CoRUM, investigated by Papola (2016). The target of the CoNL is mainly reproducing the inter-correlation effects among utilities of the alternatives in a flexible way. Furthermore, the CoNL has the computational advantage of being a closed form models, thus not requiring simulation. It has been tested in a transport mode-choice survey on a dataset of about 1000 individual choices in a seven alternatives scenario. The seven alternatives were: car, high-speed train (1st and 2nd class), Eurostar train (1st and 2nd class) and Intercity train (1st and 2nd class). This was a typical context wherein the inter-correlations among the alternatives have an high impact on the choice probabilities. In fact, six alternatives of the total seven were train alternatives, divided into 3 different kinds of train, each one with its 1st and 2nd class options. The proposed CoRUM showed both the best goodness of fit and the best prediction results, compared with other complex models available in literature.
However, the CoRUM framework seems much more general and it potentially allows accommodating other crucial choice modelling issues. To date, in fact, there are some complex general models (see Mixed Logit; McFadden and Train, 2000) that require multi-dimensional integral simulations. But, by now, the latter do not scare when the integral simulation is performed on a few dimensions. However, in order to be more flexible, they need to add many dimensions of integration, with a prohibitive computational burden. In addition, there are other difficulties in estimating them, especially when using classic gradient-based methods. Under the CoRUM framework, instead, this generality can be achieved, for example, by combining a basic complex model formulation with other easier models formulations.
Another relevant question arises when using complex models, I whatever way they are obtained. How does their adding complexity reflect on the forecasting capability? Can we measure the danger of creating an instrument very capable of fitting real observations but incapable to make correct predictions? This is an interesting aspect to be investigated, to better understand the trade-off between adding complexity and effectiveness in different choice scenarios.

Furthermore, with reference to specific problems of travel demand estimation, the advantage of such a general framework can be used for solve some important theoretical and operative problems. For example, the route choice modelling have some peculiarities. The similarities among alternative routes, given by their physical overlapping, surely represents one of these. This similarity generates a positive correlation between their perceived utilities. Moreover, the number of routes generally involved is very high, if we compare it with other choice contexts, given the size of a real-world network. That means the correlation scenario can be very complex, thereby requiring a model with robust theoretical assumptions. The CoRUM, particularly the aforementioned CoNL specification, can be the answer to this complex operational question.

1.3 Objectives and contributions of the thesis

This thesis, as far as the general discrete choice theory is concerned, focuses on some of the choice modelling problems mentioned at the end of Section 1.1. Particularly, these problems are the substitution patterns among the alternatives, the inter/intra-respondent taste heterogeneity and the heteroskedasticity. The latter have certainly received the most attention in the discrete choice literature over the last five decades. Particularly, in the current thesis, concerning the inter-intra respondent taste heterogeneity and heteroskedasticity problems, a new model formulation has been proposed, aiming at providing a more practical formulation than that currently implemented in the software and analysed in the literature, obtained by a generalization of the current CoRUM framework (Chapter 3). The inter-correlations impact, on the other hand, has been analysed both in terms of ability to reproduce observed choices and in terms of predictive unbiasedness, analysing the main closed-form models of the literature, and contrasting them with the CoNL model (Chapter 4). These proposed advances are general and crosscutting to several other scientific fields than transport. The main application of Chapter 3, however, refers to a typical inter-urban transport mode choice. Furthermore, the formulations proposed in Chapter 3 could be easily introduced as kernel within other theoretical paradigms. In fact, sometimes the framework of the mathematical models used under different discrete choice theories can be equal (for example, the random utility models and the random regret minimization models have the same structure, but their arguments are different: in the first case is the perceived utility, to be maximized, in the second case is a combination of unpleasant effects, to be minimized).
The other advances proposed in the current thesis, on the other hand, are transport field-specific. The main transport problem addressed in the current thesis relates the route choice. In fact, as described in Section 2.3, the route choice models represent the core of traffic assignment problem and, therefore, of the traffic simulation. It surely represents the more demanding *choice dimension* (Cascetta, 2009) to be modelled, given its inherent complexity. A first problem comes from the size of a real-world network and the number of feasible alternatives that a decision maker could consider. A second problem is how the decision maker considers the alternatives that are highly overlapped and what is the impact on the probability that he chooses one of them[3]. The latter is a very challenging theoretical problem. Several existing models, with different basic assumptions, try to address it. Some of them consider the overlapping problem roughly, but they have proven to fail in several contexts (see Section 2.3.2 for relevant papers quoting this aspect). Other models address it with specific theoretical assumptions, but they can be very difficult to be used, given their computational complexity.

[3] Other route choice modelling issues not directly addressed in the current thesis will be briefly recalled in Section 2.3.5.

Consequently, in order to make them operational, it is necessary to formulate simplified hypothesis, which, however, imply the introduction of a bias in the choice probabilities they compute. This thesis proposes a new route choice model, namely the CoNL route choice model (Chapter 5), which allows very complex correlation patterns. The model is made operative by means of opportune algorithms that allow it to work on any network, without requiring neither the estimation of a large number of parameters, nor simulation procedures. Chapter 6 goes further, providing some advance on route choice modelling and, particularly, on the novel CoNL route choice model, allowing implementing an algorithm, that is theoretically consistent with the CoNL route choice model, for computing traffic flows, without the need to explicitly consider the routes, thus avoiding the operation computationally more onerous at stake. This and other algorithmic advances are proposed and tested on various aspects: choice probabilities, correlations reproduced and computed traffic flows. Several applications are presented both on toy networks and on a real big network (more than 500.000 road links), to show the goodness of fit measures of the CoNL model with reference to models of literature that are mainly implemented within the commercial software.

1.4 Thesis canvas

The thesis is structured as follows:

- Chapter 2 reviews the state of the art on random utility theory and its application to route choice. In particular, the Section 2.1 provides the basic setup for the description of RUMs; Section 2.2 reviews the random utility models available in the literature, with reference to the two main problems of the error structure (inter-correlations and heteroskedasticity problems) and the inter/intra-respondent taste variation; Section 2.3 briefly summarizes the main applications of the random utility theory to the route choice problem; Section 2.4 describes the main assumptions of the Combination of random utility models (CoRUM) as a general framework for modelling discrete choices, with particular reference to travel choices.

- Chapter 3 investigates more general specifications of the CoRUM than those previously analysed, allowing accommodating also the taste heterogeneity and the heteroscedasticity, in particular by combining mixtures of RUMs. To this end, the chapter proposes a theoretical generalization of the CoRUM framework and a real-world application on data collected on a stated survey of 1688 observations of 211 respondents. This is a forthcoming paper named:

  Tinessa, F., Marzano, V., Papola, A., (forthcoming). "*Combination of random utility models for accommodating taste variation and flexible substitution patterns*".

- Chapter 4 represents an estimation exercise with applications on future scenarios on the main closed form random utility models, on synthetic datasets with variable sample sizes and complex underlying correlation scenarios. Such correlation scenarios, on the other hand, can be representative of typical mode choice or route choice contexts. The aim of this chapter is investigating the potential of the CoNL (and the other models) in terms of forecasting, and comparing it with the models goodness of fit performances. The chapter entirely reprises and extends the set of experiments of the work published as:

Tinessa, F., Papola, A., Marzano, V., 2017. *"The importance of choosing appropriate random utility models in complex choice contexts"*, M.T.-I.T.S. Conference Proceedings, 26-28 June 2017, Naples.

- Chapter 5 proposes a new route choice model obtained under the CoRUM framework. It describes an algorithm to generate a CoNL specification, allowing detecting a set and a composition for the components of the model, and a way to compute all the structural model parameters, whatever the network. It has been published as:

  Papola A., Tinessa F., Marzano V., 2018. *"Application of the Combination of Random Utility Models (CoRUM) to route choice"*, Transportation Research Part B (111), 304-326.

- Chapter 6 is currently an original contribution of this thesis and describes several advance compared to the published work in Chapter 5. In particular, an implicit enumeration algorithm theoretically consistent with the CoRUM route choice model,is proposed and tested on toy networks; an in-depth analysis of the complex route choice models is carried out on their ability to reproduce complex correlation scenarios, drawing important conclusions, both theoretical and applicative, on the novel CoNL route choice model, proposed in Chapter 5, and on the existent Link Nested Logit model; some practical advance on the original route choice model is proposed and tested both on toy networks and on a real network (Region Campania network). The goodness of fit of the CoNL route choice has been analysed and compared with the one of the other route choice models, using real observations collected by means of GPS detection of about 200 trajectories.

- Chapter 7 reports a summary of the conclusions reached in the whole thesis and proposes several ideas for future research steps.

Chapter 2: State of the Art

2.1 Random utility theory – general assumptions

The choice modelling discipline is very broad and it embraces several sub-areas. The current thesis work focuses on *random utility theory*. The random utility models (RUMs) start from neoclassic micro-economic assumption that a decision maker searches for the maximum benefit in his choices. Particularly, it is assumed that the decision maker is rational, he considers a choice-set of alternatives and chooses the one that maximized his perceived utility. The latter, in general, can be seen as a function of several quantities that are observable (sometimes measurable or, eventually, ranked or dummies) and other quantities that are not observable. Since the presence of the latter, the perceived utility is an unknown quantity for the analyst.

2.1.1 Random utility

Attempting to embrace various formalizations of the problem (McFadden, 1974; Ben-Akiva and Lerman, 1985; Swait and Bernardino, 2000; Bierlaire, 2003; Train, 2009) in a general way, this thesis starts from the assumption that the *perceived utility* of the alternative *i*, for the decision maker *n*, facing a choice scenario (or *choice task*) *t* for a number of alternatives m, can be expressed as:

$$U_i^{n,t} = U_i^{n,t}(V_i^{n,t}; \varepsilon_i^{n,t}) \quad \forall i \in \mathrm{C}^{n,t} \tag{2.1}$$

where:

- $U_i^{n,t}$ is the value of the perceived utility of the alternative *i*, for the decision maker *n*, facing the choice scenario *t*;
- $V_i^{n,t}$ is the expected value, so called *systematic utility*[4], of the perceived utility of the alternative *i*, for the decision maker *n*, facing the choice scenario *t*;
- $\varepsilon_i^{n,t}$ is the unobservable term of $U_i^{n,t}$, so called *random residual*[5] (Cascetta, 2009) or *disturbance* or *error*. This term includes, with reference to the econometric approach, all that is unknown to the analyst, while according to the psychometric approach, all that is intangible to the decision maker. The assumption is that $E[\varepsilon_i^{n,t}]=0, \forall i \in \mathrm{C}^{n,t}$ i , as in the regression models;
- $\mathrm{C}^{n,t}$ is the choice set considered by the decision maker *n*, facing the choice scenario *t*;

---

[4] Sometimes improperly called observable part of the utility, in the current work, the $V_i^{n,t}$ simply means the expected value of the $U_i^{n,t}$..

[5] Here it is assumed a different meaning for the random term $\varepsilon_i^{n,t}$, that is "less inclusive" with reference to the definitions of Train (2009) and Cascetta (2009). It does not include all the randomness inherent the behavioural analysis, because some of them will be explicitly included into the taste coefficients, for better setup the discussion of the successive topics of the thesis.

The systematic utility $V_i^{n,t}$, in turn, can be expressed as a function of quantities that are observable $X_{k,i}^{n,t}$, so called *attributes/regressors* or explanatory variables[6], and other quantities $\beta_{k,i}^{n,t}$, so called *parameters/coefficients*, representing the marginal utility for the alternative *i* with reference to the attribute *k*. In other words, the generic $\beta_{k,i}^{n,t}$ quantifies the importance that the decision maker *n* in the choice task *t* gives to the k^{th} attribute of the i^{th} alternative's perceived utility.

Consequently, the expression (2.2) can be written as:

$$U_i^{n,t} = U_i^{n,t}(X_{k,i}^{n,t};\beta_{k,i}^{n,t};\varepsilon_i^{n,t}) \ \forall i \in \mathrm{C}^{n,t} \tag{2.2}$$

2.1.2 Additive and multiplicative approaches

Two main approaches can be adopted, depending upon how the systematic term $V_i^{n,t}$ is combined with the unobservable random term $\varepsilon_i^{n,t}$:

additive R.U.M.)

$$U_i^{n,t} = V_i^{n,t}(X_{k,i}^{n,t};\beta_{k,i}^{n,t}) + \varepsilon_i^{n,t} \ \forall i \in \mathrm{C}^{n,t} \tag{2.3}$$

Wherein the random residuals $\varepsilon_i^{n,t}$ are adding noises with reference to the $V_i^{n,t}$.

multiplicative R.U.M.)

$$U_i^{n,t} = V_i^{n,t}(X_{k,i}^{n,t};\beta_{k,i}^{n,t}) \cdot \varepsilon_i^{n,t} \ \forall i \in \mathrm{C}^{n,t} \tag{2.4}$$

Wherein the random residuals $\varepsilon_i^{n,t}$ are amplifiers of the disturbances as increasing functions of the expected values of the perceived utilities $V_i^{n,t}$. Furthermore, Dagsvik (1995) generalized the definition of random utility models with the definition of *interpersonnel RUM*.

However, the current thesis focuses on the additive random utility models (*ARUMs*). Particularly, a linear in parameters specification of the systematic utility will be always assumed. In this way, the (2.3) can be expressed as:

$$U_i^{n,t} = \sum_k \beta_{k,i}^{n,t} \cdot X_{k,i}^{n,t} + \varepsilon_i^{n,t} \ \forall i \in \mathrm{C}^{n,t} \tag{2.5}$$

In the (2.5), obviously, the notation $X_{k,i}^{n,t}$ can indicate, in general, a function of observable quantities. In the following, the (2.5) will be the main vehicle to show the main differences among all the random utility models described in Section 2.2.

2.1.3 Maximizing benefit behaviour

Defining $y_i^{n,t}$ as a binary variable assuming values 0/1, respectively, if the actual choice of the individual *n* in the choice task *t* is or is not represented by the alternative *i*, what has been said in the premise of the current chapter can be easily formalized as follows:

$$y_i^{n,t} = d^{n,t}(\mathrm{U}_i^{n,t};\mathrm{C}^{n,t}) \tag{2.6}$$

where $d^{n,t}$ indicates the *decision rule*, i.e. the synthesis of all the rules according to which the decision maker *n* chooses *i* when facing *t*, in the choice-set $\mathrm{C}^{n,t}$. In the random utility theory case, $d^{n,t}$ summarizes the assumption of *homo-economicus* above described, who chooses the alternative that maximizes his/her perceived utility. Thus, the (2.6) can written more explicitly as:

[6] It will be assumed the exogeneity hypothesis, i.e. that the attributes and the random terms are stochastically independent. For a comprehensive discussion about the endogeneity problems the reader refer to Louviere et al. (2005) and Train (2009).

$$\begin{cases} y_j^{n,t} = 1 \Leftrightarrow U_j^{n,t} \geq U_i^{n,t}, \forall i \in \mathrm{C}^{n,t} \ , j \neq i \\ y_j^{n,t} = 0 \quad otherwise \end{cases} \tag{2.7}$$

Extending the concept, by suppressing the superscript *t*, we can indicate with y_i^n the result of a sequence of choices for the individual *n*.

A choice obeying the maximizing benefit behaviour must respect three intuitive axioms (Debreu, 1954; Block and Marschak, 1959):

- *Completeness*: a decision maker is always able to make a binomial choice between whatever pair of alternatives *a* and *b*. In this case, the exact term is *weak preference* between the two, because, in a repeated choice situation with the same scenario, the decision maker could actually make another choice. The *strong preference* represents the opposite concept, i.e. the decision maker do prefer *a* to *b* always if he faces the choice task infinite times.
- *Transitivity*: if a decision maker prefers *a* to *b* and *b* to *c*, he will prefer *a* to *c*;
- *Continuity*[7]: an infinitesimal variation in a relevant attribute value does not vary the preference of a certain alternative;

In the following, a general framework will be presented, i.e. the random utility theory, wherein the three above-mentioned axioms represent a pre-condition for make a model operative.

2.1.4 Random utility maximization models (RUMs)

Starting from (2.7), a generic random utility model is a mathematical relationship whose outcome is the probability of observing that a given alternative *j* is chosen by decision maker *n* in the choice task *t* within the choice set $\mathrm{C}^{n,t}$:

$$p^{n,t}(j) = \Pr\left\{U_j^{n,t} \geq U_i^{n,t}, \forall i \in \mathrm{C}^{n,t}, j \neq i\right\} = \Pr\left\{U_j^{n,t} - U_i^{n,t} \geq 0, \forall i \in \mathrm{C}^{n,t}, j \neq i\right\} \tag{2.8}$$

And given (2.3) it (2.8) can be also written as:

$$p^{n,t}(j) = \Pr\left\{\varepsilon_j^{n,t} - \varepsilon_i^{n,t} \geq V_i^{n,t} - V_j^{n,t}, \forall i \in \mathrm{C}^{n,t}, j \neq i\right\} \tag{2.9}$$

That is needed just to emphasize three relevant aspects. First, such probability depends entirely on the utilities differences, while the absolute values of the utilities do not matter. This represents a fundamental of the random utility theory, implying that the only parameters that can be estimated are the ones that allow to define the utilities differences. Second, when multiplying the utility values for a scale factor, the $p^{n,t}(j)$ does not change. This leads to another fundamental of the random utility theory, i.e. that the scale does not matter. Thus, quoting Train (2009), "*only differences in utility matter and the scale is arbitrary*". Third, the choice probability essentially depends on how the random terms within the perceived utilities is distributed. As better described in the following sub-sections, the scale is directly linked to the variance of the perceived utilities.

A different interpretation for the $p^{n,t}(j)$ can be given in a similar linear regression fashion. The probability $p^{n,t}(j)$ is the regression value of $y^{n,t}_j$ affected by an error, averagely distributed with 0 mean (Dagsvik, 2004). As it will be seen, a random utility model is effectively a non-linear regression model in regressors with a discrete outcome.

The (2.8) can be translated into an operative formulation, by assuming the vector **U** of the perceived utilities $U_i^{n,t}$ distributed with a specific density function f(**U**), as:

[7] Actually, some authors refers the third axiom as the fact the decision maker faces a finite choice set (Marzano, 2006).

$$p^{n,t}(j)=\int_{U_j^{n,t}=-\infty}^{U_j^{n,t}=-\infty}\int_{U_1^{n,t}=-\infty}^{U_1^{n,t}=U_j^{n,t}}\dots\int_{U_{m-1}^{n,t}=-\infty}^{U_{m-1}^{n,t}=U_j^{n,t}}\int_{U_m^{n,t}=-\infty}^{U_m^{n,t}=U_{j,n,t}} f(\mathbf{U})d\mathbf{U} \tag{2.10}$$

And the (2.9), indicating with $\boldsymbol{\varepsilon}$ and $\boldsymbol{\beta}$, respectively, the vectors of the $\varepsilon_i^{n,t}$'s and the $\beta_{k,i}^{n,t}$'s, leads to:

$$p^{n,t}(j)=\int_{\varepsilon_j^{n,t}=-\infty}^{\varepsilon_j^{n,t}=+\infty}\int_{\varepsilon_1^{n,t}=-\infty}^{V_j^{n,t}-V_1^{n,t}+\varepsilon_j^{n,t}}\dots\int_{\varepsilon_{m-1}^{n,t}=-\infty}^{V_j^{n,t}-V_{m-1}^{n,t}+\varepsilon_j^{n,t}}\int_{\varepsilon_m^{n,t}=-\infty}^{V_j^{n,t}-V_m^{n,t}+\varepsilon_j^{n,t}}\int_{\boldsymbol{\beta}} f(\boldsymbol{\beta},\boldsymbol{\varepsilon})d\boldsymbol{\beta}d\boldsymbol{\varepsilon} \tag{2.11}$$

wherein f($\boldsymbol{\beta}$,$\boldsymbol{\varepsilon}$) is the joint distribution of the $\boldsymbol{\beta}$ and the $\boldsymbol{\varepsilon}$. It has been noted that in the range bounds of each integral the $V_i^{n,t}$ appears as a function of the $\beta_{k,i}^{n,t}$, under this framework assumed in Section 2.1.1. Thus, the (2.10) and (2.11) integrals are multi-dimensional, with dimension m in the space of the **U**. The same is an integral with m + n_β dimensions in the space of the $\boldsymbol{\varepsilon}$ and $\boldsymbol{\beta}$, wherein n_β represents the number of the coefficients $\beta_{k,i}^{n,t}$'s. However, they can be expressed in the space of the differences between the perceived utilities, leading to an (m-1)-dimensional integral. The same applies when working in the space of the $\boldsymbol{\varepsilon}$ and $\boldsymbol{\beta}$, leading to an (m-1+ n_β)-dimensional integral. The difference can be made with reference to anyone of the alternatives utilities/random terms taken as touchstone. For example, taking the alternative j as touchstone, indicating with $\mathbf{W}^j$ the vector of the differences $U_i^{n,t} - U_j^{n,t}$, and with $\boldsymbol{\Lambda}^j$ the vector of the differences $\varepsilon_i^{n,t} - \varepsilon_j^{n,t}$, the choice probability can be expressed as:

$$p^{n,t}(j)=\int_{U_1^{n,t}-U_j^{n,t}=-\infty}^{U_1^{n,t}-U_j^{n,t}=0}\int_{U_2^{n,t}-U_j^{n,t}=-\infty}^{U_2^{n,t}-U_j^{n,t}=0}\dots\int_{U_{m-1}^{n,t}-U_j^{n,t}=-\infty}^{U_{m-1}^{n,t}-U_j^{n,t}=0}\int_{U_m^{n,t}-U_j^{n,t}=-\infty}^{U_m^{n,t}-U_j^{n,t}=0} f(\mathbf{W}^j)d\mathbf{W}^j \tag{2.12}$$

$$p^{n,t}(j)=\int_{\varepsilon_1^{n,t}-\varepsilon_j^{n,t}=-\infty}^{\varepsilon_1^{n,t}-\varepsilon_j^{n,t}=V_j^{n,t}-V_1^{n,t}}\int_{\varepsilon_2^{n,t}-\varepsilon_j^{n,t}=-\infty}^{\varepsilon_2^{n,t}-\varepsilon_j^{n,t}=V_j^{n,t}-V_2^{n,t}}\dots\int_{\varepsilon_{m-1}^{n,t}-\varepsilon_j^{n,t}=-\infty}^{\varepsilon_{m-1}^{n,t}-\varepsilon_j^{n,t}=V_j^{n,t}-V_{m-1}^{n,t}}\int_{\varepsilon_m^{n,t}-\varepsilon_j^{n,t}=-\infty}^{\varepsilon_m^{n,t}-\varepsilon_j^{n,t}=V_j^{n,t}-V_m^{n,t}}\int_{\boldsymbol{\beta}} f(\boldsymbol{\beta},\boldsymbol{\Lambda}^j)d\boldsymbol{\beta}d\boldsymbol{\Lambda}^j \tag{2.13}$$

Another representation of choice probability is very interesting for our purposes. Quoting the introduction chapter of Train (2009), and adapting it with the introduced (2.2) (wherein, differently from Train, the marginal utilities are, in general, variable with individual *n* and choice situation *t*) consistently with the notation above defined, a binary 0/1 indicator $I_j^{n,t}$ can be defined as a function of observable and non-observable quantities, as:

$$\begin{cases} \mathrm{I}_j^{n,t}\left[h(\mathbf{X},\boldsymbol{\beta},\boldsymbol{\varepsilon})=j\right]=1 & if \quad y_j^{n,t}=1 \\ \mathrm{I}_j^{n,t}\left[h(\mathbf{X},\boldsymbol{\beta},\boldsymbol{\varepsilon})=j\right]=0 & otherwise \end{cases} \tag{2.14}$$

This means, the probability of choosing the alternative *j* can be seen as the expected value of $I_j^{n,t}$ over joint density $f(\boldsymbol{\beta},\boldsymbol{\varepsilon})$. Such expected value can be expressed as:

$$p^{n,t}(j)=\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\varepsilon}} \mathrm{I}_j^{n,t}\left[h(\mathbf{X},\boldsymbol{\beta},\boldsymbol{\varepsilon})=j\right]\cdot f(\boldsymbol{\beta},\boldsymbol{\varepsilon})\,\mathrm{d}\boldsymbol{\beta}d\boldsymbol{\varepsilon} \tag{2.15}$$

Other authors (for instance, see Train,2009) refer this formulation, giving to $\boldsymbol{\varepsilon}$ the meaning of the whole random part of the perceived utility, i.e. incorporating the random part of $\boldsymbol{\beta}$. However, assuming the current thesis framework, the assumption on the joint density is generally that $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are independent, thus giving f($\boldsymbol{\beta}$,$\boldsymbol{\varepsilon}$)=f($\boldsymbol{\beta}$)f($\boldsymbol{\varepsilon}$).

Whatever definition of $p^{n,t}(j)$ is made, it is clear that (2.15) represents a multidimensional integral, in the differences or in the single alternatives random terms space, and it has to be analytically solved or simulated. As described in Section 2.2, under opportune assumptions, the integral have a closed form solution.

The following section contains a review of the state of the art of the main random utility models analysed in literature.

2.2 State of the art of RUMs

The literature on choice modelling is very wide, being the latter a powerful transversal tool for many scientific areas. A lot of general and theoretical contributes to this discipline are trackable in economic, psychometric, transport, biological, healthcare and several others scientific fields. In fact, not infrequently a general mathematical model has been proposed to answer some very specific problem. This thesis work does not embrace all relevant issues analysed in scientific literature, but mainly focuses on some well-established lines of research. In Chapter 1, several choice modelling issues have been mentioned. Particularly, two main problems has always been recognized as very relevant, not only for transport aims the error structure assumptions on random terms and the taste heterogeneity. The proposed state of the art investigates the main random utility models, with reference to these two fundamental problems.

2.2.1 Error structure

The main assumption of the current sub-section is that marginal utilities $\beta_{j,k}^{n,t}$'s are not randomly distributed across population of individuals n and choice scenarios t, i.e. f($\boldsymbol{\beta}$) and its dimensions of integration are suppressed from (2.15). As clarified in the previous section, this is just a pedagogical assumption, to create a clear cut between the two phenomena. Actually, in real-world applications there is not the possibility to create such separation between the two.
As mentioned in Section 2.1.4, a random utility model involves some multi-dimensional integrals computations. Whenever (2.15) does not have a closed form solution, this means involving simulation.
The first random utility models family totally avoids this computational problem, by making some appropriate assumptions on f($\boldsymbol{\varepsilon}$) and, therefore, on f($\mathbf{U}$). Unfortunately, it is not always possible to resort to these easier tools, and, in general, closed form models actually exhibit some limitations. Other models that do not have closed form solution need to be specified in a way to ensure the computation of the integral is feasible. Although these models are harder to work with, the advantages, whenever possible, are not negligible. Particularly, three main problems can be identified, with reference to error structure of random terms:

- Heteroskedasticity;
- Substitution across alternatives;
- Correlation of the random terms over time;

For a more comprehensible discussion, it is better to make the more didactic, but unrealistic assumption, that parameters $\beta_{k,i}^{n,t}$ do not vary across respondents and choice situations. This problem, i.e. the second one presented in the premise of this chapter (taste heterogeneity), will be discussed in Section 2.2.2. In this sub-section, the choice to isolate the two effects has exclusively a clarifying aim. Another assumption of the section, that is actually very common in literature, is that the error structure of the $\varepsilon_j^{n,t}$'s is keep fixed across the observations. The reader can refer Swait and Bernardino (2000) for a relaxation of this assumption in a three tran-oceanic air travels choice context.

2.2.1.1 GEV models

The GEV family (Mac Fadden, 1978) represents a wide class of random utility models. The underlying assumption on random residuals is that they are distributed as Multivariate extreme value (MEV). Although chronologically this brilliant generalization has come after the existence of the simplest and best-known GEV family exponent, i.e. the Multinomial Logit (Luce, 1959), the GEV family will be presented at first, and the main GEV models will be summarized in the following sub-sections, briefly describing their properties and limitations.
Specifically, a GEV model can be always represented in this way:

$$p^{n,t}(j) = \frac{\exp(V_j^{n,t})}{\mu \cdot G^{n,t}} \cdot \frac{\partial G^{n,t}}{\partial \exp(V_j^{n,t})} = \frac{Y_j^{n,t}}{\mu \cdot G^{n,t}} \cdot G_j^{n,t} \tag{2.16}$$

Wherein G is the so termed *GEV generating function* and, apart from superscripts n and t, G_j represents its first order derivative with reference to Y_j, in turn defined as the exponential of systematic utility of j, while m is a constant defining the degree of homogeneity of the function G. The latter is defined *strict utility function* in psychologists' literature (Domencich and McFadden, 1975). Thus, the GEV family models represent a wide class of closed form models, whose underlying distribution and the choice probabilities (2.16), are defined by assuming an appropriate formulation for the function G.
McFadden (1978) and successive generalizations (Ben-Akiva and Francois, 1983) exposed the main properties of the GEV generating function G, to give a GEV model, as follows:

1) Non-negativity ∀ $V_j^{n,t}$, i.e. ∀ positive value of $Y_j^{n,t}$ (Daly and Bierlaire, 2006);
2) Homogeneity of degree $\mu>0$;
3) Positive divergence when each $Y_j^{n,t}$. goes to $+\infty$;
4) Non negativity of kth cross partial derivatives in $Y_i^{n,t}$, with $j\neq i$, when k is odd number;
5) Non positivity of kth cross partial derivatives in $Y_i^{n,t}$, with $j\neq i$, when k is even number;

It is possible to demonstrate (Cascetta, 2006), that:

- a cumulative distribution function of residuals given by $F(\boldsymbol{\varepsilon}) = \exp(-G)$, where F is a function of the residuals as $G = G(\exp(-\boldsymbol{\varepsilon}))$, is MEV distributed;
- the mean of maximum between perceived utility values is equal to $E\left\{\max_{j\in C^{n,t}}[U_j^{n,t}]\right\} = \frac{1}{\mu} \cdot \left[\ln(G^{n,t}) + \gamma\right]$ where γ is the Euler-Mascheroni's constant (0.57721566…);

and, obviously, the (2.16).
Dagsvik (1994) analysed the generality of the GEV models, indicating that a GEV model could theoretically mimic effectively every possible random utility structure.
Based on the above mentioned properties, several models will be presented. The assumptions on G for the main models is reported in Table 2.1. All indexes involved in the formulations will be discussed in the following sub-sections.

Multinomial Logit

Multinomial Logit, as mentioned above, represents the first GEV family member, but its birth comes before the GEV class birth. Luce (1959) was the first to propose the following Logit formula:

Multinomial Logit)	$G^{n,t} = \sum_{i \in C^{n,t}} \exp(\frac{V_i^{n,t}}{\theta})$	(2.17)
Nested Logit)	$G^{n,t} = \sum_{k \in K} \left[\sum_{i \in k} \exp(\frac{V_i^{n,t}}{\theta_k}) \right]^{\frac{\theta_k}{\theta}}$	(2.18)
Cross Nested Logit)	$G^{n,t} = \sum_{k \in K} \left[\sum_{i \in k} \alpha_{ik}^{\left(\frac{\theta}{\theta_k}\right)} \exp(\frac{V_i^{n,t}}{\theta_k}) \right]^{\frac{\theta_k}{\theta}}$	(2.19)
Pair Combinatorial Logit)	$G^{n,t} = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \left[\alpha_{i,ij}^{\left(\frac{\theta}{\theta_{ij}}\right)} \exp(\frac{V_i^{n,t}}{\theta_{ij}}) + \alpha_{j,ij}^{\left(\frac{\theta}{\theta_{ij}}\right)} \exp(\frac{V_j^{n,t}}{\theta_{ij}}) \right]^{\frac{\theta_{ij}}{\theta}}$	(2.20)
FinMix)	$G^{n,t} = \sum_{c \in C} \left(G^{c,n,t} \right)^{\frac{\theta_c}{\theta}}$	(2.21)

Table 2.1: GEV generating functions for well-known GEV models.

$$p^{n,t}(j) = \frac{\exp(\frac{V_j^{n,t}}{\theta})}{\sum_{i \in C^{n,t}} \exp(\frac{V_i^{n,t}}{\theta})} \tag{2.22}$$

Wherein θ is the inverse of the so-called scale parameter. In additive models, the scale of systematic utilities does not affect the choice probabilities. Luce started from the assumption of *Independence of irrelevant alternatives*, or, as more intuitively called by Block and Marschak (1959), *the independence from added alternatives*. Actually, Luce saw this property as desirable, expressing the following concept: if a new alternative *k* is added to the choice set, the ratio between the choice probabilities of any pair of alternatives *j* and *s* does not change. In fact, according to (2.22), the ratio would be:

$$\frac{p^{n,t}(j)}{p^{n,t}(i)} = \exp(\frac{V_j^{n,t} - V_i^{n,t}}{\theta}) \tag{2.23}$$

i.e. it is not affected at all by the introduction of *k*. This ensures the consistence with the maximizing utility axioms (Marschack, 1960). Marley (Luce and Suppes, 1965) demonstrated that the (2.22) descends from an i.i.d. Gumbel assumption on random residuals. McFadden (1974) demonstrated the Logit formula necessarily implies extreme value distribution for random residuals. Thus, the i.i.d. Gumbel assumption is both necessary and sufficient for (2.22). Recently Ye et al. (2017) proposed a practical test for the validity of the Gumbel assumption when estimating a MNL on a given sample. This is consistent with (2.17), because:

$$F(\boldsymbol{\varepsilon}) = \exp(-G^{n,t}) = \exp\left[-\sum_{i \in C^{n,t}} \exp(\frac{-\varepsilon_i^{n,t}}{\theta}) \right] = \prod_{i \in C^{n,t}} \exp\left[-\exp(\frac{-\varepsilon_i^{n,t}}{\theta}) \right] \tag{2.24}$$

that represents the product of identical Gumbel distributions for each ε_i. This means the Multinomial Logit assumption is that residuals are independently distributed as Gumbel with the same variance parameter θ. More succinctly, this is well-known in literature as the i.i.d. (independently and identically distributed) Gumbel assumption on random residuals. Being θ

the same for all alternatives, the model (as all GEV models) is homoscedastic. The variance parameter is the inverse of the scale and it flattens the differences between choice probabilities when it rises. Conversely, when it decreases, it emphasizes the choice probability of the maximum systematic utility alternative, that goes to 1. On the other hand, an infinite dispersion means null capacity of the analyst to explain the phenomena, whilst 0 variance means perfect knowledge, i.e. a deterministic scenario.

However, (2.23) actually represents the main limitation of Multinomial Logit. In fact, (2.23) implies proportional substitution across alternatives. In other words, when a $V_j^{n,t}$ rises/decreases is immediate to recognize that the choice probabilities of *j* rises / decreases and the other choice probabilities decrease/rise the same relative quantity. This has potentially undesirable forecasting implications, because several could be the situations wherein the market share choosing an alternative could be penalized more than the one of another alternative. Famous example of the counter-intuitiveness of this result are reported in Chipman (1960), Debreu (1960), Daganzo and Sheffi (1977), Ortuzar (1983), Brownstone and Train (1999), Cascetta (2006), Train (2009). The most known are maybe the red bus/blue bus problem and the Daganzo network.

Thus, Multinomial Logit, as evident by the i.i.d. assumption (Gumbel is not relevant), does not allow for heteroscedasticity and inter-correlations between alternatives or random terms over time, i.e. none of the three problems mentioned in the premise of Section 2.2.1. A different approach for generalizing the Multinomial Logit model, whilst not renouncing to a closed-form expression for the choice probability, is represented by the *Mother* (or *Universal*) *Logit model* (McFadden, 1975). The idea is putting, in some ways, the observable attributes of the other alternatives within the utility of the examined alternative[8]. Thus, no different assumptions on random residuals are made, but, substantially, a different utility specification within the observable part. Examples of this framework are, with various declinations, the Dogit model (Gaundry and Dagenais, 1978), the Parametrized Logit Captivity model (Swait and Ben-Akiva, 1987), but, above all, the C-Logit (Cascetta et al., 1996) and the Path-Size Logit model (Ben-Akiva and Ramming, 1998). The last two, in particular, are route choice models, and they will be described in Section 2.3.2.2 and 2.3.2.3.

For an heteroskedastic formulation of Logit the reader can refer to Bhat(1995), who hypothesized a multivariate Gumbel distribution with different variance parameters for the alternative utilities. This leads to a formulation without closed form. It has received no interest in the practical applications given the successive strong development of the more general Mixed Logit model (see Sections 2.2.1.3 and 2.2.2.2). Other formulations that have not received an interest in the literature is the Negative Exponential model (Daganzo, 1978a) and the Weibull model (Beilner and Jacobs, 1971), both described in the book of Daganzo (1979). A model with an underlying Weibull assumption will be deepened in Section 2.3.3.4 for route choice applications, under some specific assumptions.

The strong limitations of Logit model, although its appealing great simplicity and the incomparable computational efficiency of its parameters estimation, has driven the researchers to move towards other more complex formulations. Although there are many situations wherein, with a good utility specification, it is possible to reduce the undesirable effects of Logit (it has been reminded that all the problems mentioned regards the unobservable components

[8] This idea is very similar to the one of the Random Regret minimization (RRM) theory, as described in Mai (2017) for route choice. However, in the RRM, the decision maker is assumed to choose the alternative who minimizes the perceived regret, rather than maximizing the perceived utility.

of the perceived utilities), unfortunately, in the practical applications, often it could be convenient to opt for other models.

Nested Logit

The Nested Logit model (partially) relaxes the assumption of independence of random residuals of alternatives, allowing for correlations among some groups of alternatives. Practically, the alternatives are clustered into groups k, called nests, identifying positively correlated alternatives. Also the Nested Logit formulation comes before GEV family and its GEV membership demonstration. Several authors (Ben-Akiva, 1974; Domencich and McFadden, 1975; Williams, 1977; Daly and Zachary, 1978), probably independently, proposed the following framework for choice probabilities, in an intuitive conditional form:

$$p^{n,t}(j) = p^{n,t}(j/k) \cdot p^{n,t}(k) \tag{2.25}$$

For creating a correlation within a nest of alternatives, the generic random residual can be viewed as portioned in two adding i.i.d. Gumbel random residuals:

$$\varepsilon_j^{\;n,t} = \varepsilon_k^{\;n,t} + \varepsilon_{j/k}^{\;\;n,t} \tag{2.26}$$

The part $\varepsilon_k^{n,t}$ represents the part that is shared among the correlated utilities. In fact, the correlation is the tendency of two random variables to go in the same direction with reference to their average values. Sharing a random portion with the same sign, a positive correlation is induced. In other words, two utilities are positively correlated if an unobservable factor influences both of them.

Thus, the integral (2.15) involves the joint distribution f($\varepsilon_k^{n,t}$, $\varepsilon_{j/k}^{n,t}$), represented by the product between f($\varepsilon_k^{n,t}$) and f($\varepsilon_{j/k}^{n,t}$), being the two random portions independent one from the other. Therefore, the integral in the second part clearly becomes the Multinomial Logit formula applied to alternatives within the nests. In other words, the first term of the (2.25) is the Logit formula involving all alternatives $a \in k$, within which in the denominator of the exponential's argument there is the variance of the parameter of the $\varepsilon_{j/k}^{n,t}$. It is clear to recognize that the variance of the total random residuals is given by the sum of the variances of the two portions in (2.26)(the covariance is null), so, the variance within k will be smaller than the total variance. This means θ_k is smaller than θ.

The second term is Gumbel distributed with variance given by the difference ($\theta^2 - \theta_k^2$) times $\pi^2/6$. For the nest probability, it is worth to think of the perceived utility of a group. It is defined as the maximum of utilities of alternatives within the group, so the systematic utility is defined as:

$$U_k^{n,t} = E\left[\max_{a\in k}\left(U_a^{n,t}\right)\right] = E\left[\max_{a\in k}(V_a^{n,t} + \varepsilon_{a/k}^{n,t}) + \varepsilon_k^{n,t}\right] \tag{2.27}$$

So, inasmuch as maximum of Gumbel random variables with the same θ_k is always a Gumbel with parameter θ_k and mean given by:

$$V_k^{n,t} = E\left[\max_{a\in k}(V_a^{n,t} + \varepsilon_{a/k}^{n,t})\right] = \theta_k \cdot \ln\left[\sum_{a\in k}\exp\left(\frac{V_a^{n,t}}{\theta_k}\right)\right] \tag{2.28}$$

The closed form expression (2.28) for the log-sum, according to McFadden (2001), seems due to the PhD thesis of Ben-Akiva (1972).

The $U_k^{n,t}$ is characterized by a random residual distributed as $\varepsilon_j^{n,t}$, i.e. with variance θ. Consequently, the choice probability for nest follows the Logit formula extended to all nest utilities (2.28) and variance parameter θ to the denominator. Definitely, given these premises and substituting (2.28) in a Logit for the second term of (2.25), the final choice probability for alternative *j* is given by:

$$p^{n,t}(j)=\frac{e^{V_j^{n,t}/\theta_k}}{\sum_{a\in k}e^{V_a^{n,t}/\theta_k}}\cdot\frac{\left(\sum_{a\in k}e^{V_a^{n,t}/\theta_k}\right)^{\delta_k}}{\sum_{k'\in K}\left(\sum_{a'\in k'}e^{V_{a'}^{n,t}/\theta_{k'}}\right)^{\delta_{k'}}}=\frac{e^{V_j^{n,t}/\theta_k}\cdot\left(\sum_{a\in k}e^{V_a^{n,t}/\theta_k}\right)^{\delta_k-1}}{\sum_{k'\in K}\left(\sum_{a'\in k'}e^{V_{a'}^{n,t}/\theta_{k'}}\right)^{\delta_{k'}}} \tag{2.29}$$

Wherein K is the set of the nests, *a* and *a'* represent, respectively, the alternatives within nests *k* and *k*'∈ K. The parameter δ_k is defined as the independence parameter and represents the ratio θ_k/θ. The name of this parameter derives from the fact that the correlation between a generic pair of alternatives *i* and *j* belonging to the same nest *k* can be expressed as:

$$Corr\left\{\varepsilon_i^{n,t},\varepsilon_j^{n,t}\right\}=\frac{Cov\left\{\varepsilon_i^{n,t},\varepsilon_j^{n,t}\right\}}{\sqrt{Var\left\{\varepsilon_i^{n,t}\right\}\cdot Var\left\{\varepsilon_j^{n,t}\right\}}}=\frac{Var\left\{\varepsilon_k^{n,t}\right\}}{\sqrt{Var\left\{\varepsilon_i^{n,t}\right\}\cdot Var\left\{\varepsilon_j^{n,t}\right\}}}=\frac{\frac{\pi^2}{6}\cdot\left(\theta^2-\theta^2{}_k\right)}{\sqrt{\frac{\pi^2}{6}\cdot\theta^2\cdot\frac{\pi^2}{6}\cdot\theta^2}}= \tag{2.30}$$

$$=1-\frac{\theta^2{}_k}{\theta^2}=1-\delta^2{}_k$$

That means when δ_k rises the covariance decreases and vice versa. Its values is constrained between 0, for the non-negativity of the variances, and 1, for the relationship $\theta_k\leq\theta$. This condition is always been considered as necessary to ensure the consistence with stochastic utility maximization, particularly with reference to stochastic transitivity and regularity (McFadden, 1978; Daly and Zachary, 1978; Borsch-Supan, 1990). Actually, more recently, Batley and Hess (2016) analysed this issue more in details. They showed how the lower bound, in some cases, may be greater than 0 to ensure the transitivity property, thus restricting inferiorly the interval. Furthermore, they showed the upper bound is not constrained to 1 neither by regularity nor transitivity.

The Nested Logit above described is also called Hierarchical Logit or one-level Nested Logit or, by several authors, two-levels Nested Logit, computing the nest level as a level. Ortuzar (2001) provided a good historical reconstruction of the Nested Logit with one level. More levels can be added (McFadden, 1978; Ben-Akiva and Lerman, 1985), to improve the correlation structure of the model, but not totally relaxing the independence assumption. In fact, within the nests the I.I.A. property remains and, furthermore, the property holds also between nest probabilities. Train (2009) calls the latter as *independence from irrelevant nests* (I.I.N.). However, in presence of multiple levels, the (2.25) and (2.29) can be easily extended, adding more conditional probabilities. The (2.30) can be updated, by substituting the ratio θ_k/θ with a product of ratios, each one referred to a level and its previous one (Daganzo and Kusnic, 1993). Definitely, Nested Logit is the first model that partially solves only one of the three problems above presented, i.e. it goes towards a more flexible substitution pattern amongst alternatives, but is still an homoscedastic model with limitations in reproducing covariance matrices and it does not allow for correlation over time of the residuals.

Cross Nested Logit and further generalizations

The idea of Cross Nested Logit is allowing for an alternative *j* belonging to more than one nest *k*. Therefore, there is no more clustering but a sort of fuzzy division of the alternatives in a set of nests. That means creating a greyscale membership of each alternative into the nests,

defining a positive *degree of membership* α_{jk}. Based on this conceptual vehicle, the Cross Nested Logit has been proposed under several names and with different particularisations. Maybe the first one who proposed it was Chu (1981) and, more notoriously, always Chu (1989). The model was named Pair Combinatorial Logit (PCL), and the alternatives was divided into pair nest, one for each feasible couple of alternatives in the choice set. The intuition was allowing for a correlation among each pair of alternatives. Another contribute, that generally each author refers, is the so-called Ordered GEV model (Small, 1987), wherein the peculiarity was that the alternatives were ranked (ordered) and the correlation depended on the closeness in the order. However, the first contribute to the Cross Nested Logit in the most similar form for what is known nowadays is due to Vovsha (1997). The name itself (abbreviate CNL) is certainly due to that contribute, wherein the author performed an application of the model on mode choice in Tel Aviv Metropolitan area. After this, several works have generalized the model and its properties (Ben-Akiva and Bierlaire, 1999; Papola, 2000; Wen and Koppelman, 2001, Bierlaire 2001, Papola, 2004, Abbe et al. 2007, Marzano and Papola 2008).
The CNL choice probability can be written in a conceptual form as:

$$p^{n,t}(j) = \sum_{k \in \mathrm{K}} p^{n,t}(j/k) \cdot p^{n,t}(k) \tag{2.31}$$

With the notation clarified in the previous sub-section, (2.31) can be expressed as a function of the degrees of membership (also called *inclusion* or *allocation parameters*):

$$p^{n,t}(j) = \frac{\alpha_{jk}^{1/\delta_k} \cdot \exp(V_j^{n,t} / \theta_k)}{\sum_{a \in k} \alpha_{ak}^{1/\delta_k} \cdot \exp(V_a^{n,t} / \theta_k)} \frac{\left(\sum_{a \in k} \alpha_{ak}^{1/\delta_k} \cdot \exp(V_a^{n,t} / \theta_k) \right)^{\delta_k}}{\sum_{k' \in \mathrm{K}} \left(\sum_{a' \in k'} \alpha_{a'k'}^{1/\delta_{k'}} \cdot \exp(V_{a'}^{n,t} / \theta_{k'}) \right)^{\delta_{k'}}} \tag{2.32}$$

Some authors, starting from Wen and Koppelman (2001), refers to it with the name of Generalized Nested Logit model (GNL), that theoretically differs from the CNL proposed by Vovsha (1997) for the presence of different values of δ_k for each nest. However, naming it CNL is not imprecise and giving a different name for a very slight increase of generality appears to be excessive.
For giving the same average value on random residuals the α_{jk} must satisfy the constraint that:

$$\sum_{k' \in \mathrm{K}} \alpha_{jk'} = h \qquad \forall j \in C^{n,t} \tag{2.33}$$

Generally, the positive constant h is normalized to 1.
The derivation of CNL from GEV theory is due to Papola (2004). The CNL totally relaxes the independence assumption of the simple MNL, allowing for a very flexible substitution pattern of correlations. Unfortunately, the covariance for CNL is not computable with a closed-form expression, involving a double-integral simulation (Abbe et al., 2007). Papola (2004) provided a conjecture that is very useful for reproducing covariances in boundary cases:

$$Cov\{i, j\} = \frac{\pi^2 \cdot \theta^2}{6} \cdot \sum_{k \in \mathrm{K}} \sqrt{\alpha_{ik} \cdot \alpha_{jk}} \cdot \left(1 - \delta^2{}_k\right) \tag{2.34}$$

Marzano and Papola (2008) explored the capability of the CNL to reproduce all the feasible homoscedastic covariance matrices domain underlying the CNL (i.e. the domain of matrices satisfying the maximum rank condition), and they showed the CNL is not able to reproduce the whole domain. They tested also the flexibility of various nesting structures, finding in the full-nest structure, i.e., a double nest error structure with nests containing all the alternatives, the most flexible structure. Thus, the Pair Combinatorial Logit specification (Chu 1989; Koppelman and Wen, 2000), that is very interesting and widely used in route choice, as

described in Section 2.3.2, does not represent the best strategy in terms of covariances flexibility. Marzano et al. (2013) demonstrated the possibility to express the covariance expression in a single-integral form, reducing considerably the computation time. In a successive work, Marzano (2014) brilliantly generalized the procedure for all GEV models.
Therefore, the CNL introduces a great flexibility in substitution across alternatives, but it is still an homoscedastic model and does not allow for correlation of random residuals over time. However, Fosgerau et al. (2013) carried out a very hard demonstration of how the Cross Nested logit model may theoretically approximate any ARUM, but no guidance on how to specify a CNL for this purpose has been provided.
Further developments on the topic of the correlations have been: the *cross correlated Logit* (Williams, 1977; Williams and Ortuzar, 1982), the *Tree Extreme Value model* (*TEV*; McFadden,1981), the *principles of differentiation model* (*PD*; Bresnahan et al., 1997), the *Generalized MNL* (also called *Choice set generation Logit* or *GenL*; Swait, 2001) and the Fuzzy Nested Logit[9].
The most known generalization of the CNL is the Recursive Nested Extreme Value model (RNEV), proposed by Daly(2001*b*) and successively reprised by Daly and Bierlaire (2006) and Newmann (2008). The model represents an extension of the CNL to more levels. The principle is that the choice can be represented by a graph, with the alternatives represented by a set of nodes at the lowest level, without successors. For each level, a set parameter can be associated, and a GEV generating function can be computed as a function of these parameters and the GEV function of the lower level. In fact, according to Daly and Bierlaire (2006), the *GEV-inheritance theorem* ensures the resulting function is always a GEV function. But, apart from route choice (Papola and Marzano, 2013), the model has not received a strong interest in real-world applications for the excessive numbers of parameters it introduces and the very quite incremental improvement in terms of correlations reproduced, when comparing it with a simpler CNL.

FinMix

The FinMix model (Swait, 2003) is recalled mainly for its structure, very similar, but not identical, to the one of the CoRUM model. The conceptual vehicle is always the same, i.e. allowing for flexible correlation patterns. With this purpose, the author derived the model by combining GEV generating function of different models in the way described in (2.21). In the formula, c is the generic generating function component and C is the set of all generating functions involved. In fact, FinMix is the abbreviation of "finite mixtures". The resulting function is always a GEV function, given by:

$$G = \sum_{c' \in \mathrm{C}} \left(G^{c'} \right)^{\frac{\theta_{c',0}}{\theta_0}} \tag{2.35}$$

Wherein c' represents the generic generating function, while the generic $\theta_{c',0}$ is the first level (or 0-level) variance parameter of the model underlying the c^{th} generating function and θ_0 is the total variance parameter.
The resulting probability expression is a weighted combination of choice probabilities, each one obtained from its underlying generating function $G^{c'}$ in a conditional form:

$$p^{n,t}(j) = \sum_{c' \in \mathrm{C}} p^{n,t}(j / c') \cdot p^{n,t}(c') \tag{2.36}$$

[9] The idea of Fuzzy Logit, i.e. a multi-level CNL, is due to Vovsha, but it is unpublished. It is briefly described in Koppleman and Sethi (2000), with reference to an email sent by the author.

However, in this case, the $p^{n,t}(c')$ has a different meaning than the nest probability $p^{n,t}(k)$ of the NL. In fact, it represents a weight indicating the probability of belonging to that generating function c'. As it will be seen, the conceptual vehicle is very similar to the one of the latent class models for reproducing taste variation (see Section 2.2.2.3). It can be seen as a latent preference for GEV function c, in a latent class fashion. However, differently from latent class models, wherein the assumption made on the variable indicating a generic class (analogously here the generating function c') is not explicit, here the assumption is explicitly made through $G^{c'}$. In other words, the latent class are generally modelled with a non-parametric assumption on the distribution of the variable that is considered to be latent, while here there is an explicit parametric assumption on each c'. Thus, the resulting random residuals distribution is a combination of known random residuals distributions.
Particularly, the formulation for $p^{n,t}(j/c')$ strictly depends upon the assumption on $G^{c'}$. For instance, assuming (2.17) gives the MNL formula (2.22), while (2.18) gives the entire NL formula (2.29) and so on, without any limitations.
The $p^{n,t}(c')$, instead, can be expressed as a function of all the G^c as:

$$p^{n,t}(c) = \sum_{c' \in \mathrm{C}} \frac{\exp\left[\frac{\theta_{c,0}}{\theta_0} \cdot \ln\left(G^{c,n,t}\right)\right]}{\exp\left[\frac{\theta_{c',0}}{\theta_0} \cdot \ln\left(G^{c,n,t'}\right)\right]} \tag{2.37}$$

In a tree representation, this first choice-level is perfectly analogous to the first level (or 0-level) of the NL, while the successive are the choice trees depending on the assumption on $G^{c'}$.
The probability $p^{n,t}(j/c)$, thus, represents an endogenous weight for the sum (2.36), given the fact that its value entirely depends on the structural parameters of all the component models. In fact, it is easy to recognize that it is impossible to identify the value of θ_0 per se, given the indifference of the random utility models with reference to the scale.
A very interesting example of FinMix is obtained mixing several (2.18) in (2.35). in this case $G^{c'}$ is the denominator of the NL probability raised to a power of the ratio $\theta_{c',0}/\ \theta_0$. This specification will be analysed on the basis of several synthetic experiments in Section 4.3.

2.2.1.2 Multinomial Probit

It can be said that the random utility theory was born with Probit model. The first intuition about the framework goes back to Thurstone (1927), who tested the binomial choice behaviour of his students with reference to a panel of repeated choice situations. He noted that, a person who faced more than once the same binomial choice scenario sometimes chose a different alternative. He came at the intuition of a similar framework to (2.3), wherein he interpreted the random term as a disturbance in the decision-maker perception, i.e. an imperfect awareness due to the human imperfection. The disturbance he considered was related to what he called *stimuli*, that represented a sort of precursor concept of the one of *perceived utility*. He assumed this disturbance to follow a Normal distribution. This led to a bivariate Normal density for the disturbances, and, thus, for the likelihood of preferring an alternative to the other. The first transport application of the binary Probit model is probably due to the PhD thesis of Lisco (1967) and the first book entirely dedicated to this topic is the one of Finney (1971).
However, the Multinomial Probit model, as it is known nowadays, is generally attributed to the paper of Hausman and Wise (1978) and the book of Daganzo (1979). The main assumption is that random residuals follow a Multivariate Normal distribution:

$$\varepsilon_i^{n,t} \sim MVN(\mathbf{0}, \mathbf{\Sigma}) \tag{2.38}$$

Where **0** represents the null vector mean and Σ is whatever feasible covariance matrix. Given (2.15) this leads to a choice probability defined as:

$$p^{n,t}(j) = \int_{\boldsymbol{\varepsilon}} \mathrm{I}_j^{n,t}\left[h(\mathbf{X},\boldsymbol{\varepsilon}) = j\right] \cdot \frac{1}{\sqrt{\left(2\cdot\pi\right)^{\mathrm{m}} \cdot \det(\boldsymbol{\Sigma})}} \cdot \exp\left(-\frac{1}{2}\cdot\boldsymbol{\varepsilon}\cdot\boldsymbol{\Sigma}^{-1}\cdot\boldsymbol{\varepsilon}^{\mathrm{T}}\right) d\boldsymbol{\varepsilon} \tag{2.39}$$

where the integral form holds because a closed form is not available for (2.39).
The MNP can be considered a benchmark in terms of flexibility in reproducing correlations effects. In fact, the MNP overcomes each one of the three problems mentioned in the premise of this section. The MNP is an heteroskedastic model, it is flexible in reproducing covariances alternatives and it allows for random terms to be correlated over time. In fact, it has been noted that **ε** indicates the vector of $\varepsilon_i^{n,t}$, and in a context wherein random residuals are expected to be correlated over time, the integral (2.39) becomes an (m x T) dimensional integral and **Σ** is the (m x T) quadratic covariance matrix of each residual *i* in a choice scenarios *t*.
It has been noted that, differently from GEV models, here the covariance matrix **Σ** explicitly appears. Actually, reminding that only utility differences matters, without any importance for level and scale, (2.39) can be written as a function of differences in random residuals (Hausman and Wise, 1978):

$$p^{n,t}(j) = \int_{\varsigma} \mathrm{I}_j^{n,t}\left[h(\mathbf{X},\varsigma) = j\right] \cdot \frac{1}{\sqrt{\left(2\cdot\pi\right)^{\mathrm{m\text{-}1}} \cdot \det(\boldsymbol{\Omega})}} \cdot \exp\left(-\frac{1}{2}\cdot\varsigma\cdot\boldsymbol{\Omega}^{-1}\cdot\varsigma^{\mathrm{T}}\right) d\varsigma \tag{2.40}$$

Wherein **ζ** is the (m-1)-vector of the differences in random residuals, while **Ω** represents the covariance matrix of them, with reference to whatever alternative random residual. This implies the independent covariance matrix elements of (2.39) that are identifiable are not m(m+1)/2, but (m-1)m/2-1 (the -1 indicates the possibility to normalize the value of one of them, because, as already said, scale does not matter). Thus, when one wants to estimate a MNP model can follow two strategies: directly specifying the error structure in the space of differences among utilities, that is a (m-1) dimensional space, or in the space of utilities, that is a m-dimensional, but conscious of the fact that only (m-1)m/2-1 terms of Σ are actually identifiable. However, specifying an MNP in terms of differences and, in general, specifying an MNP in contexts with an high number of alternatives could be prohibitive. The reason will be better explained below. Daganzo (1979) explains also a procedure for reducing the number of integration to (m-2) with a transformation that exploits the **ζ** and the Choleski lower triangular matrix of **Ω**.
The integral solution, as mentioned, needs performing simulation. The first method is using the classical quadrature method for computing multi-dimensional volume of (2.39) or (2.40) (Butler and Moffit, 1982; Guilkey and Murphy, 1993; Geweke, 1996) but, unfortunately, it is too onerous to perform in contexts with more than 4/5 alternatives. The Monte-Carlo simulation (with the A/R (acceptance/rejection) method is the first simulated procedure adopted for MNP (Lerman and Manski, 1977). Practically, the MNP probability is simulated drawing many times from the density f(**U**), therefore, under the assumption of the current section (no taste variation), from the density f(**ε**), and assigning a deterministic probability of 1 to the alternative who exhibits the maximum $U_i^{n,t,r}$ in the drawing *r* and 0 to the other. The simulated MNP probability results in a ratio between the favourable cases (i.e. the sum of the $p(i)^{n,t,r}$ with assigned value of 1) and the total of draws R:

$$p_{sim}{}^{n,t}(j) = \sum_{r=1}^{\mathrm{R}} \frac{p_{det}{}^{n,t,r}(j/\boldsymbol{\varepsilon}^r)}{\mathrm{R}} = \frac{n_j}{\mathrm{R}} \tag{2.41}$$

wherein $\boldsymbol{\varepsilon}^r$ indicates the generic vector of $\varepsilon_i^{n,t,r}$ drawn from f($\boldsymbol{\varepsilon}$) and $p(j/\boldsymbol{\varepsilon}^r)_{det}{}^{n,t,r}$ is the conditional deterministic probability.
A very comprehensive analysis on simulation procedure for MNP is reported in Hajivassiliou et al. (1996), whose paper investigates a lot of simulators, namely: Crude Frequency (CFS), Normal Importance Sampling (NIS), Kernel-Smoothed Frequency (KFS; McFadden, 1989), Stern Decomposition (SDS; Stern, 1992), GHK(Geweke, 1992; Borsch-Supan and Hajivassiliou, 1993; Keane, 1994; Hajivassiliou et al., 1996), Parabolic Cylinder Function (PCF), Deak Chi-Square (DCS; Deak, 1980a,b), Acceptance/Rejection (A/R), Gibbs Sampler (GSS; Hajivassiliou and McFadden, 1990), Sequentially Unbiased (SUS) and Approximately Unbiased (AUS). The authors stated the superiority of the GHK simulator that seems to be the most used simulator for MNP (Train, 2009). Another simulator frequently used in the first MNP application is the Clark's algorithm (Clark, 1961), adopted for MNP first by Daganzo et al. (1977a). But Horowitz et al. (1982) showed this algorithm may be inaccurate in some situations.
However, the MNP needs to be specified for allowing a maximum degree of flexibility for $\boldsymbol{\Sigma}$ or, more precisely, for $\boldsymbol{\Omega}$. The first way to specify the MNP for allowing a flexible correlation pattern is the *error component specification*. The concept is analogous to that of NL model when a correlation among two alternatives is introduced sharing a portion of random residual. Each random residual $\varepsilon^{n,t}{}_i$ is specified as a sum of mono-variate normally distributed error components $\eta^{n,t}{}_{ij}$~N(0,$\sigma^2{}_{\eta ij}$), where the subscript *ij* indicates the membership of the error component to the specification of residuals $\varepsilon^{n,t}{}_i$ and $\varepsilon^{n,t}{}_j$. Obviously, each random term is specified with an own $\varepsilon^{n,t}{}_{ii}$, allowing for reproducing alternative specific variances. Thus, m(m+1)/2 random terms can be inserted into the random residuals specification. However, taking into account the identification issues due to scale and level above mentioned, only (m-1)m/2-1 parameters (namely the variances $\sigma^2{}_{\eta ij}$) can be estimated. This means specifying the error component structure in the space of differences among utilities or specifying in the space of utilities with some devices. For instance, a random residual can be imposed to be 0 for each individual and choice scenario, leaving only (m-1)m/2 error components in the (m-1) remaining residuals, and whatever remaining error component variance $\sigma^2{}_{\eta ij}$ can be set to 1, or whatever positive value. Bunch (1991) provided an easy explanation of this procedure expressing covariances in utilities differences as a function of covariances in utilities.
Another well-known method for drawing from a Multivariate Normal density is the *Factor Analysis* (Fishman, 1973). It is a similar specification to Error Component, expressing $\varepsilon_j^{n,t}$ as a combination of coefficient f_{jk} and Standard Normal variables $z_j^{n,t}$. The coefficients are computed by the *Choleski decomposition* of a given covariance matrix $\boldsymbol{\Sigma}$. Thus, when one wants to estimate a MNP model he must hypothesize, for each estimation iteration, a matrix $\boldsymbol{\Sigma}$, compute a Choleski lower diagonal matrix $\mathbf{F}_\Sigma$ and operate the scalar product $\varepsilon_j^{n,t} = \mathbf{F}_\Sigma \cdot \mathbf{z}$, where $\mathbf{z}$ is the vector of the standard Normal draws $z_j^{n,t}$. However, the problems due to level and scale holds and the analyst must preliminarily verify his assumption on error structure on $\boldsymbol{\Sigma}$ to ensure its identification.
In a panel data context, one wants to take into account that each individual *n* faces several scenarios *t*, treating all choices of the same individual as a unique choice being a sequence of them. It can be made the assumption that $\varepsilon_j^{n,t}$ does not vary across choice tasks *t*. Practically, for each individual, the same random residual ε_j^n is considered.
The simulation issues above described explodes when estimating parameters of the MNP. In fact, different estimators than the closed form models case must be used. Generally, the MNP

is estimated through the maximum simulated log-likelihood estimator. Unfortunately, this estimator has several issues of stability, as widely described in Train (2009) in his chapter 8. Furthermore, especially when using the A/R simulator, the simulated choice probability is a discontinuous function in parameters. In fact, it is easy to recognize that computing the probability as quantity of successes on the total number of draws means obtaining a step function of the $p^{n,t}(j)$ in parameters. This means the first order derivatives of simulated choice probability is 0 or undefined, making impossible using the classical gradient-based algorithms for maximizing the simulated log-likelihood function. This is the reason that drove McFadden (1989) creating a simulator for computing MNP probabilities that were smoothed in parameters, i.e. twice differentiable. The simulator is obtained simply computing a MNL conditional probability for each vector of drawn perceived utilities, instead of a deterministic (accept/reject, i.e. 0/1) probability. The resulting simulated log-likelihood represents an efficient, unbiased and, as said, smooth estimator for the parameters. This operative intuition seemed to introduce a bias in the MNP choice probability but, actually, it represents the first intuition towards a new general and powerful model, i.e. the Mixed Logit.

2.2.1.3 Mixed models

A mixed model assumes random residuals distributed as sum of two terms (Ben-Akiva and Bolduc, 1996):

$$\varepsilon_i^{n,t} = \omega_i^{n,t} + \gamma_i^{n,t} \quad \forall i \in \mathrm{C}^{n,t} \tag{2.42}$$

wherein $\omega_i^{n,t}$ is generally Multivariate Normal distributed, while $\gamma_i^{n,t}$ is, generally, an extreme value term (see Section 2.2.3 for other more general examples such as Mixed Probit model).
The perceived utility, according to the assumption made on marginal utilities in this section, becomes:

$$U_i^{n,t} = \sum_k \beta_{k,i} \cdot X_{k,i}^{n,t} + \omega_i^{n,t} + \gamma_i^{n,t} \quad \forall i \in \mathrm{C}^{n,t} \tag{2.43}$$

This means, the general (2.15) becomes an integral in joint density f($\boldsymbol{\omega}$,$\boldsymbol{\gamma}$), that are generally assumed to be independent from each other, giving f($\boldsymbol{\omega}$,$\boldsymbol{\gamma}$)=f($\boldsymbol{\omega}$)f($\boldsymbol{\gamma}$). Thus, the integral becomes (Train, 1995; Train, 2009):

$$\begin{aligned} p^{n,t}(j) &= \int_{\boldsymbol{\omega}}\int_{\boldsymbol{\gamma}} \mathrm{I}_j^{n,t}\left[h(\mathbf{X},\boldsymbol{\omega},\boldsymbol{\gamma}) = j\right]\cdot f(\boldsymbol{\omega})\cdot f(\boldsymbol{\gamma})\,\mathrm{d}\boldsymbol{\omega} d\boldsymbol{\gamma} = \\ &= \int_{\boldsymbol{\omega}} \left\{ \int_{\boldsymbol{\gamma}} \mathrm{I}_j^{n,t}\left[h(\mathbf{X},\boldsymbol{\omega},\boldsymbol{\gamma}) = j\right]\cdot f(\boldsymbol{\gamma}) d\boldsymbol{\gamma} \right\} f(\boldsymbol{\omega})\,\mathrm{d}\,\boldsymbol{\omega} = \\ &= \int_{\boldsymbol{\omega}} g(\boldsymbol{\omega}, f(\boldsymbol{\gamma}))\cdot f(\boldsymbol{\omega})\,\mathrm{d}\,\boldsymbol{\omega} \end{aligned} \tag{2.44}$$

The integral within g(.) is a conditional function of $\boldsymbol{\omega}$. According to the i.i.d value assumption on $\boldsymbol{\gamma}$, g(.) is solvable in closed form.

Error Component Logit

The Mixed Logit (also called mixture of Logit's and abbreviated as MMNL) is by far the model who received more interest in literature in the last decades. Here is presented in a pure error component specification, so it will be called error component Logit. The error component Logit assumption is that $\gamma_i^{n,t}$ is an i.i.d. Gumbel term. This leads the g(.) becoming the Multinomial Logit formula, conditional on values of $\boldsymbol{\omega}$ (that needs to be drawn).
Thus, integral in (2.44) becomes:

$$p^{n,t}(j)=\int_{\boldsymbol{\omega}}\frac{\exp\left(\frac{V_j^{n,t}+\omega_j^{n,t}}{\theta}\right)}{\sum_{i\in C^{n,t}}\exp\left(\frac{V_i^{n,t}+\omega_i^{n,t}}{\theta}\right)}\cdot f(\boldsymbol{\omega})\cdot \mathrm{d}\boldsymbol{\omega} \tag{2.45}$$

The integral (2.44) needs to be simulated, but it is easy to recognize that the Smoothed A/R estimator described in the previous section is opportune for this purpose. This means that the Mixed Logit seems a random utility model born for justifying a biased estimator for the Multinomial Probit.

Ben-Akiva and Bolduc(1996) provided the first formalization of the model proposing the (2.42), naming it "*Probit with Logit kernel*", and presenting the Factor analytic specification for $\boldsymbol{\omega}$ and the random coefficient form for (2.45) (that will be discussed in Section 2.2.2.2). However, the generality of the Mixed Logit model has been demonstrated by McFadden and Train (2000). The authors showed that a Mixed Logit allows to approximate any RUM with a certain degree of closeness. This very hard mathematical demonstration is often used as a justification to use Mixed Logit model for any context. However, it is worthy to note that Train (2008a) itself pointed out that no guidance was proposed to obtain a specification of the Mixed Logit for ensuring it to mimic a target model. By the way, Train (2009) gives a more intuitive explanation of this powerful property in chapter 6 of his book.

The specification of $\boldsymbol{\omega}$ can be made in the same ways described for MNP, namely Error component and Factor analysis. However, in the error component Logit, one more parameter than the MNP is present, i.e. the homogeneous variance of the $\gamma_i^{n,t}$'s.

A very in-depth analysis on the identification issues of the Normal error component Logit model (NECLM) is reported in Walker (2002) and, above all, Walker et al.(2007), who investigate all the conditions required for the identification of the parameters.

The proposed procedure for identification of a NECLM is summarized as follows in a what-if logic:

- Hypothesize a covariance matrix (in a factor analytic fashion);
- Determine the covariance matrix of utility differences;
- Apply the *order condition*: this practically means checking that the number of parameters does not exceed m(m-1)/2-1, as described in Section 2.2.1.2 for MNP. It actually represents only an upper limit to the number of identifiable parameters (Bunch,1991);
- Apply the *rank condition*: this is a check on the rank of the Jacobian matrix of all the variances/covariances of utility differences expressed as a function of the covariances of utilities. Practically, the analyst must compute each first-order cross derivative of variances/covariances in utility differences (on the rows) with reference to all variances/covariances in utilities (on the columns). The rank of the resulting matrix minus 1 is another upper bound to the number of identifiable parameters. This condition is more restrictive than the order condition.
- Apply the *equality condition*: it is a check that the scale effectively does not affect the probabilities. This means the analyst must verify that normalization imposed to the covariance matrix of the utility differences (for instance fixing the variance parameter of Gumbel disturbance), does not make change in the value of the probabilities with reference to the non-normalized case. In other words, to keep the probabilities to the same value, the matrix of covariances of utility differences after normalization must be exactly equal to the one before normalization.

Notwithstanding the very high interest in literature for Mixed Logit, the error component specification remains a not so easy instrument to manage. In fact, it easy to recognize that, although all mentioned upper bounds to the identifiable parameters, their number rises when the number of alternatives raises. This means the NECLM can be easily estimated in contexts wherein there are many alternatives but a few parameters (Brownstone and Train, 1999). But, it is intuitive to recognize that using the NECML with this constraints can considerably reduce its flexibility. However, the Mixed Logit always remains the most used random utility framework for modelling taste variation across/within respondents (see Section 2.2.2.2).

Normal alternative specific Error Component GEV

The model (2.44) can be partially solved in closed form also with other GEV distribution assumptions for the $\gamma_i^{n,t}$'s. The resulting probability statement represents a Mixed GEV model. The utility of a more complex Error component GEV formulation seems very irrelevant, given the generality of the Error component with a Logit kernel. However, it can be justified when one wants to catch the inter-alternatives variance effects with a Normal error component structure, but the covariances effects with a less burdensome closed form GEV model. This allows to considerably reduce the random terms involved in the integral simulation (the integral involves a maximum of m-1 dimensions) and also to relax the homoscedastic assumption of the simple GEV model.
Although the advantage can be relatively smaller, with reference to the complexity to implement a more sophisticated GEV model, in a joint random coefficient error component specification this can represent a gain in terms of goodness of fit and allows for reducing the misleading about the taste variation effects and the variances/covariances effects. This framework will be deepened in 0, with reference to the CoRUM model.

2.2.2 Taste heterogeneity

The second relevant problem faced by this thesis work refers the taste heterogeneity. In general, the concept, as briefly introduced in Chapter 1, is that each decision maker gives a different importance to a generic variable involved in a utility function. It is knowable that several personal variables could affect the perception of a given attribute: social status, demographic condition, attitude, habits and so on. For instance, when the analyst wants to put the monetary cost in systematic utility, he should consider that the marginal (dis)utility due to monetary cost can be heavily affected by: personal income with reference to the level of welfare of the community, sex, attitude to spend money for that specific alternatives, personal inclination to spend money in general and so forth. Furthermore, the individual can vary his inclination over time or decision tasks, because one of the mentioned conditions can vary, or simply because his personal inclination could arise differently in different choice scenarios.
The failure to take account this inter/intra-personal variation can strongly affect the assessment on the weight of some attribute in forecasting.
However, similar to Section 2.2.1, considering one by one all the random effects, at less from a conceptual standpoint , it is assumed that all randomness lies in the $\beta_{I,k}^{n,t}$, while $\varepsilon_i^{n,t}$ is fixed to be 0:

$$U_i^{n,t} = \sum_k \beta_{k,i}^{n,t} \cdot X_{k,i}^{n,t} \qquad \forall i \in \mathrm{C}^{n,t} \tag{2.46}$$

It has been pointed out that (2.46) does not exclude the correlation effects. In fact, the (2.46) is random within the coefficients, thus a variance into the utility is present, and a covariance among the utilities sharing the coefficients is induced (see Daganzo, 1979; Ben-Akiva and Lerman, 1985; Cascetta, 2009). However, the focus of this section is on the taste variation. For further details on the variances and covariances induced by (2.46) the reader can refer the insights of Chapter 3.

Under the assumption (2.46), the current section will consider a particularization of (2.15) given by:

$$p^{n,t}[j] = \int_{\boldsymbol{\beta}} \mathrm{I}_j^{n,t}\left[h(\mathbf{X},\boldsymbol{\beta}) = j\right] \cdot f(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta} \tag{2.47}$$

The following sub-sections illustrate the procedures generally adopted for taking into account this problem. Firstly, they are presented some approximate techniques for treating as systematic the taste variation (see Section 2.2.2.1). Second, all the stochastic approaches are described, namely the parametric, nonparametric and semi-nonparametric approaches. All models implementing these approaches will be treated with reference to all main contributes to the scientific literature.

2.2.2.1 *Observable taste heterogeneity*

As an initial approximation, the random taste heterogeneity can be treated as systematic heterogeneity. Although this is not a complete approach from a theoretical standpoint, this can considerably improve the goodness of fit measure and, in some cases, it avoids to moving toward formulations that are more complex.

In fact, the reader must remember that all the integration variables in (2.15) refer the unobservable components. The impossibility to observe some attribute can be considered always a limit of the analyst in that particular choice context and with that particular dataset. In fact, it can be said that Multinomial Logit does not fail in a certain choice context per se, while it fails given the lack of knowledge of the variables influencing choice probabilities. In other words, from a pure theoretical standpoint, if the analyst is able to observe all the variables that induces a choice, the problems of random correlations / random heteroscedasticity /random taste heterogeneity and so on reduce their importance. In fact, given the additive framework (2.3), it is immediate to recognize that everything that is observable to the analyst is put into the systematic utility and it is taken off from random residual. Thus, the relative importance of $\varepsilon_i^{n,t}$ in $U_i^{n,t}$ decreases, and the random effect fades. Practically, if one put the right variables into the utility of bus and car in the famous red bus/blue bus paradox, maybe the systematic utilities of the three alternatives considerably change their value, the random term tend to be null and the model reproduces the expectations in terms of choice probabilities.

The same concept applies to taste heterogeneity. When treating it as function of observable attributes, the necessity of resorting to complex multi-dimensional integrations consistently reduces. Practically, doing it means adding more attributes to the systematic utility specification. Each choice context is substantially different from another, so giving general rules is obviously an unrealistic pretence. However, some interesting example for travel demand can be shown. Considering, for instance, the monetary cost above mentioned. It can be entered into the systematic utility of a certain transport mode (car, train, air, bike etc.) and it is surely expected to be very variable across the population. Firstly, it can vary as a function of the personal (or family) income. Particularly, it is expected that the income increases the utility, so, when putting the income in the formulation as an adding variable the analyst must estimate another parameter, expected to be positive. The estimation of another parameter can be avoided, obtaining the income effect on the utility by putting it in the denominator of the monetary cost. This implies when the income rises, the ratio between monetary cost and income drops, giving a plausible effect on the systematic utility, without requiring other parameters to be estimated, or worse, a randomness introduced in the marginal utilities. This strategy can be surely refined. For example, a ratio between average income and individual income can be put into the utility for adjusting the relative importance of rising individual income. This ratio can be also used in a non-linear way, elevating to the power of a certain exponent. Other socio-demographic information can be put into the utility, multiplying it for the same monetary cost coefficient. Some examples are the age, the gender (dummy variable) or the professional condition (ranked variable). The reason for what a decision maker makes a trip could certainly influence his willingness to spend money (willingness to pay), so another dummy variable can be inserted multiplying it for the monetary coefficient.

Regarding the time coefficient the same concept applies. The analyst can try to improve his specification, detecting all the relevant variables that are observable. It is expected, for example, that a time spent on board a train can be influenced from some service attributes. If a decision maker knows he can access to a Wi-Fi, or a bar/restaurant, or a comfortable seat, he will be more willing to spend time on board for that class of that train. And surely, the mentioned variables could also impact on willingness to spend money.

Apart from the addition of attributes, the analyst can try to use the same attribute, but introducing more than one parameter for it (*segmentation*). For example, when using the income together with the cost, one can estimate more values, one for each income bracket. The same can be made on the age and so on.

Definitely, before resorting to models that are more complex (see Sections 2.2.2.2, 2.2.2.3 and 2.2.2.4), the analyst should improve the utilities specification adding more observable information within them. This reduces the presence of randomness in the perceived utilities, limiting the benefits (but also the burdens) that one obtains from models allowing for random taste variation. Unfortunately, sometimes it is impossible for the analyst observe all the sources of taste variation, potentially leading to inconsistent parameters estimates (Chamberlain, 1980). This is the reason why, after a good specification work, often the analyst has to resort to formulations accounting for the random sources of taste heterogeneity.

2.2.2.2 Parametric approach

This sub-section provides a brief recap of the random utility models that explicitly take into account the random taste heterogeneity. A first statistical approach is the parametric one. With this approach, the analyst must explicitly make an assumption on the shape of the random distribution of the marginal utilities. Thus, for operationalizing the model, he must estimate the parameters from which the parametric distribution depends on.

When using a parametric approach the underlying model specification is generally called *random coefficient*. The latter has a long history in econometrics applications, before being applied to discrete non-linear models. Initially, it was used mainly within the linear regression models, to relax the assumption that the coefficients of the regression were fixed (see Hildreth and Houck, 1968, Swamy, 1972, Rao, 1973 and Greene, 2001 for a summary of the main contributes). In the following sub-sections the most relevant random utility models specified as random coefficient will be discussed, namely the MNP and the MMNL.

Probit with random coefficient formulation

The Probit with random coefficient specification (Hausman and Wise, 1978; Akin et al., 1979) represents the first model taking explicitly into account the random taste variation. The framework is perfectly analogous to that of (2.39) but, in this case, the random vector $\boldsymbol{\beta}$ replaces $\boldsymbol{\varepsilon}$:

$$p^{n,t}(j) = \int_{\boldsymbol{\beta}} \mathrm{I}_j^{n,t}\left[h(\mathbf{X},\boldsymbol{\beta}) = j\right] \cdot \frac{1}{\sqrt{(2\cdot\pi)^{\mathrm{m}} \cdot \det(\Sigma_{\boldsymbol{\beta}})}} \cdot \exp\left(-\frac{1}{2}\cdot\boldsymbol{\beta}\cdot\Sigma_{\boldsymbol{\beta}}^{-1}\cdot\boldsymbol{\beta}^{\mathrm{T}}\right) d\boldsymbol{\beta} \tag{2.48}$$

wherein $\Sigma_{\boldsymbol{\beta}}$ represents the covariance matrix among the $\beta^{n,t}_{j,k}$'s considered to be distributed as $\boldsymbol{\beta} \sim MVN(\bar{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})$. Generally, the Normal distributed $\beta^{n,t}_{i,k}$ are simply assumed to be independently distributed. Thus, this means simulating the integral (2.48) by drawing each parameter from a mono-variate Normal density as:

$$\beta_{i,k}^{\ n,t} = \bar{\beta}_{i,k} + \sigma_{\bar{\beta}_{i,k}} \cdot z_{i,k}^{n,t,r} \tag{2.49}$$

wherein $z_{i,k}^{n,t,r}$ is the Standard Normal r^{th} draw from the density of the generic $\beta^{n,t}_{i,k}$.

When hypothesizing a correlation among the parameters, the Choleski factorization can be adopted. Obviously, it implies estimating also the Choleski constants and, indirectly, also the covariances among the parameters.

However, nowadays the MNP is no longer used, because of its already mentioned problems of estimation, while the pure random coefficient Logit (or Mixed Logit random coefficient) is by far the most used parametric model in the practical applications.

Random coefficient Logit

Although the Mixed Logit with random coefficient is often considered born with Ben-Akiva and Bolduc (1996) paper, an earlier literature faces the problem of mixtures of MNL with random coefficient. For example, Cardell and Dunbar (1980) and Boyd and Mellman (1980) tried to simulate an integral referred to market share, rather than a single decision maker, in an automobile demand model. In McFadden and Train (2000) several other previous works are mentioned (Talvitie, 1972; Westin, 1974; McFadden and Reid, 1975; Westin and Gillen, 1978) and many other successive works.

However, the random coefficient Logit (or Mixed Logit random coefficient, or mixtures of MNL) has incredibly increased its popularity in the first decade of the 2000's. It is very hard to embrace all the relevant literature on it.

Anyway, the random coefficient is perfectly similar to the error component Logit (sometimes said to be mathematically equivalent to it, except for putting random term inside or outside the systematic utility) and analogous to the MNP random coefficient. The choice probabilities descends from the assumption that the perceived utility is defined as:

$$U_i^{n,t} = \sum_k \beta_{k,i}^{n,t} \cdot X_{k,i}^{n,t} + \gamma_i^{n,t} \qquad \forall i \in \mathrm{C}^{n,t} \tag{2.50}$$

Wherein all the involved variables has the meaning already clarified in the previous sections, but the $\beta^{n,t}_{i,k}$ can be distributed in any way. It is clear that the MNP model can be seen as a particular case of (2.50), wherein $\gamma^{n,t}_i$ is fixed to be 0 and the $\beta^{n,t}_{i,k}$'s are Normally distributed.

The choice probability formula is given by:

$$p^{n,t}(j) = \int_{\boldsymbol{\beta}} \frac{\exp\left(\dfrac{V_j^{\ n,t}(\boldsymbol{\beta})}{\theta}\right)}{\sum\limits_{i\in C^{n,t}} \exp\left(\dfrac{V_i^{n,t}(\boldsymbol{\beta})}{\theta}\right)} \cdot f(\boldsymbol{\beta}/\boldsymbol{\Phi}) \cdot \mathrm{d}\boldsymbol{\beta} \tag{2.51}$$

wherein $f(\boldsymbol{\beta}/\boldsymbol{\Phi})$ represents the density function of $\boldsymbol{\beta}$ characterized by a set of parameters $\boldsymbol{\Phi}$ that need to be estimated. The (2.51) can be seen as the average value of the MNL choice conditional probabilities (conditional to $\boldsymbol{\beta}$), i.e. a sum of infinite terms, each one weighted with the corresponding f($\boldsymbol{\beta}/\boldsymbol{\Phi}$)d$\boldsymbol{\beta}$.

The choice of distribution shape f($\boldsymbol{\beta}/\boldsymbol{\Phi}$) is the most important step to make the model operative. In fact, starting from such assumption, the integral is simulated in different ways. The Normal distribution for $\boldsymbol{\beta}$, anyway, still represents a widespread used assumption. However, a lot of other distributions have been tested in literature. The issue of the Multivariate Normal is that it is unrealistic because of it is an unbounded and asymmetric distribution. Apart from the problem of considering also very big values, it is undesirable because, often, one would estimate a positive/negative definite distribution. For example, in the monetary cost case, it is weird to observe positive values of its marginal utility, because it would mean a decision maker gets a benefit from spending more money. The same applies to the time coefficient and many other cases. Unfortunately, drawing from a truncated Multivariate Normal can be computationally intensive, because an A/R simulator for a vector of random variables $\boldsymbol{\beta}$ should reject all the values outside the intervals, and this may give a high number of rejects, stretching the computational time. Thus, the analyst can use semi-definite distributions or limited distributions. To name a few, the log-Normal, the Triangular (symmetric or not), the Rayleigh, the Gamma, the skew Normal, the logistic transformation of Normal and the general Sb-Jhonson and even the Uniform are applied. The reader can see Mehndiratta (1996), Revelt and Train (1998), Bhat (1998a), Revelt (1999), Revelt and Train (2000), Bhat (2000), Train (2001), Siikamaki (2001), Siikamaki and Layton (2001), Hensher and Greene (2003), Train and Sonnier (2005), Burda et al. (2008), Bhat (2011), Bhat and Sidharthan (2012) Keane and Wasi (2013) and Dekker (2016) for examples of applications.

Another problem influencing the choice of the distribution is the space considered for the estimation. When using the preference space, i.e. estimating the distribution of coefficients $\beta^{n,t}_{i,k}$ as put into (2.50), it is possible that the analyst cannot understand the distribution of some interesting transformation of them. The classical example is the willingness to pay (WTP), generally fundamental in the transport mode choice. If one want to know the actual distribution of the WTP, but he has made the assumption of Normal distribution for the cost and the time coefficient, he cannot say anything about the distribution of WTP, because the ratio between two Normal is no longer a Normal. This can be solved by hypothesizing a log-Normal distribution, because of the ratio between two log-Normal is still a log-Normal (given the exponential properties). Another way is to estimate the model directly in the WTP (or evaluation) space (Ben-Akiva et al., 1993), i.e. explicitly putting the WTP measure (for example the value of travel time VTTs) into the systematic utility. Making a parametric assumption on it allows for estimating its distribution. Unfortunately, neither log-Normal distributions nor estimation in WTP space are always simple to be estimated.

However, the random coefficient Logit allows for incredible improvements in goodness of fit when using panel data (Revelt and Train, 1998), as already mentioned in Section 2.2.1.2 for MNP model. This is the main reason for what it is the most used parametric model for capturing the random taste variation.

Posterior analysis and individual-level parameters on continuous distributions

The Mixed Logit allows for taste variation and estimating a precise distribution, dependent from the assumption made on it. However, in terms of prediction, this represents a limited utility (Hess et al., 2010), because of the analyst, when applying the estimated MNP or (more suitably) MMNL, must simulate the integral over the density across the observations. This means he cannot know exactly where a particular decision maker lies in that distribution.

To overcome this limitation, a powerful tool for prediction is the *posterior analysis*. The concept is that the information about past choices of the individual can be used to make inference on the individual, in a Bayesian fashion. If the analyst considers m alternatives, and thus m observed market shares in a revealed choice experiment, the revealed choices can be used to build m distributions conditional on the revealed choice. Thus, for each decision maker, it will be more appropriate to draw from the density representative of his own marker share. This means passing from an unconditional distribution on the whole population f($\boldsymbol{\beta}/\boldsymbol{\Phi}$) to a conditional distribution g($\boldsymbol{\beta}/y^{s,t}, \boldsymbol{\Phi}$), indicating the distribution of people *s* that would choose the alternative *y* when faces the choice scenario *t* when the parameters of the whole population are $\boldsymbol{\Phi}$.

This easy concept can be stressed and extended at the individual level. In fact, given a sequence of choices (for example in a panel data of observed choices or in an SP survey) for a single individual y^n, the conditional likelihood L($\boldsymbol{\beta}/y^n,\boldsymbol{\Phi}$) of the decision maker to have a certain vector of values for $\boldsymbol{\beta}$ can be computed. Practically, applying the Bayesian rule, this is:

$$\mathrm{L}(\boldsymbol{\beta} / \mathrm{y}^n, \boldsymbol{\Phi}) = \frac{p(\mathrm{y}^n / \boldsymbol{\beta}) \cdot f(\boldsymbol{\beta} / \boldsymbol{\Phi})}{p(\mathrm{y}^n / \boldsymbol{\Phi})} \tag{2.52}$$

The term $p(y^n/\boldsymbol{\beta})$ represents the product of T conditional MNL probabilities of the choices *j(t)* that a decision maker makes in his choice sequence:

$$p(\mathrm{y}^n / \boldsymbol{\beta}) = \prod_{t=1}^{\mathrm{T}^n} \frac{\exp\left[V_{j(t)}^{n,t}(\boldsymbol{\beta})\right]}{\sum_{i \in C^{n,t}} \exp\left[V_i^{n,t}(\boldsymbol{\beta})\right]} \tag{2.53}$$

f($\boldsymbol{\beta}/\boldsymbol{\Phi}$) is the estimated density function and the $p(y^n/\boldsymbol{\Phi})$ represents the Mixed Logit formula applied to a sequence of choices y^n. In other words, it is:

$$p(\mathrm{y}^{n,t} / \boldsymbol{\Phi}) = \int_{\boldsymbol{\beta}} \prod_{t=1}^{\mathrm{T}^n} \frac{\exp\left(\frac{V_j^{n,t}(\boldsymbol{\beta})}{\theta}\right)}{\sum_{i \in C^{n,t}} \exp\left(\frac{V_i^{n,t}(\boldsymbol{\beta})}{\theta}\right)} \cdot f(\boldsymbol{\beta} / \boldsymbol{\Phi}) \cdot \mathrm{d}\boldsymbol{\beta} \tag{2.54}$$

The formula (2.52) needs simulation but all the quantities within it are computable after the model estimation. The mean of the individual level parameter can be computed as:

$$\overline{\boldsymbol{\beta}} = \int_{\boldsymbol{\beta}} \boldsymbol{\beta} \cdot \mathrm{L}(\boldsymbol{\beta} / \mathrm{y}^n, \boldsymbol{\Phi}) \cdot \mathrm{d}\boldsymbol{\beta} \tag{2.55}$$

This represents the basic idea on what the Bayesian estimation is founded. In fact, in a Bayesian procedure the f($\boldsymbol{\beta}/\boldsymbol{\Phi}$) represents the *prior*, i.e. the summary of all the information that are available for the analyst, previously to the estimation. Starting from such prior, the procedure searches for the parameter of the new conditional distribution, avoiding optimization problems and relative issues.

This can be an advantage in terms of prediction, but it has been noted that it can be made only with individuals whose past choices are available. However, with the increasing availability of micro-data, this problem looks set to become relatively less significant.

2.2.2.3 Nonparametric approach

The parametric approach for taste variation and, particularly, the use of Mixed Logit, has an "original sin". In fact, making a specific assumption on the shape of the density function is very restrictive. Whatever distribution hypothesized by the analyst, the estimation procedure will provide some parameters for that distribution, even if it is a very wrong assumption. Generally, a Mixed Logit model exhibits a better goodness of fit than the simple MNL, but a wrong distribution may lead to biased predictions. Furthermore, the posterior analysis is strongly dependent from the density function f($\boldsymbol{\beta}/\boldsymbol{\Phi}$) estimated.

Thus, it is easy to recognize that an approach avoiding a preliminary assumption on the distribution can be suitable for many applications, especially when a knowledge of the approximated shape of distribution is not available a priori. From an operative standpoint, this means assuming the continuous function f($\boldsymbol{\beta}/\boldsymbol{\Phi}$) approximated by a discrete set of mass points, for what no assumption on the reciprocal position is made. For each mass point, the analyst must estimate the value of the variable and its mass probability. This practically means transforming the multi-dimensional integral in (2.48) in a finite sum over a discrete set of points(*segments*) *s*, as:

$$p^{n,t}(j) = \sum_{s=1}^{S} \mathrm{I}_j^{n,t}\left[h(\mathbf{X}, \boldsymbol{\beta}^s) = j\right] \cdot p(\boldsymbol{\beta}^s) \tag{2.56}$$

Where S is the total number of discrete points and $p(\boldsymbol{\beta}^s)$ is the mass probability of the vector $\boldsymbol{\beta}^s$. The next sub-sections show some applications of (2.56).

Latent class MNL

Quoting Greene and Hensher (2003), just like Mixed Logit, the latent class models history starts before the one generally reported in transport literature. However, the first applications of the latent class model with Logit kernel for a multinomial choice seems owing to the works of Kamakura and Russel (1989), Swait (1994), Gupta and Chintagunta (1994) and the PhD thesis of Gopinath (1994).

The latent class models represent an interpretation of the finite sum in (2.56) assuming the membership to a segments *s* as a latent characteristic of the individuals. The segment is called class, with reference to an opportune combination of parameters in vector $\boldsymbol{\beta}^s$. When the analyst is interested in reproducing only taste variation, the indicator I in (2.56) can assume the simple MNL formula (LC-MNL). Thus, the resulting expression is very similar to the integral (2.51), except for the sum that replaces the integral:

$$p^{n,t}(j) = \sum_{s=1}^{S} \left\{ \frac{\exp\left[\dfrac{V_j^{n,t}(\boldsymbol{\beta}^s)}{\theta}\right]}{\sum_{i \in C^{n,t}} \exp\left[\dfrac{V_i^{n,t}(\boldsymbol{\beta}^s)}{\theta}\right]} \right\} \cdot p(\boldsymbol{\beta}^s) \tag{2.57}$$

wherein the parameters to be estimated are the S vectors $\boldsymbol{\beta}^s$ and (S-1) mass probabilities p($\boldsymbol{\beta}^s$) (called *class allocation probabilities*). Therefore, the number of parameters rises quickly when S rises. For a discussion on the correlations and the elasticities with reference to the taste parameters it can made reference to Hess et al. (2009). Thus, the latent class model can be difficult to estimate when the number of classes S exceeds three or four, especially when the segment is referred to a big number of parameters (Yuan et al., 2015). The number of classes itself should be a variable to estimate. Generally, a trial and error process is carried out and some validation tests, like the likelihood ratio test or Akaike Information Criterion (Louviere et al., 2000), is performed *a posteriori*.

Furthermore, the classic estimation procedures for searching the optimum can be instable with this finite mixture of models (Redner and Walker, 1984; McLachlan and Basford, 1988; Bhat, 1997). Bhat (1997) proposed the Expectation Maximization (EM) algorithm (Dempster et al., 1977) for such model, which has become a common practice. In the same work, he also

proposed an endogenous segmentation version of the (2.57). However, it is often referred under the semi-parametric approach (Hensher and Greene, 2003), thus it will be shown in Section 2.2.2.4.
Notwithstanding also the latent class model can exhibit some estimation issues, several cases in literature show it can be superior, in terms of fit, with reference to the random coefficient Logit (see Greene and Hesnher, 2003; Shen, 2009; Sagebiel, 2011 to cite someone).

Discrete mixture of MNL

The models in (2.57) can be made more flexible by relaxing the assumption that a segment s must have the parameter value within the vector $\boldsymbol{\beta}^s$. This kind of specification is generally referred in literature as discrete mixture, and, when taking an MNL as a kernel (DC-MNL), can be formalized as follows:

$$p^{n,t}(j) = \sum_{s_1=1}^{S_1}\sum_{s_2=1}^{S_2}\cdots\sum_{s_k=1}^{S_k}\prod_{i=1}^{K} p(\beta^{s_i})\cdot\left\{\frac{\exp\left[\frac{V_j^{n,t}(\boldsymbol{\beta}\ )}{\theta}\right]}{\sum_{i\in C^{n,t}}\exp\left[\frac{V_i^{n,t}(\boldsymbol{\beta})}{\theta}\right]}\right\} =$$
$$= \sum_{s_1=1}^{S_1}\sum_{s_2=1}^{S_2}\cdots\sum_{s_k=1}^{S_k} p(\beta^{s_1})\cdot p(\beta^{s_2})\cdot\ldots\ldots\cdot p(\beta^{s_k})\left\{\frac{\exp\left[\frac{V_j^{n,t}(\boldsymbol{\beta}\ )}{\theta}\right]}{\sum_{i\in C^{n,t}}\exp\left[\frac{V_i^{n,t}(\boldsymbol{\beta})}{\theta}\right]}\right\} \qquad (2.58)$$

Indicating with $\boldsymbol{\beta}$ the generic combination of the K coefficients for which the heterogeneity is assumed, each one with its mass point β^{si} and $p(\beta^{si})$ is the mass probability of each one of them. The product of the mass probabilities represents the mass probability of observing the segment with the combination of that β^{si}.
This surely generalizes the simpler Latent Class MNL, but it also introduces a big number of adding parameters to estimate. In this case, the number of mass points to estimate is always S times K, but the number of mass probabilities is (S-1) for each parameter, i.e. a total of K(S-1). It means that, with only 3 parameters and 4 classes, a latent class model needs 12 mass points and only 3 mass probabilities to be estimated, while a discrete mixture of MNL needs always 12 mass points but 9 mass probabilities to be estimated. Furthermore, the sum in (2.58) becomes larger. In the example mentioned, with the LC-MNL the sum consists of 4 terms, while in the DM-MNL case it involves 12 terms. This becomes quickly impractical when the number of classes or random parameters rise.

Variation on the theme

Several variations have been presented in literature, with reference to the model described in this sub-section.
For example, Bajari et al. (2007) and Train (2008) proposed a finite mixtures distribution where the mass points were fixed. This led to a multi-dimensional grid in the coefficients space, wherein all mass points were independent variables with values fixed by the analyst. The advantage is that only mass probabilities need to be estimated. Train (2008) successfully estimated a multi-dimensional discrete distribution with 233.280 mass points. The question arises on how the prior analyst choice about the location of the mass points may influence the estimation of the model. Successive works (Bastani and Weeks, 2013; Train, 2016) tried to face

this problem by fixing exogenously the grid, after a prior estimations carried out on smaller grids.
Dong and Koppelman (2014), instead, tried to fix endogenously the grid and the mass probabilities. Starting from this framework, a recent work of Vij and Krueger (2017) implemented a Mixed Logit with a nonparametric multi-dimensional grid in the coefficient space, with equal or unequal intervals between successive points along the same dimension.

Posterior analysis with discrete distributions

The Bayesian rules can be applied also when estimations are performed on discrete mass probability functions, as in the LC-MNL and DC-MNL models case. Analogously to MMNL model, wherein it is interesting to compute the conditional likelihood $L(\boldsymbol{\beta}/y^n,\boldsymbol{\Phi})$, here it is interesting to compute the conditional probability of an individual *n* belonging to a class *s*, given his previous choices y^n. Indicating this probability as $p^n(\boldsymbol{\beta}^s/y^n)$, for the LC-MNL it is:

$$p^n(\boldsymbol{\beta}^s / y^n) = \frac{p^n(y^n / \boldsymbol{\beta}^s)\cdot p^n(\boldsymbol{\beta}^s)}{p^n(y^n)} \tag{2.59}$$

On the right side, $p^n(y^n / \boldsymbol{\beta}^s)$ is the conditional probability of observing the sequence of choices y^n, given the vector of values $\boldsymbol{\beta}^s$, i.e. the Logit formula that appears in the sum in (2.57) and (2.58), $p^n(\boldsymbol{\beta}^s)$ is the mass probability of the vector $\boldsymbol{\beta}^s$ (estimated) and the denominator is simply the unconditional LC-MNL individual choice probability. In this case, the superscript *t* is suppressed by virtue of fact that the probability is referred to the sequence of all the individual choices *t*. In other words, here $p^n(y^n)$ is:

$$p^n(y^n) = \prod_{t=1}^{T^n} p^{n,t}[i(t)] \tag{2.60}$$

where $p^{n,t}[i(t)]$ is the choice of individual *n* in the choice situation *t*.
The advantage of (2.59) is that it does not involve any integral simulations. Its generalization for DC-MNL is easy, but in this case, the probability does not refer to a class *s*, but to a generic vector $\boldsymbol{\beta}$ containing a combination of values β^{s^j}.

2.2.2.4 *Semi-nonparametric approach*

In Sections 2.2.2.2 and 2.2.2.3 the advantages and limitations of parametric and nonparametric approaches have been deepened. It is clear that, choosing one is generally based on trying both of them and verifying the one who provides the best goodness of fit. The parametric approach gives the advantage of estimating a continuous distribution, but needs performing simulation, and making a precise choice of the shape of distribution, with the impossibility to handle particular realistic shapes of distribution (e.g. multi-modal). The nonparametric approach theoretically allows mimicking any kind of distribution but, when the required precision rises, it wrestles with the operative constraints of estimating an high number of parameters. Thus, it can be inadequate "to capture the full extent heterogeneity in the data" (Allenby and Rossi, 1998). Thus, it is plausible trying to merge the advantages of both approaches in a unified framework (*semi-nonparametric approach*). In fact, also a finite mixture of continuous distributions theoretically allows handling any shape of distribution.
This section describes the main formulations adopting this approach, with reference to the applications to the models already analysed.

Mixed Logit as mixture of parametric distributions

Generally, the semi-nonparametric approach refers the applications of Mixed Logit with a distribution built on some transformations of parametric distributions. In the last ten years, several interesting examples have been proposed. Although the first applications of Mixed Logit with mixtures of Normal densities were referred as nonparametric Mixed Logit (Fosgerau and Hess, 2007) it will be referred here as semi-nonparametric, because it represents a combination of parametric distributions.
A very comprehensive state of the art of the recent works on this topic is reported in Vij and Krueger (2017). Fosgerau and Hess (2009) and Bujosa et al.(2010) combine mono-variate Normal distributions, while Greene and Hensher (2013) combine mono-variate triangular distributions. An interesting proposal for mono-variate distributions is due to Fosgerau and Mabit (2013), who propose this framework for drawing from a density of a parameter with an unknown distribution:

$$\beta_{i,k}^{n,t,r} = \sum_{h=0}^{H} a_h \cdot \left(z_{i,k}^{n,t,r} \right)^h \tag{2.61}$$

Wherein the generic r^{th} draw for $\beta_{i,k}^{n,t,r}$ is obtained by combining Normal draws, as a function of coefficient a_h to estimate in a power series. The (2.61) includes the Multinomial Logit case (h=0) and the Normal Mixed Logit case (h=1). The use of h≥2 for even numbers, tends to add only strictly positive values to the distribution, making more appropriate the shape of distribution in the case a positive distribution (or negative if the estimated coefficient is <0) is expected. The (2.61) is an example of the so called *sieve estimators*, i.e. estimators that tries to approximate a function whose distribution is unknown, with a series of basic functions. Unfortunately, the (2.61) allows only for considering mono-variate distributions.
Other applications of this concept are the Legendre polynomials (Fosgerau and Bierlaire, 2007), the cubic B-Spline (Bastin et al., 2010), but they are applicable only to mono-variate distributions, as in the previous case.

Latent class MNL with endogenous segmentation

As mentioned in Section 2.2.2.3, Bhat (1997) proposed an endogenous segmentation version of the (2.57), for computing the mass probability of each $\boldsymbol{\beta}^s$. Particularly, he computed the mass probability p($\boldsymbol{\beta}^s$) as a function of socio-demographic variables. The fact that the distribution of the $\boldsymbol{\beta}$ depends on observable attributes, leads often to classify it under the semi-parametric approach, and this the reason why it is presented in this sub-section. Particularly, (2.57) becomes:

$$p^{n,t}(j) = \sum_{s=1}^{S} \frac{\exp\left[\dfrac{V_j^{n,t}(\boldsymbol{\beta}^s)}{\theta}\right]}{\sum_{i\in C^{n,t}} \exp\left[\dfrac{V_i^{n,t}(\boldsymbol{\beta}^s)}{\theta}\right]} \cdot \frac{\exp\left[\dfrac{\boldsymbol{\gamma}^s \cdot z^{\,n})}{\theta}\right]}{\sum_{s=1}^{S} \exp\left[\dfrac{\boldsymbol{\gamma}^s \cdot z^{\,n})}{\theta}\right]} \tag{2.62}$$

Wherein γ^s is the vector of other coefficients that are relative to the class of people s and z^n is the vector of the individual socio-demographic and trip related variables. In this case, the advantage is that class allocation probabilities are individual-specific. Often, the latent class models are used under this framework.

Logit-Mixed Logit model

Train (2016) proposed the so called Logit Mixed Logit model, unifying two frameworks, i.e. the parametric nature of the Mixed Logit and the nonparametric nature of the latent class Logit. Particularly, the concept is that a Mixed Logit consists of two parts: the conditional individual Logit formula given a vector of values for $\boldsymbol{\beta}$ and the density function $f(\boldsymbol{\beta}/\boldsymbol{\Phi})$ (see Section 2.2.2.2 for notation). This means assuming a continuous cumulative function $F(\boldsymbol{\beta}/\boldsymbol{\Phi})$ whatever shape of distribution is hypothesized. Train assumed this cumulative function as a set of pre-fixed mass points, computing the mass probability of each mass point with the Logit formula.
The definitive formula is:

$$p^{n,t}(j)=\sum_{s=1}^{S}\frac{\exp\left[\frac{V_j^{n,t}(\boldsymbol{\beta}^s)}{\theta}\right]}{\sum_{i\in C^{n,t}}\exp\left[\frac{V_i^{n,t}(\boldsymbol{\beta}^s)}{\theta}\right]}\cdot\frac{\exp\left[\frac{\boldsymbol{\alpha}'\cdot g(\boldsymbol{\beta}^s)}{\theta}\right]}{\sum_{s'=1}^{S}\exp\left[\frac{\boldsymbol{\alpha}'\cdot g(\boldsymbol{\beta}^{s'})}{\theta}\right]} \tag{2.63}$$

Wherein $\boldsymbol{\alpha}'$ is a vector of parameters of the second Logit formula and they have to be estimated, while $g(\boldsymbol{\beta}^{s'})$ is a vector-valued function of $\boldsymbol{\beta}^{s'}$ itself.
The framework seems very similar to Bhat (1997), but it has been noted that Bhat expressed the mass probability as endogenous function of socio-demographic and trip specific attributes, while Train proposes a mass probability that is a function of the location of parameters themselves. He also demonstrated how this framework can accommodate, in a boundary case, the common Normal, log-Normal, step functions and splines.
This framework is still an open topic in the literature and recently Bansal et al. (2018) generalized it for the simultaneous presence of fixed and random parameters.

Latent class as a finite mixture of continuous distributions (Latent class Mixed Logit)

The Mixed Logit, whatever distribution assumed, does not allow for a multi-modal distribution, because of its nature of parametric model. This can be accommodated by assuming as a kernel of the Mixed Logit the latent class Logit formula (Bujosa et al., 2010; Greene and Hensher, 2012). This means, once again, mixing two approaches, namely the parametric approach of the MMNL and nonparametric of the LC-MNL. This can be viewed, on the contrary, also as a LC model with an MMNL inside each class.
The resulting probability statement is:

$$p^{n,t}(j)=\sum_{s=1}^{S}\int_{\boldsymbol{\beta}^s}\frac{\exp\left[\frac{V_j^{n,t}(\boldsymbol{\beta}^s)}{\theta}\right]}{\sum_{i\in C^{n,t}}\exp\left[\frac{V_i^{n,t}(\boldsymbol{\beta}^s)}{\theta}\right]}\cdot f(\boldsymbol{\beta}^s/\boldsymbol{\Phi}^s)\cdot \mathrm{d}\boldsymbol{\beta}^s \tag{2.64}$$

wherein the meaning of all parameters has been already clarified, except for $f(\boldsymbol{\beta}^s/\boldsymbol{\Phi}^s)$ that is the continuous distribution within each class. This can give an high improvement when searching for multi-modalities of taste distribution within the sample. However, this could imply very hard computations, because of the presence of: more parameters (each set $\boldsymbol{\Phi}^s$ of parameters of each class distribution), integral simulation and the already mentioned problems of the LC model with reference to the classical gradient-based optimization algorithms.

2.2.2.5 Variation on a theme – the scale heterogeneity

Considering the general framework (2.5), it is easy to recognize that all the randomness within the formula represents a lack of knowledge of the analyst. The relative importance of the random terms with reference to the perceived utility says how much the analyst is far from a perfect knowledge (and so, a deterministic scenario) of the behavioural phenomenon. In other words, when the $\varepsilon_i^{n,t}$'s rise, or the $\beta_{i,k}^{n,t}$'s are farther from their mean value, the individual *n*, for that choice task *t* is more random from the analyst standpoint. It is clear that, in general, an analyst has not an equal knowledge about the individuals and choice tasks specific choices. This aspect is known, in literature, as *scale heterogeneity*. Thus, the latter is directly related to the error variance of the random terms, that, in turn, appear within a ratio with the systematic utility (see Logit formula with the exponential of $V_i^{n,t}/\theta$). Consequently, all the models present a ratio among the marginal utility and a variance coefficient (that is the inverse of the scale).

A way to consider the scale heterogeneity is assuming a *scaled Mixed Logit model* (*S-MMNL*), i.e. a random coefficient Logit (2.51) with fixed coefficient and θ variable across individuals and choice tasks. It is easy to understand that it represents a very strong assumption, imposing unrealistic homogeneous marginal utilities across the observations.

Several researchers (Louviere et al., 1999; Fiebig et al., 2010; Greene and Hensher, 2010; Hess and Rose, 2012; Hess and Train, 2017) asked themselves how much of the randomness lies in the differences in tastes and how much lies in the differences in scale. The first one refers the $\beta_{i,k}^{n,t}$, while the second one refers the error variance (for example, the θ within the Mixed Logit formula). Assuming a homogenous scale means potentially overstating the taste heterogeneity in estimation, because all the heterogeneity will refer the marginal utilities.

The assumption of Multivariate distribution for the marginal utilities is particularly suitable to prevent the confounding effects between random taste heterogeneity and scale heterogeneity. Although it is impossible to disentangle between the two phenomena, estimating the correlations between two coefficients can say how two parameters tend to increase or decrease together. It can help the analyst to understand how much of this variation is due to differences in tastes and how much to differences in scale. In fact, in general, some randomness is coefficient specific, while another part can be shared across the coefficients. A model that allows for correlated coefficients thus allows for estimating this effect by assuming only the coefficients as random, i.e. avoiding assuming the scale as random. As shown in (Hess and Train, 2017), this is a critical assumption that, under no circumstances allows for disentangling among them. For fixing the ideas, it can be considered that a greater/smaller variance (i.e. scale) for an individual implies a smaller/greater value of all the coefficients for that individual. When two coefficients are correlated, they tend to rise/decrease together. Thus, the scale heterogeneity is itself a form of correlation among the utility coefficients. Another way to view this phenomenon is thinking of the (2.5), considering the random term in as product of a random term and its standard deviation. Since the scale of the utility does not matter, the same is obtained by dividing it by the standard deviation. Thus (2.5) will present all the coefficients as divided by the standard deviation, which is equivalent to view all of them as multiplied by the same coefficient. It is easy to recognize that, in the same error component fashion, it induces a positive correlation among the random coefficient.

However, it is not possible to disentangle the various sources of correlation among coefficients. In fact, an individual could perceive as correlated two coefficients, not only for the influence of the unobserved factors. Thinking of transport problems, a person who is not willing to spend money could be more willing to spend time on board, i.e. the two coefficients can be (negatively) correlated per se.

Differently from the S-MMNL, the *Generalized Multinomial Logit model* (*G-MNL*; Keane, 2006; Fiebig et al., 2010; Greene and Hensher, 2010) defines a random coefficient as a combination of two parts, taking into account the two phenomena. But, the scale part is able to reproduce the whole correlation among the coefficients, without possibility to disentangle among the sources of correlation. Thus, the problem persists and the same applies to the Mixed Logit, both in preferences and in WTP space, to Latent class models and to another models called Scaled Adjusted Latent class (Magidson and Vermunt, 2005). In fact, none of these are able to disentangle the various forms of correlations. However, assuming a multivariate distribution for the coefficient with a full correlation matrix represents the best options to incorporate the effects of correlation due to scale heterogeneity and the other sources. The conclusions of this paragraph are entirely deepened in Hess and Train (2017), who give several suggestions for interpreting the results when different models are used.

2.2.3 Unified framework

The two main problems faced by this thesis have been presented in separate sub-sections, assuming a more restrictive framework for the general (2.5) for each of them, with the purpose of presenting them as isolate phenomena. It has just a representative value. In the real-world applications, it is practically impossible to disentangle between the two phenomena and, in general, between all the phenomena mentioned in Chapter 1. However, all the frameworks presented separately in Section 2.2.1 and Section 2.2.2 can be merged in an opportune way. In this section the most relevant applications of this unified framework models are briefly described. Thus, the more general (2.5) represents the assumption of this section.

Probit and Mixed Logit with joint EC/RC specification

It is immediate to recognize that the general integral in (2.15) lends itself immediately to the Multinomial Probit model, making the assumption that the joint density function f($\boldsymbol{\beta}$,$\boldsymbol{\varepsilon}$) refers a Multivariate Normal distribution. Adopting the independence between f($\boldsymbol{\beta}$) and f($\boldsymbol{\varepsilon}$) does not represent any loss of generality.
However, as widely described in the previous sections, the Multinomial Probit can be actually seen as a particular, restrictive, and quite undesirable case (from a computational standpoint) of the Mixed Logit. This is the main reason for what the MNP is "out of fashion" in all practical applications. The Mixed Logit surely represents the most used framework for capturing the effects of the error structure and the random taste heterogeneity.
When using a joint error component random coefficient formulation for it, the underlying assumption is that (2.5) becomes:

$$U_i^{n,t} = \sum_k \beta_{k,i}^{n,t} \cdot X_{k,i}^{n,t} + \omega_i^{n,t} + \gamma_i^{n,t} \qquad \forall i \in \mathrm{C}^{n,t} \tag{2.65}$$

Where, differently from (2.43), the parameters $\beta_{i,k}{}^{n,t}$. are individual and choice task specific, and differently from (2.50), the residuals involves also the Multivariate error term $\omega_i{}^{n,t}$. The integral takes the general form:

$$p^{n,t}(j) = \int_{\boldsymbol{\omega}}\int_{\boldsymbol{\beta}} \frac{\exp\left[\dfrac{V_j^{n,t}(\boldsymbol{\beta}) + \omega_j^{n,t}}{\theta}\right]}{\sum_{i \in C^{n,t}} \exp\left[\dfrac{V_i^{n,t}(\boldsymbol{\beta}) + \omega_i^{n,t}}{\theta}\right]} \cdot f(\boldsymbol{\omega}/\boldsymbol{\Omega}) \cdot f(\boldsymbol{\beta}/\boldsymbol{\Phi})\, d\boldsymbol{\omega}\, \mathrm{d}\boldsymbol{\beta} \tag{2.66}$$

This surely represents the most flexible parametric formulation. However, combining the two formulations in a unified framework means also combining all their problems above described. Particularly, adding more random parameters could worsen the computational issues of the error component. It is particularly relevant when the number of alternatives is high and the error components structure is as general as possible. Definitely, resorting to (2.66) can be computationally efficient and reliable in a context with a small number of alternatives, or when some constraint is applied to the covariance matrix of the $\boldsymbol{\omega}$ (see Section 2.2.1.3). In different cases, moving towards a different assumption on $\boldsymbol{\omega}$ and $\boldsymbol{\gamma}$ can be a best option.

Mixed GEV

The problems described in the previous sub-section makes the integral in (2.66) practically impossible to manage in several situations. Thus, it has been noted that, by suppressing $\boldsymbol{\omega}$ and assuming a MEV distribution for $\boldsymbol{\gamma}$, leads to a Mixed GEV formulation. In fact, reminding the logic behind the (2.44), the assumption leads to the general choice probability statement:

$$p^{n,t}(j) = \int_{\boldsymbol{\beta}} L^{n,t}(j/\boldsymbol{\beta}) \cdot f(\boldsymbol{\beta}/\boldsymbol{\Phi}) \cdot \mathrm{d}\boldsymbol{\beta} \tag{2.67}$$

Where the integral in $\boldsymbol{\beta}$ still holds, but the integral in $\boldsymbol{\omega}$ disappears and the integral in $\boldsymbol{\gamma}$ resolves in $L^{n,t}(j/\boldsymbol{\beta})$, representing the conditional GEV likelihood of the alternative j, for individual n and choice task t, given the values of vector $\boldsymbol{\beta}$. It can be particularly useful when one wants to catch both the phenomena, without increasing the number of random terms. In other words, the assumption of whatever density function $f(\boldsymbol{\beta}/\boldsymbol{\Phi})$ deals with the random taste heterogeneity, while the $L^{n,t}(j/\boldsymbol{\beta})$ deals with catching the correlations between alternatives.
A first application on this framework is due to Bhat and Guo (2004), who proposed a so-called *Mixed Spatially Correlated Logit*. This model consisted of a Mixed Pair Combinatorial formulation, wherein the alternatives (residences) were assumed to be correlated or not if they were contiguous. Thus, the specification related a not flexible correlation pattern and, furthermore, they estimated a unique parameter (the dissimilarity parameter of the PCL), with a very poor gain in goodness of fit with reference to a simple MMNL. Sener et al. (2011) reprised the model relaxing the assumption on the contiguity as a condition to consider two alternatives as correlated. Hess et al. (2005*a*) tried to use a Mixed NL and a Mixed CNL formulation on the Swiss Metro stated survey, wherein the three alternatives involved were the train, the car and the hypothesis to use the Swiss Metro. The mixed models were specified adding a few parameters with reference to the MMNL. However, the Mixed NL exhibited a good improvement in goodness of fit, while the Mixed CNL, although better than MMNL, did not go better than the Mixed NL. The Mixed NL formulation, indeed, seems to be the preferred solution, until now, because of the problems of the complexity in estimation introduced by a Mixed CNL formulation. For other examples of applications of random coefficient NL the reader can refer to Teye et al. (2014), Cheng and Yang (2015), Haghani et al. (2015).
It has been noted that the random coefficient specification, per se, induces also heteroscedasticity and correlation among perceived utilities. Thus, this framework adds generality to the correlation reproduced.
Furthermore, although apparently not so interesting, the (2.67) can be further generalized, adding the presence of the Multivariate Normal error term $\omega_i^{n,t}$:

$$p^{n,t}(j) = \int_{\boldsymbol{\omega}}\int_{\boldsymbol{\beta}} L^{n,t}(j/\boldsymbol{\beta},\boldsymbol{\omega}) \cdot f(\boldsymbol{\omega}/\boldsymbol{\Omega}) \cdot f(\boldsymbol{\beta}/\boldsymbol{\Phi}) \cdot \mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{\beta} \tag{2.68}$$

The (2.68) could seem a very redundant formulation, since flexible correlation patterns are already captured by (2.67). However, apart from adding even more flexibility, an interesting particularization of (2.67) occurs when $\boldsymbol{\omega}$ is assumed to be independently distributed. In this way, it is possible to give an alternative specific variance to each perceived utility, thus improving the capability to handle heteroscedasticity. But, it is also a double edge sword. In fact, it is easy to recognize that adding an alternative specific variance term tend to reduce the correlation reproduced (the variances are present within the denominator of the Pearson correlation coefficient). Therefore, it is necessary to pay attention when using such formulation. The potential of a specification mixing GEV models will be proposed and deepened within Chapter 3.

Mixed Probit

Another way to implement a different formulation in a (2.67) fashion, is introducing a standard Multinomial Probit formula for the conditional likelihood $L^{n,t}(j/\ \boldsymbol{\beta})$. The first advantage is that a MNP is less difficult to estimate when it is specified as error component (Train, 2009), because there is no presence of i.i.d. Gumbel error term. The second advantage is that a conditional MNP formula for $L^{n,t}(j/\ \boldsymbol{\beta})$, when computing it with the GHK simulator, needs much less draws from the Multivariate density than an unconditional MNP formula (Train, 2009). However, it is not faster to estimate than a Mixed GEV model. Recently, Bhat and Sidhartan (2012) and Bhat and Lavieri (2018) analysed the Mixed MNP with different non-normally marginal coefficients distributions for different coefficients, and propose a methodology for estimating the model.

Latent class GEV

A natural way for extending the (2.57) and, in a sense, particularizing the (2.67) for the discrete case, is assuming a more flexible GEV formulation within the latent class model. Therefore, the formulation for the choice probability becomes, in general:

$$p^{n,t}(j) = \sum_{s=1}^{S} L^{n,t}(j/\boldsymbol{\beta}^s)\cdot p(\boldsymbol{\beta}^s) \tag{2.69}$$

being $L^{n,t}(j/\ \boldsymbol{\beta}^s)$ the conditional GEV likelihood for the alternatives j, individual n and choice task t, given the vector of parameters $\boldsymbol{\beta}^s$ of the class s. The first example of it is represented by the latent class NL (LC-NL) applied by Kamakura et al. (1996), followed by Swait (2003) and Bodapati and Gupta (2004). Most recently this framework has been reprised by Wen et al. (2011) and Oviedo and Yoo (2017). Wen et al. (2013) applied a Latent class GNL (LC-GNL) for modelling carrier choice, estimating it with a simple gradient based procedure for maximizing the log-likelihood and giving slight improvement in goodness of fit than LC-NL.

2.2.4 **Going further**

The previous sub-sections describe the random utility models under the framework of maximizing benefit behaviour. The rationale behind the RUMs lies in the assumptions that 1) an individual considers a discrete choice set of alternatives 2) investigates each alternative according to each observable attribute 3) expresses a preference for the alternative that maximizes his perceived benefit / utility and 4) makes the choice. However, there are many other interpretations of the choice problem in literature, spacing from economics, psychometric and many other disciplines contributes. In fact, preserving the assumptions 1), 2) and 4), there are other paradigms considered for modelling choice behaviour. For example, the *random regret* *minimization theory* (*RRM*; Chorus et al., 2008) considers the decision maker, in some circumstances, as a minimizer of regret, rather than a minimizer of benefit. It founds on the random regret theory under uncertainty (Loomes and Sugden, 1982), but it has been recently operationalized by Chorus et al. (2008) and Chorus (2010), in a way that enables the use of the same models described within the current section and the possibility to use smooth estimators for them. In fact, the framework is perfectly analogous to that of RUMs, but it replaces the systematic utility with a figure of merit termed *regret*, that represents the opposite concept. Furthermore, it considers the attributes of all the other alternatives within the regret of the considered alternative, as a way to incorporate all the trade-off effects among the various attributes playing a role. This ensures the non-occurrence of the IIA property, also when using a simple MNL model. Another interesting property of the RRM models lies in the fact that they are able to accommodate also semi-compensatory effects. In fact, in the classic RUM models, any attribute of quality may numerically compensate whatever other attribute of service. In other words, in the classic linear in parameters RUM framework, a surplus of an attribute induces an equal deficit into the utility of the other alternatives (becase only difference in utilities matters). In the RRM models, instead, this effect is alleviated, thanks to the consideration of the so-called binary regrets. Practically, the regret is expressed as a logarithm of all the cross differences among the attributes, each one weighted with coefficients that are consistent with the ones in (2.5). It is receiving a strong interest for transport studies and several recent contributes to generalize the RRM models have considered the possibility to build an hybrid RRM-RUM (Chorus et al., 2013), or a latent class RUM-RRM (Hess et al., 2012), i.e. a model that considers two classes of preference paradigms (respectively RRM and RUM) that are latent. Other interesting paradigms are the *Prospect theory* (*PT*; Kahneman and Tversky, 1979) and the *Elimination by Aspects theory* (*EBA*; Tversky 1972*a*,*b*). Their fundamentals will be briefly described in Section 2.3.5 for route choice applications. Another framework that is receiving a growing interest is represented by the *Decision Field Theory* (*DFT*; Busemayer and Townsend, 1992; 1993; Diederich, 1997; Roe et al., 2001). The DFT has been inherited from the mathematical psychology, and it is a dynamic cognitive model that relaxes the assumption of individual preference as stationary concept across the time. The fact of being a dynamic over time model allows treating those choice contexts characterized by risk or time pressure. In fact, substantially, the decision maker of the DCT makes choices in two ways: when he reaches an internal threshold value or when some external factors occurs, like the response time that finishes (Hanckock et al., 2018). The main assumption is that a preference at each instant t can be expressed as a function of the preference in the previous instant (t-1), weighted by a so-called feedback matrix, and a so-called valence vector at each instant t, which depends on the observable attributes. Thus, the preference at each time t can be expressed as a series of powers. This makes the model very general, but also very complex to manage, and its potential must be still investigated, although some comparison with RUM and RRM models has been already carried out (see Hanckock et al., 2018 for a comprehensive review of all the main case studies proposed in the recent literature), showing good performances of the DFT. Another theory that explores a different paradigm than 3) is the *Satisficing theory* (*ST*). It is based on the works of Simon (1955; 1956), who postulated three principles, particularly realistic in complex choice situations, like the ones characterized by a big number of alternatives. In fact, in these contexts, a decision maker tends to choose the first alternative that allows for a reference value of satisfaction or, in other words, the first that is "good enough". In such contexts, it takes a crucial importance the concept of *search order*, which decisively influences the final choice. The first principle states that a decision maker does not actually consider a continuous outcome (like the perceived utility), but rather some simplified outcomes, like the fact that the alternative is acceptable or not and, in the last case, he immediately discards it. The second principle relates the cost of the information. Substantially, it states that higher is the cost of the information of the choice context and simpler will be the cognitive process leading to the choice. It means that a decision maker does not actually consider all the attributes or alternatives (as well as EBA), thus relaxing the assumption 2), while it considers only a subset of the total choice set, strictly dependent upon the search order. The third principle states that some attributes are not easily comparable, such as quality and service attributes. Thus, according to ST, the decision maker considers the attributes independently, evaluating if the alternatives are acceptable or not with reference to them. The operationalization of a model consistent with all the three principles appears as an open research topic, but recently Gonzalez-Valdez and Ortuzar (2018) tried to provide a solution. However, this framework can be particularly appropriate when more than one alternative allows reaching the minimum degree of satisfaction, while in the other cases, other paradigms like RUM or RRM can be more suitable, given their operational simplicity.

When the hypothesis of discrete choice set of alternatives in 1) is not appropriate, one may resort to other continuous formulations, such as the mixed discrete continuous models (MDCEV). In fact, some transport choice dimensions (e.g., the departure time) can be directly considered as continuous choice. The most prominent contribution to the field is the paper of Bhat (2005).

Finally, when the main objective is not only reproducing choices, i.e. extending the 4), other paradigms are available. In fact, there are aspects other than choices an analyst could be interested in. In the last twenty years, there has been a growing interest in reproducing also the attitudes of the decision makers, to accommodate all the factors that are unknown not only to the analyst, but also to the decision makers themselves. This means relaxing the assumption the random utility models are founded on, i.e. perfect knowledge of the decision maker and imperfect knowledge of the analyst, allowing the decision maker being affected by some factors he ignores (*latent variables*). This implies an endeavour of cross fertilization among two different approaches, namely the econometric approach, more interested in prediction, and the psychometric approach, more interested in deconstruction of the cognitive process (Ben-Akiva et al., 1999; 2002). These kinds of models, so-termed *hybrid choice models (HCM)*, try to incorporate other psychological factors than the classic RUMs, such as history or latent psychological constructs (generally attitudes or perception). The classic RUMs are a simplified version of that, wherein the operational path starts from the observable explanatory variables, passes through the decision process (utility maximization) and arrives to the choice as final step. The HCM construct is more general, allowing to incorporate also latent variables or latent segmentations (like latent classes) in an unifying framework. They allows reproducing the choices and also some attitudinal indicators (e.g. risk attitude), and they are often termed *Integrated Choice and Latent Variable* (*ICLV*; Vij and Walker, 2016). However, their higher complexity is the main reason to accurately assess the gains and losses of using them rather than the classic, easier choice models. In fact, when the analyst is not so interested in understanding the unobservable factors behind the cognitive process, maybe it makes more sense resort to the choice models falling into the target 4)

2.3 Application of RUMs to route choice

The level of service (*generalized costs*) and the level of congestion (*flows*) of a transport system represent the main outcomes of a transport problem. Generally, the analysts is interested in computing these quantities for all the elemental components of the transport system. The topology of a transport network is modelled through a graph G≡{L,N}, wherein L is the set of links and N is the set of nodes of the network. Therefore, the elemental components are represented by the links belonging to the set L. The generalized costs and link flows computation are performed in an *assignment procedure*, representing a unique system of equations sharing all the modelling components equations of a transport system (*supply model equations* and *demand model equations*). A general behavioural concept is recognizable, namely that the supply system and the demand system interacts themselves. In other words, when the level of traffic increases, the costs increase but, in turn, the latter influence the choices of people and, thus, again, the level of traffic. This concept is the basis of the *User Equilibrium* problem. Thus, for computing the link flows, a general problem must be solved, involving all the mathematical relationships for computing the costs of each single component of the network.

The assignment approaches may be several. A link flow is the sum of trips that, at a disaggregate level, the users make for moving from an origin *o*, towards a destination *d*, with a transport mode m. This trip can be represented in various ways. The classic approach consists of assuming the total trip of each user as a sequence of links of the network, connecting the o-d pair, with o and d representing, respectively, the first and last nodes of the path[10]. This work will refer such sequence as *path* or *route* (the two terms will be used indifferently) and it will be indicated with the notation k. Furthermore, k will assume the meaning of *acyclic path*, i.e. a path that does not pass more than once through whatever node of the network. Thus, the computation of the link flows needs, in some way, the computation of the *path flows*. The latter represents a main prerogative of the demand models.

The travel demand estimation involves the simulation of several users choice steps, defined as *choice dimensions* (Cascetta, 2009). The following notation will be fixed with reference to the *explicit paths enumeration approach*. This approach needs the definition of each path with reference to each o-d pair, for the specific mode m whose the analysed network refers. Generally, with reference to the simple *four-stages model* (without any loss of generality), once computed the travel demand for each o-d pair, with reference to each transport mode m, one wants to compute the paths flows F_k for each path connecting the o-d pair with the mode m. The random utility models are generally applied for all the choice dimensions and, particularly, for the route choice.

Under the random utility framework, the perceived utility of a path can be surely assumed as a combination of parameters and attributes. Generally, the assumption made is:

$$U_k^{n,t} = -C_k^{n,t} + \varepsilon_k^{n,t} \qquad \forall k \in \mathrm{K}_{\mathrm{od}}^{n,t} \tag{2.70}$$

being $C_k^{n,t}$ [11] the systematic dis-utility fo the path k, here called *generalized path cost*, for individual n in choice task t, and $\mathrm{K}_{\mathrm{od}}^{n,t}$ represents the choice-set of paths connecting an o-d pair, by the individual n in the choice task t (it represents the particularization, for the route choice problem, of the $C^{n,t}$ of the previous section). The assumption made on the perceived utility is that:

[10] Another more complex representation, very common into the transit systems simulation, consists of assuming the trip within an hyper-path, i.e. a set of paths that the user considers, without a priori deciding which one will adopt. This kind of *travel strategy*, defined as en-route choice behaviour, considers the user adapting himself to the events happening. The hyper-path assumption can be used also for road systems, but the purpose of the thesis is analysing the route choice behaviour without en-route choices.Other approaches refer a trip in different, more simplistic ways (see Section 2.3.5 for some examples).

[11] The notation $C_k^{n,t}$ is quite general, although in the route choice literature, generally the superscripts n and t does not explicitly appear. However, the application of a route choice model at a disaggregate level on a panel dataset requires this distinction and I hold it. In the following, when presenting the route choice models, the two superscripts will be deleted for the sake of simplicity.

$$U_k^{n,t} = \sum_l a_{lk} \cdot u_l^{n,t} + U_k^{NA,n,t} \qquad \forall k \in \mathrm{K}_{\mathrm{od}}^{n,t} \tag{2.71}$$

in which two parts play a role. The first sum expresses the part of the path utility that is perceived as a sum of the perceived utilities of the links *l* belonging to the path *k* (*additive costs*), being a_{lk} the generic 0/1 element of the *link-path incidence matrix*[12]. The second term (*non-additive path cost*) represents the contribution to the utility of all the quantities that cannot be related to the links *l* (for example a fixed road toll, a quantity of crossing or lights that is bigger than a prefixed value, the *u*-turns, the left-turns and so on). Therefore, the random residual lies both in the additive and in the non-additive parts of the perceived utility. Furthermore, the variance of the perceived utility of the cost, indicating with the notation ε the random residuals of the components playing a role, is:

$$\begin{aligned} Var\left[U_k^{n,t}\right] &= Var\left[\sum_l a_{lk} \cdot u_l^{n,t} + U_k^{NA,n,t}\right] = \sum_l a_{lk} \cdot Var\left[u_l^{n,t}\right] + Var\left[U_k^{NA,n,t}\right] = \\ &= \sum_l a_{lk} \cdot Var\left[\varepsilon_l^{n,t}\right] + Var\left[\varepsilon_k^{NA,n,t}\right] \qquad \forall k \in \mathrm{K}_{\mathrm{od}}^{n,t} \end{aligned} \tag{2.72}$$

since a_{lk} are 0/1 element, so a_{lk}^2 is equal to a_{lk}. The same applies to the covariance among two routes *k* and *h*, assuming the random residual $\varepsilon_l^{n,t}$ between the various links being stochastically independent:

$$\begin{aligned} Cov\left[U_k^{n,t}, U_h^{n,t}\right] &= E\left[(U_k^{n,t} - V_k^{n,t}) \cdot (U_h^{n,t} - V_h^{n,t})\right] = \\ &= E\left[\varepsilon_k^{n,t}, \varepsilon_h^{n,t}\right] = E\left[\left(\sum_l a_{lk} \cdot \varepsilon_l^{n,t} + \varepsilon_k^{NA,n,t}\right) \cdot \left(\sum_l a_{lh} \cdot \varepsilon_l^{n,t} + \varepsilon_h^{NA,n,t}\right)\right] = \\ &= E\left[\sum_l a_{lk} \cdot \varepsilon_l^{n,t} \cdot \sum_l a_{lh} \cdot \varepsilon_l^{n,t}\right] + E\left[\varepsilon_k^{NA,n,t} \cdot \varepsilon_h^{NA,n,t}\right] = \\ &= \sum_l a_{lk} \cdot a_{lh} \cdot E\left[\left(\varepsilon_l^{n,t}\right)^2\right] + E\left[\varepsilon_k^{NA,n,t} \cdot \varepsilon_h^{NA,n,t}\right] = \\ &= \sum_l a_{lk} \cdot a_{lh} \cdot Var\left[\varepsilon_l^{n,t}\right] + Cov\left[\varepsilon_k^{NA,n,t}, \varepsilon_h^{NA,n,t}\right] \qquad \forall k,h \in \mathrm{K}_{\mathrm{od}}^{n,t} \end{aligned} \tag{2.73}$$

It has been noted that, when the non-additive component of the utility is considered to be null, the covariance reduces to the variance of the shared links among routes *k* and *h*.
The generalized path cost within (2.70) can be expressed as a function of the generalized link costs $c_l^{n,t}$, in the same way of (2.71):

$$C_k^{n,t} = \sum_l a_{lk} \cdot c_l^{n,t} + C_k^{NA,n,t} \qquad \forall k \in \mathrm{K}_{\mathrm{od}}^{n,t} \tag{2.74}$$

Each term $c_l^{n,t}$ generally includes the travel time and the monetary cost for the link *l*. In general, the link cost is a function of the link flows. This means assuming the following relationship:

$$c_l = c_l(\mathbf{f}) \qquad \forall l \in \mathrm{L} \tag{2.75}$$

being **f** is the (n_Lx1) vector of flows influencing the cost of the link *l*. The (2.75) is called *link cost function*. Assuming the *non-separable-cost function hypothesis*, (2.75) simplifies because **f** represents only f_l.
By virtue of what previously said, the generic f_l is computed with the so called *Network Flow Propagation* (*NFP*; Cascetta, 2009) model as:

[12] It represents the matrix **A** whose generic a_{lk} assumes value of 1 when the generic link *l* belongs to the path *k*, 0 otherwise.

$$f_l = \sum_k a_{lk} \cdot F_k \qquad \forall l \in \mathrm{L} \tag{2.76}$$

Narrowing it down to a specific transport mode m in the specific time interval h, the path flow F_k is the output of the demand model:

$$F_k = d_{\text{o-d}} \cdot p(k / \text{o-d}) \qquad \forall k \in \mathrm{K}_{\text{od}}^{n,t} \tag{2.77}$$

in which $d_{o\text{-}d}$ represents the generic o-d matrix entries for the underlying transport mode, while p(k/o-d) is the route choice probability applied at an aggregate level.
Indicating with $\mathbf{C}^{n,t}$ the vector of path costs that the individual n considers in the choice task t, a route choice model at disaggregate level is a relationship representing the probability of choosing k as a function of $\mathbf{C}^{n,t}$:

$$p^{n,t}(k) = p^{n,t}(-\mathbf{C}^{n,t}) \tag{2.78}$$

The assignment equilibrium system can be formalized as follows:

$$\begin{cases} \mathbf{C} = \mathrm{A}^{\mathrm{T}} \cdot \mathbf{c}(\mathbf{A} \cdot \mathbf{F}) + \mathbf{C}^{NA} \\ \mathbf{F} = \mathbf{P}(-\mathbf{C}) \cdot \mathbf{d}(-\mathbf{C}) \end{cases} \tag{2.79}$$

being:

- **C** is the (n_k x 1) vector of path costs;
- **A** is the (n_L x n_k) link-path incidence matrix;
- **c** is the (n_L x 1) vector of link costs;
- $\mathbf{C}^{NA}$ is the (n_k x 1) vector of non-additive path costs;
- **F** is the (n_k x 1) vector of path flows;
- **P** is the (n_k x $n_{o\text{-}d}$) matrix of path choice probabilities for each o-d pair ($n_{o\text{-}d}$ represents the total amount of o-d pairs)
- **d** is the ($n_{o\text{-}d}$ x 1) vector containing all the o-d matrix entries for the specific mode considered;

These general assumptions will be used in the following sub-sections, for describing the major issues of route choice modelling and the main route choice models proposed in the state of the art.

2.3.1 Route choice modelling: general issues

Concerning what said in the introductory part, the route choice problem represents the core of the assignment problem. However, several problems characterize this choice dimension, making it the more complex dimension to model in the demand analysis.
The first issue in choice modelling concerns the size of the network. A graph of a real network is generally modelled with several thousands of nodes and links, and several hundreds of *zones* or *centroids* (Cascetta, 2009), representing the origins and the destinations of the network. Thus, the complexities are manifold, because a big number of centroids implies a big number of o-d pairs and, at the same time, a big number of nodes and links implies a big number of feasible paths connecting each o-d pair. Consequently, the first relevant problem lies in modelling the real set $K_{od}^{n,t}$ considered by an individual n in a specific choice scenario t. In fact, it is not plausible that a user considers thousands of feasible paths when choosing his itinerary for moving from o to d. However, selecting a sub-set of paths could be a double-edge sword, because it reduces the complexity but it creates the risk of not including the real path that a user may chooses. Two main approaches are available in literature: the explicit paths enumeration and the implicit paths enumeration. The *explicit paths enumeration* needs the paths involved being explicitly detected as a sequence of links of the network. This means storing in memory a great

amount of data and increasing the computational time. The *implicit paths enumeration*, instead, provides the link flows f_l without need of path enumeration. This is possible by means of algorithms that, given a specific assumption on (2.78), are able to compute the link flows directly once computed the (2.75) and the $d_{o\text{-}d}$, totally avoiding the computation of (2.78) itself. Actually, these algorithms, whenever possible (and often, under some restrictive assumptions), operates a sequence of equations giving a perfectly equivalent result, in terms of flows, than one obtained when using explicitly (2.78). Each one of these approaches can be declined in two different ways: the *exhaustive approach* and the *selective approach*. The two approaches consider (explicitly or implicitly), respectively, the set of all feasible paths or an opportune sub-set of it. Generally, the explicit enumeration paths methodology uses the selective approach (for a comprehensive list of criteria for selecting paths see Cascetta, 2009), while the implicit enumeration methodology, depending on the model, can be used with both of them. In the following section, several example of models using the exhaustive or the selective approach will be described.

The second relevant problem is due to the overlapping among the routes. The actual perception of the alternatives can be strongly influenced by this aspect. In fact, although thousands of routes connecting an o-d pair may exist, the most of them will be not considered as perfectly distinct from each other, because overlapped to some degree. Therefore, from a behavioural standpoint, the users do not perceive the utilities of the alternatives as independent, but as stochastically correlated. Given the number of alternatives involved, the correlation structure may be very complex. As discussed in the next Section, there are route choice models that take into account deterministically, i.e. as a function of the observable attributes, the effects of the correlations, and route choice models that take into account them stochastically, i.e. as hypothesis on the error structure.

In summary, a route choice model should own the following desirable characteristics:

- A closed form expression for (2.78), in order to avoid computational burdens due to simulation;
- Flexibility in reproducing the overlapping effects on the choice probabilities;
- Easy computation of the structural parameters by means of the available observable information (i.e. low number of parameters to estimate);
- The possibility to implement an implicit enumeration algorithm for computing the link flows, , consistent with the model, in order to avoid the burdens of the preliminary paths enumeration;

This section described the peculiarities of the main route choice models available in literature. Section 2.3.2 describes the application of the models that take into account the overlapping effects deterministically, using the Multinomial Logit framework; Section 2.3.3 describes the main applications of models belonging to the GEV family; Section 2.3.4 describes the models involving simulation; in the end, Section 2.3.5 provides some references of alternatives approaches for the route choice problem.

2.3.2 Logit based route choice models

The MNL for route choice derives from the assumption of i.i.d. Gumbel distribution of perceived utilities of the paths belonging to the considered choice set. The general formula for the route choice probability is the application of (2.22) for route choice context:

$$p(k) = \frac{\exp\left(\frac{V_k}{\theta}\right)}{\sum_{k' \in \mathrm{K_{od}}} \exp\left(\frac{V_{k'}}{\theta}\right)} \tag{2.80}$$

wherein the superscript *n* and *t* will be suppressed, *k* automatically identify the o-d pair connected by it and θ is the variance parameter for that o-d. The MNL-based formulations will be described in the following sub-sections.

2.3.2.1 Multinomial Logit - Dial's algorithm

The MNL as presented in (2.80) is generally implemented by means of the Dial's algorithm (Dial, 1971). Assuming that of only additive link impedances exist, (2.80) becomes:

$$p(k) = \frac{\exp(\frac{-C_k}{\theta})}{\sum_{k' \in \mathrm{K_{od}}} \exp(\frac{-C_{k'}}{\theta})} = \frac{\exp\left(\frac{-\sum_l a_{lk} \cdot c_l}{\theta}\right)}{\sum_{k' \in \mathrm{K_{od}}} \exp\left(\frac{-\sum_l a_{lk'} \cdot c_l}{\theta}\right)} = \tag{2.81}$$

$$= \frac{\prod_l \exp\left(-\frac{a_{lk} c_l}{\theta}\right)}{\sum_{k' \in \mathrm{K_{od}}} \prod_l \exp\left(-\frac{a_{lk'} c_l}{\theta}\right)} = \frac{\prod_{l \in k} \exp\left(\frac{-c_l}{\theta}\right)}{\sum_{k' \in \mathrm{K_{od}}} \prod_{l' \in k'} \exp\left(\frac{-c_{l'}}{\theta}\right)} = \prod_{l \in k} p\left[l / h(l)\right]$$

Defining the generic p[l/h(l)] as the conditional probability of choosing the link l when coming from the node $h(l)$. Indifferently, (2.81) can be written as a function of p[l/t(l)] with analogous meaning.

The algorithm allows for computing link flows consistent with the (2.81), without need of explicitly enumerating the paths. Dial defined the following recursive equations:

$$\begin{cases} w_l = \exp\left(\frac{-c_l}{\theta}\right) \cdot W_{t(l)} \\ W_{t(l)} = \sum_{l' \in EBS[t(l)]} w_{l'} \end{cases} \tag{2.82}$$

The w_l's and $W_{t(l)}$'s take the names, respectively, of Dial link weights and node weights. As better described in Section 6.2.1, these assumptions lead to the Logit formula in (2.81). Dial provided a two steps - algorithm for the computation of the weights and the link flows. The first step consist of computing the weights in (2.82). The second step allows computing the link flows with the following recursive equations:

$$\begin{cases} p[l / h(l)] = {w_l}/{W_{h(l)}} \\ f_l = f_{h(l)} \cdot p[l / h(l)] \\ f_{t(l)} = \sum_{l' \in EFS[t(l)]} f_{l'} \end{cases} \tag{2.83}$$

It has been noted that none of the (2.82) and (2.83) depend on any routes definition.

The algorithm works on a subset of the exhaustive set of paths connecting the considered o-d pair, named *efficient paths*. An efficient path, with reference to the origin *o*, is a sequence of *efficient links*, i.e. links *l* verifying the following condition:

$$C_{o-t(i)} < C_{o-h(l)} \tag{2.84}$$

Where $C_{o-t(l)}$ and $C_{o-h(l)}$ represent the minimum path costs for reaching, respectively, *t(l)* and *h(l)*, starting from the origin *o*. In other words, an efficient link allows for walking away from the origin *o*. The efficient network does not include a sub-set of links for each considered origin *o*, i.e. it is a sub-network of the whole network, given an origin *o*. The same concept applies with reference to the destination *d*. Particularly, a link is efficient with reference to a destination, when it allows for moving toward it. Applying both definitions means working on a sub-network that is *doubly efficient*, namely it is efficient with reference to the origin *o* and the destination *d*. Then, the resulting sub-network has a lower cardinality than the sub-network with the simple efficiency.

Thus, the first step of the algorithm allows computing the weights, given an o-d pair, for the efficient links *l*, with reference to that o-d pair, in a forward exploration of the network, starting from o. In the second step, the link flows are computed, by assigning the $d_{o\text{-}d}$ in a backward exploration of the network (from d). This is called *double-step Dial's algorithm*, where the definition "double-step" indicates that the algorithm makes the two exploration steps for each o-d pair. Indicating with n_C the number of centroids of the network, this algorithm has a computational complexity growing up with n_C^2.

A simplified version of the algorithm is often used in the real applications, namely the single-step Dial's algorithm, by simplifying the variance parameter for each o, rather than each o-d pair. Keeping fixed the variance parameter θ for each o-d pair, the procedure can be performed with reference to each origin *o*, considering simultaneously each destination *d*. In this case, it is sufficient applying the condition of efficiency with reference to the origin *o*. This algorithm's version has a computational complexity growing up with n_C.

More precisely, the algorithm works from each origin *o* as follows:

- As a pre-algorithm, a minimum path costs algorithm is performed;
- a vector of minimum path costs $\mathbf{C}_o$ and an ordered list of nodes L, with increasing minimum path cost from o, is computed;
- moving forward in the list, the first two equations of (2.82) are computed for each links of the forward star of the node and each head of the links;
- moving backward into the list, the link flows are computed by means of the equations (2.83), assigning all the $d_{o\text{-}d'}$ for each d' in an unique step;

In an equilibrium assignment, the efficient sub-network with reference to each origin *o* (or the o-d pair) must be kept fixed, to ensure an exact solution for the S.U.E.-Logit problem (Fisk, 1980; Van Vliet, 1981; Leurent, 1997). This means keeping fixed the minimum path costs into (2.84), without making them varying with the traffic congestion.

Unfortunately, the Dial algorithm has strong limitations. First, it implements a MNL formula, not allowing for overlapping effects among alternatives. Second, it restricts the choice set only to the efficient routes. Third, the single-step algorithm introduces the unrealistic assumption of equal variance parameter θ for each o-d pair (i.e. the absolute value of the dispersion of the costs is supposed to be the same when o and d are very close or very far). However, its very low complexity explains why, despite its big limitations, the MNL with the Dial's algorithm represents, even now, one of the most implemented route choice models in common assignment software.

2.3.2.2 C-Logit

As discussed in the introduction, a way to incorporate the effects of the overlapping is trying to incorporate them into the path systematic utility. The C-Logit model (Cascetta et al., 1996) assumes the Logit formulation, with the following expression within the systematic utility:

$$p(k)=\frac{\exp\left(\frac{-C_k}{\theta}-CF_k\right)}{\sum_{k'\in \mathrm{K}_{\mathrm{od}}}\exp\left(\frac{-C_{k'}}{\theta}-CF_{k'}\right)} \tag{2.85}$$

wherein the generic $CF_{k'}$ represents the so called *Communality Factor* for path *k'* [13]. The authors expressed the communality factor in different ways, but the most known is the first one:

$$CF_k=\beta_0\cdot\ \ln\left[\sum_{k'\in \mathrm{K}_{\mathrm{od}}}\left(\frac{L_{k-k'}}{\sqrt{L_k\cdot L_{k'}}}\right)^{\gamma}\right] \tag{2.86}$$

In which L_k, $L_{k'}$ and $L_{k-k'}$ represent, respectively, the size of path *k*, *k'* and of the overlapped portion between *k* and *k'*, and γ and β_0 are parameters to be estimated. Frequently γ is set to be 1. The β_0 is the parameter indicating the importance that the users give to the overlapping between the paths. It has been noted that each term of the sum in the logarithm is very similar to the Pearson correlation coefficient.

In the same work, the authors proposed also the formulations:

$$CF_k=\beta_0\cdot\ \ln\left(\sum_{l\in k}w_{lk}\cdot N_l\right) \tag{2.87}$$

$$CF_k=\beta_0\cdot\ \sum_{l\in k}w_{lk}\cdot\ln\left(N_l\right) \tag{2.88}$$

Wherein w_{lk} is the ratio between the length of *l* and the total length of *k*, while N_l is the number of paths sharing *l*. Particularly, the (2.88) is an interesting formulation, on what Russo and Vitetta (2003) proposed an implicit enumeration algorithm, in a Dial fashion. In fact, it can be obtained by correcting the first of the (2.82) as:

$$w_l=\left(N_l\right)^{w_l{}^*}\cdot\exp\left(\frac{-c_l}{\theta}\right)\cdot W_{t(l)} \tag{2.89}$$

The $w_l{}^*$ expresses the ratio between the length of *l* and the length of the minimum path, while N_l has the meaning already clarified. However, the latter has been computed through another double exploration of the network, but always in an implicit way. Although a simplification has been proposed by the authors, allowing for a single-step procedure, Marzano (2006) showed that the results are not always satisfactory.

Another formulation has been proposed by Cascetta and Papola (1998), as a generalization of (2.86):

$$CF_k=\beta_0\cdot\ \ln\left[1+\sum_{\substack{k'\in \mathrm{K}_{\mathrm{od}}\\ k\neq k'}}\left(\frac{L_{k-k'}}{\sqrt{L_k\cdot L_{k'}}}\cdot\frac{L_k-L_{k-k'}}{L_{k'}-L_{k-k'}}\right)^{\gamma}\right] \tag{2.90}$$

The C-Logit formulation for S.U.E. programming has been investigated in Zhou et al. (2012) and Xu and Chen (2013).

[13] The original formulation (2.85) did not contain the variance parameter θ. This is a dimensionally corrected formulation, derived from Marzano (2006).

2.3.2.3 *Path-Size Logit*

On the heels of the C-Logit, another Logit formula with deterministic correction for overlapping has been proposed first by Ben-Akiva and Ramming (1998), but often attributed to Ben-Akiva and Bierlaire (1999*b*). The model, perfectly analogous in the framework to (2.85) proposed the so-called *path-size factor*, derived from notion of elemental alternatives and size variables (Ben-Akiva and Lerman, 1985):

$$p(k) = \frac{\exp\left(-\frac{C_k}{\theta} + \beta_{PS} \cdot \ln PS_k\right)}{\sum_{k' \in K_{od}} \exp\left(-\frac{C_{k'}}{\theta} + \beta_{PS} \cdot \ln PS_{k'}\right)} \tag{2.91}$$

where the PS_k is computed as (Ben-Akiva and Bierlaire, 1999*b*):

$$PS_k = \sum_{l \in k} \frac{L_l}{L_k} \cdot \frac{1}{\sum_{k \in K_{od}} \frac{L_{C_{\text{o-d,min}}}}{L_k} \cdot a_{lk}} \tag{2.92}$$

where $L_{C_{\text{o-d,min}}}$ represents the length of the path with minimum cost connecting the o-d pair. A generalized expression of (2.92) is due to Ramming (2002) and Hoogendorn-Lanser et al. (2005), with the introduction of a coefficient γ to be estimated:

$$PS_k = \sum_{l \in k} \frac{L_l}{L_k} \cdot \frac{1}{\sum_{k \in K_{od}} \left(\frac{L_{C_{\text{o-d,min}}}}{L_k}\right)^{\gamma} \cdot a_{lk}} \tag{2.93}$$

However, the introduction of the γ adds burdens in the estimation process and the behavioural interpretation of the value estimated, often within a range of 10 and 15 (Prato, 2009), can be difficult. Bovy et al. (2007) proposed the following logarithmic expression for the so-called *Path Size Correction*:

$$PSC_k = \sum_{l \in k} \frac{L_l}{L_k} \cdot \ln\left(\frac{1}{\sum_{k \in K_{od}} a_{lk}}\right) \tag{2.94}$$

Turning (2.91) into:

$$p(k) = \frac{\exp\left(-\frac{C_k}{\theta} + \beta_{PS} \cdot PSC_k\right)}{\sum_{k' \in K_{od}} \exp\left(-\frac{C_{k'}}{\theta} + \beta_{PS} \cdot PSC_{k'}\right)} \tag{2.95}$$

It has been noted that the PSC_k varies within the range of all negative values, while the original PS_k varies from 0 to 1.
Finally, holding (2.91), Frejinger et al. (2009) proposed a correction of the (2.93) called the individual *Expanded Path Size factor*, as:

$$EPS_k^n = \sum_{l \in k} \frac{L_l}{L_k} \cdot \frac{1}{\sum_{k \in K^n_{od}} \Phi_{k,n} \cdot a_{lk}} \tag{2.96}$$

being the choice-set considered by the individual n, and $\Phi_{k,n}$ is the *expansion factor* defined as a function of a sampling protocol, depending on the individual probability of considering k in the choice-set K^n_{od}.

The S.U.E formulation for the PS-Logit has been analysed in Chen et al. (2012).

2.3.2.4 IAP Logit

The framework of the *Implicit Availability Perception Logit* (Cascetta and Papola, 2001; Cascetta et al., 2002) is analogous to the C-Logit and PS-Logit ones. However, the target is different. While (2.85) and (2.91) occur to reproduce the effects of the physical overlapping between the routes, the IAP Logit want to simulate the effect of the perception of an alternative within the individual choice set. This effect is taken into account by means of a degree of membership μ_k^n of the alternative k in the individual choice-set K^n_{od}, adopting a fuzzy logic.
Particularly, the formulation of the route choice probability is:

$$p^n(k) = \frac{\exp\left[\dfrac{-C_k^n + \ln(\mu_k^n)}{\theta}\right]}{\sum_{k' \in K^n_{od}} \exp\left[\dfrac{-C_{k'}^n + \ln(\mu_{k'}^n)}{\theta}\right]} \tag{2.97}$$

The μ_k^n represents a latent random variable whose individual value is unknown for the analyst. By expressing it as a function of a vector **Y** of availability/perception attributes, and expressing its expected value as Taylor series stopped to the second term, the (2.97) can be expressed as:

$$p^n(k) = \frac{\exp\left(\dfrac{-C_k^n + \ln(\bar{\mu}_k^n) - \dfrac{1-\bar{\mu}_k^n}{2\cdot\bar{\mu}_k^n}}{\theta}\right)}{\sum_{k' \in K^n_{od}} \exp\left[\dfrac{-C_{k'}^n + \ln(\bar{\mu}_{k'}^n) - \dfrac{1-\bar{\mu}_{k'}^n}{2\cdot\bar{\mu}_{k'}^n}}{\theta}\right]} \tag{2.98}$$

The IAP correction occurs to penalize the path utility for which the users are not much aware of. However, no S.U.E. formulations have been proposed for the IAP Logit until today.
Definitely, several modified Logit formulations exists, with the aim of taking into account some stochastic effects in a deterministic way, by correcting the systematic utility of a route. But, the lack of theoretical assumptions behind these models, often is the cause of counter-intuitive results provided by them (see Prashker and Bekhor, 1998; Prashker and Bekhor, 2004; Marzano, 2006, Frejinger and Bierlaire, 2007; Papola and Marzano, 2013; Papola et al., 2018).

2.3.3 **GEV models for route choice**

The GEV class, as described in Section 2.2.1.1, is a very flexible and useful class of homoscedastic random utility models, characterized by a closed form of the choice probabilities. Several applications of GEV models have been proposed in literature for route choice but, unfortunately, they did not receive a strong interest in the practical applications. As described in the previous sub-section, the Logit model, with or without correction accounting for overlapping of other phenomena, represents an appealing formulation, due the simplicity of its framework. In fact, a very low number of parameters must be estimated. Conversely, moving toward more complex specifications, to take into account the overlapping problem under stochastic hypothesis on unobservable components, needs a practical way for computing the structural parameters of the model. In fact, given the number of the involved o-d pair and the potential very complex specification of the error structure, the number of parameters can easily overcome the thousands of units. This means that, although the closed form statement for the probabilities, the main problem of such formulations is that they may not be operative. Thus, for making practical the use of these models, the computation of the parameters must be necessarily dependent on the network and its observable quantities. This is the reason why, in the route choice modelling, often the term "route choice model" implies a peculiar specification for the general model to which it is related. For this purpose, some significant route choice GEV operationalisations will be described in the following sub-sections.

The first two mentioned models are derived under the Cross Nested Logit framework, the third one refers the more general Network GEV model and the fourth one is a recent development of the Weibit model for route choice.

2.3.3.1 Link Nested Logit

The Link Nested Logit (LNL; Vovsha and Bekhor, 1998) represents the natural particularization of the Cross Nested Logit for the route choice. In fact, an intuitive specification of a CNL can be obtained by considering all the links l of the network as nests, and the routes k sharing a link l as alternatives belonging to the nest l. In this way, the inclusion parameter α_{kl} for the route k in the nest l can be computed as the ratio:

$$\alpha_{kl} = \frac{c_l}{C_k} \cdot a_{kl} \tag{2.99}$$

The presence of a_{lk} on the right side clearly indicate that a route k is allocating into nest l if and only if k includes l, and clearly (2.99) is consistent with the constraint (2.33), assuming in the latter a value b=1. In the original work, in addition to provide the S.U.E. formulation for the LNL, the authors have been limited to analyse the boundary case of δ_l=0 for each nest. This particular allows for a stochastic network loading procedure with the implicit path enumeration, by considering that the nest probability becomes:

$$\begin{aligned} p(l) &= \lim_{\delta_{l'} \to 0} \frac{\left[\sum_{k\in l} \alpha_{kl}^{1/\delta_l} \cdot \exp(-C_k / \theta_l)\right]^{\delta_l}}{\left[\sum_{k'\in l'} \alpha_{k'l'}^{1/\delta_{l'}} \cdot \exp(-C_{k'} / \theta_{l'})\right]^{\delta_{l'}}} = \frac{\alpha_{kl_l} \cdot \exp(-C_{k_l} / \theta)}{\sum_{l'\in \mathrm{L}} \alpha_{kl'_{l'}} \cdot \exp(-C_{k_{l'}} / \theta)} = \\ &= \frac{\exp(-C_{k_l} / \theta + \ln \alpha_{kl_l})}{\sum_{l'\in \mathrm{L}} \alpha_{kl'_{l'}} \cdot \exp(-C_{k_{l'}} / \theta + \ln \alpha_{kl'_{l'}})} \end{aligned} \tag{2.100}$$

wherein the further subscript l into the path costs and the inclusion parameters indicate, respectively, the minimum cost of the minimum path passing through the link l and the inclusion parameter of the minimum cost passing through the link l. Furthermore, the conditional route choice probability is:

$$p(k) = \lim_{\delta_l \to 0} \frac{\alpha_{kl}^{1/\delta_l} \cdot \exp(-C_k / \theta_l)}{\sum_{k'\in l} \alpha_{k'l}^{1/\delta_l} \cdot \exp(-C_{k'} / \theta_l)} = \begin{cases} 1/n_l & if\ C_k = \min_{k'\in \mathrm{K}_{\mathrm{od},l}} \{C_{k'}\} \\ 0 & \end{cases} \tag{2.101}$$

Where n_l is the number of paths verifying the minimum cost condition on the right side, while $\mathrm{K}_{\mathrm{od},l}$ is the set of paths connecting the o-d pair passing through l.

The expressions (2.100) and (2.101) allow for a double-step procedure. However, the implications of using $\delta_l \to 0$ can be undesirable, as deepened in Section 6.3.

For obviating to this lack for the nesting parameters, Bekhor and Prashker (2001) proposed another expression, expressing them as a function of the inclusion parameters:

$$\delta_l = 1 - \frac{1}{n_l} \sum_{k \in \mathrm{K}_{\mathrm{od}}} \frac{c_l}{C_k} \cdot a_{kl} \tag{2.102}$$

where the nesting parameter of l is supposed to decrease with the presence of multiple paths sharing the link l. Another formulation is proposed in Marzano (2006), computing a geometric mean, instead of the arithmetic mean of the previous formulation:

$$\delta_l = 1 - \sqrt{\sqrt[n_l]{\prod_{k \in \mathrm{K}_{\mathrm{od}}} \frac{c_l}{C_k} \cdot a_{kl}}} \tag{2.103}$$

The S.U.E. problem with the LNL model has been analysed in Bekhor and Prashker (1999), Bekhor and Prashker (2001), Prashker and Bekhor (2004) and Bekhor et al. (2008).

2.3.3.2 Pair Combinatorial Logit

The Pair Combinatorial Logit (PCL) specification, although not appealing for many other discrete choice applications, principally because of the proliferation of parameters and the limited flexibility in reproducing covariances (Marzano and Papola, 2008), represents, instead, a more suitable formulation for route choice. The concept is, differently from the LNL, creating a nest for each pair of routes, thus not involving the links as nests. It implies the impossibility of working with an implicit path enumeration. However, the structural parameters can be operationalised through the observable quantities, as in the LNL case.

The PCL route choice probability is defined as:

$$p(k) = \frac{\exp\left[\dfrac{V_k}{(1-\sigma_{kk'})}\right]}{\exp\left[\dfrac{V_k}{(1-\sigma_{kk'})}\right] + \exp\left[\dfrac{V_{k'}}{(1-\sigma_{kk'})}\right]} \cdot \frac{\left\{\exp\left[\dfrac{V_k}{(1-\sigma_{kk'})}\right] + \exp\left[\dfrac{V_{k'}}{(1-\sigma_{kk'})}\right]\right\}^{\frac{1-\sigma_{kk'}}{\theta}}}{\sum_{r=1}^{n_k-1} \sum_{r'=r+1}^{n_k} \left\{\exp\left[\dfrac{V_r}{(1-\sigma_{rr'})}\right] + \exp\left[\dfrac{V_{r'}}{(1-\sigma_{rr'})}\right]\right\}^{\frac{1-\sigma_{rr'}}{\theta}}} \tag{2.104}$$

or, alternatively:

$$p(k) = \frac{\exp\left[\dfrac{V_k}{(1-\lambda_{kk'})}\right]}{\exp\left[\dfrac{V_k}{(1-\lambda_{kk'})}\right] + \exp\left[\dfrac{V_{k'}}{(1-\lambda_{kk'})}\right]} \cdot \frac{(1-\lambda_{kk'}) \cdot \left\{\exp\left[\dfrac{V_k}{(1-\sigma_{kk'})}\right] + \exp\left[\dfrac{V_{k'}}{(1-\sigma_{kk'})}\right]\right\}^{\frac{1-\lambda_{kk'}}{\theta}}}{\sum_{r=1}^{n_k-1} \sum_{r'=r+1}^{n_k} (1-\lambda_{rr'}) \cdot \left\{\exp\left[\dfrac{V_r}{(1-\sigma_{rr'})}\right] + \exp\left[\dfrac{V_{r'}}{(1-\sigma_{rr'})}\right]\right\}^{\frac{1-\lambda_{rr'}}{\theta}}} \tag{2.105}$$

Gliebe et al. (1999) assumed the *similarity parameter* of the second formulation as:

$$\sigma_{kk'} = \frac{C_{k-k'}}{C_k + C_{k'} - C_{k-k'}} \tag{2.106}$$

Prashker and Bekhor (1998) proposed a second formulation as:

$$\lambda_{kk'} = \frac{C_{k-k'}}{\sqrt{C_k \cdot C_{k'}}} \tag{2.107}$$

The S.U.E. – PCL formulation has been analysed in Prashker and Bekhor (1999).

2.3.3.3 Link Based - Network GEV

A formulation for adapting the Network GEV model for route choice has been proposed in Papola and Marzano (2013), under the name of Link Based-Joint Network GEV (LB-JNG) model. The basic idea is to obtain a particular Network GEV formulation by putting in sequence the choices at each choice stage of the so called *joint choice context*. A joint choice context is characterized by m choice dimensions, and for each one of these a set of alternatives is available. In a route choice context such sequence can be related to the sequence of links. The probability of choosing an alternative *(i,j)* in a specific choice stage *i* is a conditional probability expressed as a function of the GEV generating functions at that stage, and some structural parameters. The latter are computed as a function of the minimum path cost from an origin *o* towards the tail of the considered link, the cost of the link and the number of paths connecting o with *t(l)*. It has been noted that the latter can be implicitly computes and the authors provided an implicit enumeration procedure for the stochastic network loading on the efficient sub-network, in a Dial fashion.
The S.U.E. problem with the Network GEV model for route choice has been analysed in Hara and Akamatsu (2014).

2.3.3.4 Multinomial Weibit

It is known that the Weibull distribution, as long as the Gumbel distribution, is a particular case of the more general Generalized Extreme Value distribution. The latter, in general, depends on three real parameters (v,θ,γ). The Weibull distribution is obtained when the parameter $\gamma<0$[14]. Castillo et al. (2008) have led to a closed-form expression for the route choice probability by assuming the random residuals (of the perceived path costs) as independently Weibull distributed. The identical distribution assumption is then relaxed, leading to an heteroskedastic (but independent) model. Particularly, the choice probability expression of the resulting model, called Multinomial Weibit (MNW), is:

$$p(k) = \frac{\left(\theta_{\,k}\right)^{-\gamma}}{\sum_{k'\in \mathrm{K_{od}}} \left(\theta_{k'}\right)^{-\gamma}} = \frac{\left(t_k - v^0\right)^{-\gamma}}{\sum_{k'\in \mathrm{K_{od}}} \left(t_{k'} - v^0\right)^{-\gamma}} \tag{2.108}$$

Where the v^0 is assumed by the author with the meaning of minimum possible travel time and the generic $t_{k'}$ represents the total travel time for the path k'. Furthermore, the authors made note that assuming a logarithmic transformation in the MNL formula, the (2.108) could be obtained. In other words, the Weibit can be seen as a Multinomial Logit formula wherein the Gumbel distributed utility is expressed as the natural logarithm of the a Weibull distributed term (in fact, a logarithm of a Weibull is a Gumbel). This is the reason why often the Weibit is related to the multiplicative random utility framework (Fosgerau and Bierlaire, 2009). In fact, as

[14] The Gumbel case is obtained for $\gamma \rightarrow 0$, thus the remaining v and θ represent, respectively, the mean and the variance parameters. When $\gamma>0$ the GEV distribution gives the Frechet distribution.

seen in Section 2.1.2, a multiplicative random utility model can be expressed as equivalent additive utility model by means of a transformation of the utility in the logarithm of another random variable.

The model, proposed for route choice to handle the heteroskedastic effects on the choice probabilities, does not allow for correlation among utilities. Furthermore, the authors did not provide any applications for validating the model. Kitthamkerson and Chen (2013) extended the potential of such heteroskedastic specification by exploring the Weibit with a Path-Size attribute in the systematic utility (PS-MNW). The author also provided the S.U.E. mathematical programming formulations for the MNW and the PS-WNW, assuming the total path travel cost being the product of the link costs. Successively, Kitthamkerson and Chen (2014) analysed a mathematical programming problem for SUE-MNW without constrained optimization, while Kitthamkerson et al. (2015) analysed the SUE-MNW with elastic demand. Nakayama and Chikaraishi (2015) also investigated the possibility of using an unifying framework (*generalized logit route choice model*) with a GEV distribution, obtaining the latter by replacing the exponential in the Gumbel cdf with the *q*-exponential function (a type of generalized exponential function; Tsallis, 1994; Umarov et al., 2008), including the MNL and the MNW as special cases.

2.3.4 Error component models

In this section, the route choice models involving simulations are described. The Multinomial Probit model and successive applications of the Error component Logit model for route choice will be presented. The idea is always the same: providing a theoretically robust model, allowing for taking into account the overlapping effects on the choice probabilities, but allowing for the computation of the parameters in an easy way. The MNP and EC-MNL frameworks are very similar. In fact, the only difference lies in the indicator $I_j^{n,t}$ in the (2.15), wherein it is assumed the binary 0/1 value for MNP and the conditional MNL formula for the EC-MNL. However, the crucial step is the computation of the covariance matrix for the Multivariate Normal distribution of the random terms. In the following, the operationalization of these two models is presented.

2.3.4.1 Multinomial Probit and Monte-Carlo algorithm

The Multinomial Probit model for route choice has been proposed in Burrel (1968) and Daganzo and Sheffi (1977). The main assumption they made is the variance of the perceived link costs as proportional to the link costs themselves (sometimes referred as UPC, i.e. *Utilities proportional covariances*; Papola, 2004). In vector notation, the diagonal variance/covariance matrix of the link costs can be written as:

$$\mathbf{\Sigma}_l = \alpha \cdot diag(\mathbf{c}) \tag{2.109}$$

Being α the proportionality constant and diag($\mathbf{c}$) the diagonal matrix of the link costs. This position implies, under the hypothesis of additive link costs, that the perceived utilities are multivariate Normal distributed, with variance/covariance matrix $\mathbf{\Sigma}$ given by:

$$\mathbf{\Sigma} = \mathbf{A}^{\mathrm{T}} \cdot \mathbf{\Sigma}_l \cdot \mathbf{A} = \alpha \cdot \mathbf{A}^{\mathrm{T}} \cdot diag(\mathbf{c}) \cdot \mathbf{A} \tag{2.110}$$

The single term of the (2.110) is the numerator of each term of the matrix correlation, whose generic element is:

$$\begin{aligned}\rho_{kk'} &= \frac{\sigma_{kk'}}{\sqrt{\sigma_{kk}\sigma_{k'k'}}} = \frac{1}{\sqrt{C_k C_{k'}}} \cdot \sum_{l \in \mathrm{L}_k \cap \mathrm{L}_{k'}} c_l = \\ &= \frac{1}{\sqrt{C_k C_{k'}}} \cdot \sum_l a_{lk} \cdot a_{lk'} \cdot c_l \qquad \forall k, k' \neq k \in \mathrm{K}_{\mathrm{od}}\end{aligned} \tag{2.111}$$

being σ_{kk} and $\sigma_{k'k'}$ the path variances, with $\sigma_{kk'}$ the covariance among k and k', and with L_k and $\mathrm{L}_{k'}$ the set of the links belonging to the routes k and k'.
This assumption allows for a practical implicit enumeration computation of the probabilities, by means of drawings of the generic link cost from a mono-variate Normal distribution $N(c_l, \alpha \cdot c_l)$. The Monte-Carlo algorithm computes directly the link flows, without need of paths enumeration, with the following steps:

- Drawing of perceived link costs;
- Minimum path costs from each o;
- Deterministic network loading performed for each o;
- Averaging of the deterministic link flows;

The minimum path costs algorithm are generally single-step procedures, thus enabling the Monte-Carlo algorithm operating with a complexity that is proportional with n_C. However, the MNP has the adding problem of requiring many iterations of drawing for reaching stable results. Furthermore, it has been noted that the Daganzo and Sheffi assumption (2.109) depends on a unique variance parameter α. It represents undoubtedly an advantage in terms of estimation (in fact, only one parameter defines all the covariance matrices of the routes for each o-d pair), but it could lead to biased results. In fact, the (2.109) implies that the coefficient of variation decreases with the path cost. In other words, the dispersion of the link costs is assumed to decrease with the distance between an o-d pair, that could be unrealistic. Finally, a S.U.E. formulation (Powell and Sheffi, 1982; Sheffi, 1985, Maher and Hughes, 1997), would need repeating drawings from the link costs densities, that makes the searching for the fixed point solution of the system (2.79) (and, equivalently, the implicit enumeration version) a very computational burdensome process.
Yai et al. (1997) proposed a Multinomial Probit application to the city of Tokyo, using the concept of *structured covariance matrix*, assuming the total variance matrix $\mathbf{\Sigma}$ as a sum of two matrices: the first referred to the length of the routes and the second being independent from the lengths of the routes. However, in their study, the MNP was estimated on a maximum of four alternatives for each o-d pair.

2.3.4.2 Mixed Logit

The Mixed Logit model for route choice has been formalized by Bekhor et al. (2002). The structure was perfectly analogous to that of Probit model, because it assumed the variance of the link costs to be proportional to the link lengths. Particularly, the perceived utilities vector $\mathbf{U}_n$ of the individual n was defined as:

$$\begin{aligned}\mathbf{U}_n &= \mathbf{V}_n + \boldsymbol{\varepsilon}_n = \boldsymbol{\beta}^{\mathrm{T}} \cdot \mathbf{X}_n + \mathbf{A}_n{}^{\mathrm{T}} \cdot \mathbf{T} \cdot \boldsymbol{\zeta}_n + \mathbf{v}_n = \\ &= \boldsymbol{\beta}^{\mathrm{T}} \cdot \mathbf{X}_n + \alpha \cdot \mathbf{A}_n{}^{\mathrm{T}} \cdot diag(\mathbf{c}^{1/2}) \cdot \boldsymbol{\zeta}_n + \mathbf{v}_n\end{aligned} \tag{2.112}$$

being the first product the vector of the systematic utilities, the second product the Choleski factorization of the matrix $\mathbf{\Sigma}$ and the third vector the i.i.d. Gumbel disturbances. Particularly, in the second product, according to Frejinger and Bierlaire (2007), the incidence matrix A appears as individual matrix (the choice-set is individual-specific) and the product $\mathbf{T} \cdot \boldsymbol{\zeta}_n$ is the Choleski matrix on the $\boldsymbol{\Sigma}_l$. In fact, it is easy to recognize that is $\boldsymbol{\Sigma}_l = \mathbf{T}\cdot\mathbf{T}^{\mathrm{T}}$, and the covariance matrix of the paths, with this factor-analytic specification, is:

$$\boldsymbol{\Sigma} = \mathbf{A}^{\mathrm{T}} \cdot \mathbf{T} \cdot \mathbf{T}^{\mathrm{T}} \cdot \mathbf{A} \tag{2.113}$$

and it represents the same result already shown for the MNP.

Frejinger and Bierlaire (2007) proposed a methodology for capturing the correlation effects on the choice probabilities, within the Error Component Logit framework, through the so-called *subnetwork components*. Practically, the overlapping effect was not computed as a function of all the elemental links of the network, but as a function of some relevant network roads. The behavioural interesting representation founded on the basic idea that an individual does not actually perceive all the irrelevant links of the network, but only the main ones. Mathematically, this means simplifying the (2.113) as:

$$\mathbf{U}_n = \mathbf{V}_n + \boldsymbol{\varepsilon}_n = \boldsymbol{\beta}^{\mathrm{T}} \cdot \mathbf{X}_n + \mathbf{F}_n{}^{\mathrm{T}} \cdot \mathbf{T}^{\mathrm{Q}} \cdot \boldsymbol{\zeta}_n + \boldsymbol{\nu}_n \tag{2.114}$$

by means of a factor matrix loading $\mathbf{F}_n$ whose dimension is (n_k x n_Q), where n_Q is the number of the subnetwork components and its generic element is the square root of the overlapping length of the generic sub-network component q with the generic path k, while $\mathbf{T}^{\mathrm{Q}}$ is the generic (n_Q x n_Q) matrix of covariance parameters of the sub-network q to be estimated. The advantage of (2.114) is that it significantly reduces the quantities involved, passing from a size n_L to a size $n_Q << n_L$ for the matrix computation of the Multivariate Normal error terms. Furthermore, the authors tested a specification obtained by introducing a Path Size attribute, showing how the latter could provide better goodness of fit results.

The Mixed Logit model provides some advantages with reference to the MNP when using the explicit enumeration path approach, given the smoothness of its probability estimator. However, a Mixed Logit formulation with implicit path enumeration is not possible[15], so the MMNL is preferable to the simple MNP only with the explicit enumeration approach.

2.3.5 Going further

In the sections 2.3.2, 2.3.3 and 2.3.4 the main route choice models have been described. The taxonomy has proposed only the relevant route choice models under the classic approach described in the Section 2.3.1, i.e., for example, not considering the presence of cyclic paths.

Within the domain of the route choice models using acyclic paths, this state of the art has not faced the problem of how to generate a choice-set in the route choice application. It is easy to recognize that the latter is a crucial step in route choice modelling. In fact, as in all the discrete choice applications, the actual choice set considered by the decision maker is unknown to the analyst. But, in the route choice context, the problem explodes. The choice set are generated by the analyst by algorithm that are network-based. Generally, they are not exhaustive, but they consider only a subset of all feasible acyclic paths set. Mainly, two approaches exist in literature: the deterministic approach and the stochastic approach. The deterministic approach generally use the minimum shortest path tree algorithm many time, by changing the network costs each time. For example, Ben-Akiva et al. (1984) proposed the labelling approach, consisting of putting into the choice set all the minimum cost paths, each one computed with a different label, i.e. with reference to a different quantity (length, monetary cost and so on). The *link elimination* approach (Azevedo et al., 1993) removes, one for each step, a link of the shortest path. At each step a new searching for the minimum path on the modified network (i.e. without

[15] Some tests on the use of an algorithm computing the Mixed Logit flows averaging Dial's algorithm flows at each iteration, instead of the deterministic flows, are shown in Marzano and Papola (2004). However, the Dial's algorithm, differently from MNP with Monte-Carlo algorithm, works only on efficient routes, providing null flows on the non-efficient links.

the removed links) is performed, adding the new shortest path to the considered choice set. De la Barra (1993) proposed the *link penalty* approach, i.e. adding a big cost to strongly penalize the link of the shortest path, thus not changing the network. Van der Ziypp and Fiorenzo-Catalano (2005) proposed the constrained k-shortest path. Differently, Prato and Bekhor (2006) adapted the existent *branch-and-bound* approach for road traffic networks. The latter does not use the shortest path algorithm. In the domain of the stochastic approaches, surely the Monte Carlo algorithm can be mentioned (Powel and Sheffi, 1982), which generates routes by means of link costs draws. Bovy and Fiorenzo-Catalano (2007) proposed the so-called double stochastic approach, specifying the utilities of the path as combination of parameters and attributes that are both random draws. Furthermore, the approach for considering the route choice probability itself, conditional on the adopted choice set, can be double. It can be banally used the deterministic approach, directly computing the probability with anyone of the route choice models presented in the previous sub-sections. Instead, a probabilistic approach can be used, by computing the route choice probability as the combination of the $p(k/K_{od}^{n,t})$ of the route k, assuming the choice set, and the probability $p(K_{od}^{n,t})$ of observing the choice set $K_{od}^{n,t}$. It becomes quickly impractical given the size of the network and the feasible permutations of the sub-set $K_{od}^{n,t}$ (see Ben-Akiva and Swait, 1984, Swait and Ben-Akiva 1987*a*,*b*; and Ben-Akiva and Boccara, 1995). Actually, the IAP Logit presented in Section 2.3.2.4 is an example of the probabilistic approach treated as deterministic in perceiving routes in the own choice-set. Recent attempts to treat the exhaustive choice set, by using more effective sampling strategies within it, are shown in Frejinger et al. (2009) and Flotterod and Berlaire (2013).
A recent literature focused on the other issue, i.e. the inclusion of cyclic paths in the choice set, by applying the Markovian chains concept for the stochastic network loading problem. Practically, such methodologies view the network loading as sequence of choices in various states. Thus, in these cases, the path concept itself is improper, and it is replaced by the concept of chain in a Markovian fashion, i.e. a sequence of events without memory of the previous events. This framework is called *Markovian Traffic Assignment* (*MTA*) and it was proposed for the first time by Sasaki (1965). However, the first link with the random utility theory is due to Bell (1995), Akamatsu (1996 and 1997), who proposed a Markovian chain process where at each stage (node), a link choice was performed with a MNL formula. The main difference between the classic SNL and the MTA methodologies lies in the underlying choice-set. The MTA does not restrict the choice set only to the acyclic paths, or some sub-sets of them, but it extends the choice set to all the feasible cyclic paths. This means the choice-set has an indefinite size. The Akamatsu's procedure was generalized by Baillon and Cominetti (2008). More recently, Fosgerau et al. (2013) started a new research strand proposing a recursive Logit model with unrestricted choice set, by linking this model to the dynamic discrete choice models (Rust, 1987) and allowing for a disaggregate estimation of the model, based on GPS trajectories collection. They also tested the addition of the so called Link Size attribute to the specification of the utilities, allowing for taking into account the overlapping problem in a deterministic (similar to PS-Logit) fashion. Successively, Mai et al. (2015) extended it to a recursive Nested Logit, assuming an MNL formulation for each link choice with a link specific variance parameter, and Mai (2016) to the general Network GEV model, by means of the concept of *contraction mapping* (Rust, 1987). However, the MTA seems to exhibit various problems. First of all, considering an indefinite choice set of cyclic paths can be a double-edge-sword. In fact, it could consider also very unrealistic paths, with cycles that are behaviourally inexplicable. Second, the MTA could be impossible to solve, due to numerical problems (Oyama and Hato, 2017). Definitely, the MTA approaches have been certainly deepened and their consideration of the cyclic paths have to be adopted with the opportune constraints.

Another problem is the collection and use of data in route choice. With the technological advance, the diffusion of smartphones applications, the GPS detection and the increasing availability of data (*network free data*), the problem is moving towards how to store in memory and adapting this big quantity of data. The route concept itself, defined as sequence of network links, could be revised. In fact, a trajectory is not collected directly as a sequence of links, but as a sequence of points to be matched with the topological network model (*map-matching).* It is easy to recognize that the GPS detection needs a high precision, especially within dense urban networks, wherein an error of a few meters can mislead about the actual link of the network chosen by the user. A comprehensive discussion on the problem is reported in Bierlaire and Frejinger (2007) and in the PhD thesis of Frejinger (2008). In the latter work, also a comprehensive discussion on the state of the art about the problem of the *adaptive route choice* (i.e. the route choice in the presence of real-time information; see also Gao et al, 2008 and 2010), and some papers applying the fuzzy logic and the neural networks approach to route choice have been reported.

Different paradigms have been used in route choice modelling. The maximizing utility framework could be not sufficient in some route choice situations. For example, the Prospect theory considers the possibility that a user is more risk adverse than other users. Thus, maybe he would reduce the risk when choosing a path, notwithstanding its cost is perceived as minimum. From the analyst standpoint, it means minimizing the risk, i.e. the variance of the perceived cost of that route (for example, in an urban area, the minimum cost could rapidly become a worsen path, because of the increasing of congestion in the rush-hour.). The original prospect theory framework is due to Kanheman and Tversky (1979). Recent contributions are due to Katsikopoulos et al. (2000), Avinieri and Prashker (2004), de Palma et al. (2008), Gao et al.(2010) and de Luca and Di Pace (2015). Another paradigm is represented by the *Elimination by Aspects theory* (EBA), proposed by Tversky (1972*a*,*b*). The framework does not have a precise assumption on the error structure (apart from a recent contribution of Kolhi and Jedidi, 2015), but it describes the choice cognitive process as a sequence of successive eliminations of alternatives within the original choice-set. The eliminations at each step are carried out by considering a specific characteristic, namely the *aspect*. When the alternatives within the considered sub-set for that step does not own that specific aspect, it is deleted from the sub-set. Batley and Daly (2006) successively integrated the framework, but actually the applications to route choice and to discrete choice in general, although the recognized appealing properties of the EBA (McFadden, 1981), are very limited. Recently, the Random Regret minimization (RRM) framework has been analysed for route choice modelling by Prato (2014) and Mai et al. (2017). Finally, Kazagli et al. (2016) proposed another simplified representation of the route by using the concept of *Mental representation items* (*MRIs*). The framework draws upon the cognitive sciences and intuitively searches for a representation of the route as perceived by the user that is more consistent with the limits in perceiving the space around him. This framework has not been investigated yet but, surely, the concept of route itself seems to be debated again, searching for more effective and (possibly) simplified representations.

2.4 The Combination of random utility models (CoRUM) as a unified framework

Recently, Papola (2016) proposed a new class of additive random utility models, whose main assumption is that the underlying cumulative distribution function (cdf) of the random residuals is a convex combination of other underlying cumulative distribution functions (cdf's). Starting from the fact that a finite mixture of cdf's is itself a cdf, and that a sum of continuous functions is itself a continuous function, this class of models (named CORUM) is defined as the class of ARUMs whose underlying cdf is a finite mixture of absolutely continuous cdf's. The assumption can be formalized as follows:

$$F(\varepsilon_1^{n,t},...,\varepsilon_m^{n,t}) = \sum_{c\in C} w^c \cdot F^i(\varepsilon_1^{n,t},...,\varepsilon_m^{n,t}) \tag{2.115}$$

being F^c the c^{th} generic cdf component and w^c are exogenous weights (to be estimated) such as:

$$\sum_{c\in C} w^c = 1 \tag{2.116}$$

$$w^c \geq 0, \forall c \in C \tag{2.117}$$

Since the first order derivatives of F entries into the integral in (2.15), and given that w^c are constants, the assumption (2.115) leads to the following choice probability statement:

$$p^{n,t}(j) = \sum_{c\in C} w^c \cdot p^{n,t,c}(j) \tag{2.118}$$

where $p^{c,n,t}(j)$ represents the choice probability for the ARUM component c. Moreover, the linear in w^c expression holds for the variances and covariances:

$$Var[\varepsilon_j] = \sum_{c\in C} w^c \cdot Var^c[\varepsilon_j] \quad \forall j \in M \tag{2.119}$$

$$Cov[\varepsilon_j, \varepsilon_m] = \sum_{c\in C} w^c \cdot Cov^c[\varepsilon_j, \varepsilon_m] \quad \forall j, m \in M \tag{2.120}$$

where the superscript c for the variances and the covariances indicates the variances and covariance underlying the generic c^{th} ARUM component.

For the model's micro-elasticities of the $p^{n,t,c}(j)$ with reference to an attribute of another alternative utility specification $X_{k,h}$, (2.115) also implies:

$$E_{X_{k,h}}^{n,t}(j) = \sum_{c\in C} w^c \cdot \frac{p^{n,t,c}(j)}{p^{n,t}(j)} \cdot E_{X_{k,h}}^{n,t,c}(j) \tag{2.121}$$

being $E_{X_{k,h}}^{n,t,c}(j)$ the generic c^{th} ARUM component micro-elasticity.

Substituting Nested Logit cdf's into (2.115), the following cdf is obtained:

$$F(\varepsilon_1^{n,t},...,\varepsilon_m^{n,t}) = \sum_{c\in C} w^c \cdot \exp - \sum_{k^c} \left(\sum_{a'\in k^c} e^{-\varepsilon_{a'}^{n,t}/\delta_k^c} \right)^{\delta_k^c} \tag{2.122}$$

And a very useful CoRUM specification can be obtained. In fact, the choice probability expression can be expressed as convex combination of NL choice probabilities:

$$p^{n,t}(j) = \sum_{c\in C} w^c \cdot = \frac{e^{V_j^{n,t}/\theta_k} \cdot \left(\sum_{a\in k} e^{V_a^{n,t}/\theta_k} \right)^{\delta_k - 1}}{\sum_{k'\in K} \left(\sum_{a'\in k'} e^{V_{a'}^{n,t}/\theta_{k'}} \right)^{\delta_{k'}}} \tag{2.123}$$

Defining $\delta_{k(j,h)}^c$ as the independence parameter of the nest $k(j,h)$ that includes the pair of alternatives j and h, within the NL component c, the variances and covariances can be expressed as:

$$Var[\varepsilon_j] = \sum_{c\in C} w^c \cdot \left(\frac{\pi^2}{6} \right) = \frac{\pi^2}{6} \quad \forall j \in M \tag{2.124}$$

$$Cov[\varepsilon_j, \varepsilon_h] = \frac{\pi^2}{6} \cdot \sum_{c\in C} w^c \cdot \left[1 - \left(\delta_{k(j,h)}^c \right)^2 \right] \quad \forall j, h \in M \tag{2.125}$$

yielding to the following correlation expression:

$$Corr[\varepsilon_j, \varepsilon_h] = \sum_{c \in \mathrm{C}} w^c \cdot \left[1 - \left(\delta^c_{k(j,h)} \right)^2 \right] \quad \forall j, h \in \mathrm{M} \tag{2.126}$$

The resulting model, named CoNL (*Combination of Nested Logit*), allows for a very flexible correlation pattern, as shown in the estimation tests performed both on real and on synthetic datasets in Papola (2016).

The CoRUM potential have been investigated only with reference to the flexibility of the correlation pattern reproduced. The CoRUM framework, on the other hand, may be much more, and this thesis work intends to investigate all its potential with reference to other crucial choice modelling problems, introducing the possibility to accommodate other sources of heterogeneity, such as the taste heterogeneity and the heteroscedasticity section (Chapter 3).

The Chapter 4 investigates how the flexible substitution patterns of the CoNL could provide benefits in terms of forecasting capability, contrasting it with other flexible RUMs formulations. Moreover, the CoNL is the unique random utility model in literature to have this flexibility while maintaining a closed form expression for correlations among alternatives. The latter property, as already pointed out in Papola (2016), opens up interesting scenarios in contexts wherein a priori expectations on the values of the covariances can be made. Thus, the route choice dimension seems the natural candidate to exploit the benefits of such formulation. The Chapter 5 explains how the CoNL model can be operationalized on a real-world network, providing a way for building the specification, in terms of Nested Logit components and in terms of nests within them, and a way for computing all the structural parameters involved, namely the $\delta^c_{k'}$ and the w^c. The Chapter 6 goes further, providing an implicit paths enumeration procedure for the CoNL route choice model and an in-depth analysis of its properties, by means of several tests on both small size and a real size networks.

Chapter 3: CoRUM for taste heterogeneity

This chapter introduces and investigates the properties of mixtures of the Combination of Random Utility Models (CoRUM) model proposed by Papola (2016). Leveraging the well-established literature on the family of Mixed Logit models, the chapter illustrates that a Mixed CoRUM model, particularly if specified as a combination of Nested Logit models (CoNL) as kernel mixing distribution, is effectively capable to handle inter-intra respondent taste heterogeneity and flexible substitution patterns of correlation. Experimental results on real data show the Mixed CoRUM to exhibit an appreciable improvement in goodness-of-fit with respect to Mixed MNL and Mixed NL models, without proliferation of random parameters due to its error component specification of the error structure and with not increased computational burden.

3.1 Background and motivation: the Mixed GEV as a practical solution to joint EC/RC Mixed Logit model

Random utility theory represents, even now, the most common theoretical framework for modelling decision makers' choices. The research in this field has been pursuing two meaning paths: reproducing the effects of inter-correlation between perceived utilities of the alternatives and capturing random inter-intra respondent taste variation.

Multinomial Logit model (Luce, 1959; McFadden, 1974), although its appealing simplicity, does not allow taking into account any of the above-mentioned aspects. The difficulties are well summarized in Train (2009). The first obstacle is the well-known limitation due to I.I.A. property, implying proportional substitution across alternatives. The second obstacle is represented by the possibility to take into account only systematic, but not random, taste variation. The third obstacle lies in its impossibility to handle repeated choice situation wherein unobserved factors are correlated over time.

Scientific literature until early ninety focused primarily on the first problem. The GEV family models (McFadden, 1978; Ben-Akiva and Francois, 1983; Dagsvik, 1994) represent the first generalization of the simple Logit, overcoming its impossibility of reproducing correlation effects. Particularly, the first attempt to partially overcome the I.I.A. property was the Nested Logit formulation (Domencich and MacFadden, 1975; Williams, 1977; Mac Fadden, 1978; DalyandZachary, 1979; Borsch and Supan, 1990). The most notorious generalization of the latter, proposed under several names and in slightly different forms, is the Cross Nested Logit model (Small, 1987; Chu, 1989; Vovsha et al., 1997; Wen and Koppelman, 2001; Bierlaire 2001; Ben-Akiva and Bierlaire, 2003, Daly and Bierlaire 2003; Papola, 2004; Abbé et al. 2007;

Marzano and Papola 2008), with a particular specification being the Pair Combinatorial Logit model (Chu, 1989; Gliebe et al., 1999; Wen and Koppelman, 2000). Further generalizations of the CNL, namely the RNEV (Daly, 2001) and the Network G.E.V. (Daly and Bierlaire, 2006; Newman, 2008), have been proposed, but they received some interest in real-world applications only in route choice context (Papola and Marzano, 2013). Discrete mixture of models is another way to attempt reaching the same target of CNL, i.e. maximum flexibility in reproducing correlation effects with a closed-form probability statement. Different models can be mentioned in literature. The FinMix model (Swait, 2003), for instance, is based on the assumption of a generating GEV function that is a finite mixture of generating functions of known G.E.V. models. The resulting model provides a probability statement that is combination of probabilities of models component, with the values of the weights depending on the structural parameters of the component themselves. The CoRUM model (Papola, 2016), instead, assumes its cdf being a convex combination of R.U.M. cdf's. The weights are, in this case, parameters to be estimated. The advantage is that the resulting probabilities, correlation and elasticities expressions can be expressed as a convex combination of the models component probabilities, correlations and elasticities expressions. Particularly, a CoRUM specified as combination of Nested Logit – named CoNL– has the advantage of being a closed-form model both in terms of probabilities and in terms of covariances, but unlike the simple Nested Logit model, provides also a very flexible substitution pattern of correlations. Recent applications (Papola et al., 2018) show how this can represent a big advantage in context, like route choice, wherein prior expectations on the covariances values can be made through observable attributes (Daganzo and Sheffi, 1977).

Multinomial Probit model (Daganzo, 1979) was the first to be considered a benchmark, inasmuch it theoretically allows to reproduce effects of covariances among alternatives with maximum degree of flexibility under error component specification. The computational burdens of simulating the not closed form probabilities, together with the not-smoothed nature of the likelihood estimator, has discouraged its use in practical applications, in favour of the Mixed Logit model (Mc Fadden, 1989; Ben-Akiva and Bolduc, 1996). McFadden andTrain (2000) demonstrated the generality of the Mixed Logit framework, and the possibility to approximate any R.U.M., including Multinomial Probit, with the chosen degree of closeness. However, as author stated successively (Train, 2008), no guidance was proposed to find a mixing distribution that attains an arbitrarily close approximation. So, even now, the advantage of the error component Logit seems more theoretical than practical. Furthermore, the Normal Error Component Logit model suffers from several identification issues (Walker, 2002; Walker and Ben-Akiva, 2006). So, higher the number of alternatives, greater is the number of random parameters to be estimated and the burden to specify the model.

The second problem relates the random taste variation. It has been faced from earlier applications of Multinomial Probit in the random coefficient specification (Hausman and Wise, 1978; Daganzo, 1979). However, the Normal assumption restriction may be inappropriate for many cases and the above-mentioned problem of not-smoothed estimator for the likelihood holds. The random coefficient Logit (Boyd and Mellman, 1980; Cardell and Dunbar, 1980; Revelt and Train, 1998) has been preferred and, in the last twenty years, has surely represented the most popular instrument for capturing random taste heterogeneity, but some problems remain. The first one is just due to the parametric nature of this formulation. In fact, prior hypothesis must be done on distributional shape, without a priori knowing anything on the real distribution of tastes. This means the analyst should prove many distributions and then evaluate their validity, generally on the basis of the goodness of fit improvements, or on a range of values restriction that one distribution ensures with respect to another. A wide variety of distribution, besides Normal and logNormal, have been tested for random parameters (Train and Sonnier, 2005; Bhat, 2011; Bhat and Sidharthan, 2012; Keane and Wasi, 2013; Dekker, 2016), both in preference and in willingness to pay space (Ben-Akiva et al., 1993). Nonparametric approaches has been proposed, avoiding to make specific assumptions on parameters shape distribution, trying to approximate continuous distribution of parameters as discrete distribution, as in the case of latent class models (Swait, 1994; Gopinath, 1995; Kamakura et al., 1996; Bhat, 1997; Greene and Hensher, 2003; Bajari et al, 2007; Bujosa et al., 2010; Greene and Hensher, 2012). Unfortunately, the difficulties of estimating the discrete distribution increases significantly with the number of random parameters and mass points (latent classes). In practical applications, it is not easy to find estimates for more than three or four mass points for each parameter. A mixing framework to exploit the advantages of the parametric and nonparametric approaches is the Mixed Logit with semi-nonparametric approach (Fosgerau and Bierlaire, 2007; Train, 2008*a*; Fosgerau and Hess, 2009; Bastin et al., 2010; Bujosa et al., 2010; Greene and Hensher, 2013; Fosgerau and Mabit, 2013; Yuan et al., 2015; Train, 2016; Bansal et al., 2018). Recently, Vij and Krueger (2017) proposed a nonparametric Mixed Logit formulation that avoids a specific assumption on shape distribution, but reduces parameters to estimate.

However, while a lot of solutions on better estimating the true distribution of random parameters have been proposed, very little has been said about a second problem, that arises in contexts wherein unobservable factors correlation is not due to randomness in marginal utilities. In this case, a simple random coefficient Logit formulation may lead to biased estimation of the distribution of parameters (Hess and Polak, 2004), due to the impossibility to disentangle random taste variation effects from inter-correlation effects.

This problem has been faced primarily in Bhat and Guo (2004), with a mixed spatially correlated Logit, wherein residential location choice was studied with a mixed Pair Combinatorial Logit. The model was specified to treat as correlated only the utilities of contiguous residential spatial units. However, the model was estimated with three random taste parameters and only one structural parameter, i.e. the dissimilarity parameter of the P.C.L., and the goodness of fit improvement was not so evident with respect to a simple MNL. The model was successively reprised by Sener et al. (2011) and the spatial limitation was relaxed, allowing a more general correlation pattern. Hess et al. (2005*a*) investigated the possibility of using mixed Nested Logit and Mixed Cross Nested Logit for a stated survey on mode choice for Swiss Metro rail in Switzerland. While Mixed NL provided good gains in fitting with respect to Mixed Logit, the Mixed CNL, although it was specified with a very basic nesting formulation, showed some problems in reaching global optimum, finding a better solution than Mixed MNL but a worst solution than Mixed NL. Other applications of random coefficient NL can be found in literature (see Teye et al., 2014; Cheng and Yang 2015; Haghani et al., 2015 for instance). Summarizing, there is a scientific evidence that mixed GEV formulations represents a good solution for disentangling inter-correlation and random taste variation effects, but the advantages showed satisficing results only with Nested Logit as kernel model, while several estimation issues or limitations have been encountered when Cross Nested Logit or its particularizations (P.C.L.) have been used.

The third problem can be explicitly taken into account, only using the parametric or semi-nonparametric approach, because of correlated unobserved factors involve multi-dimensional integral computation over the distribution of them (Train, 2009).

This chapter proposes a parametric approach, therefore adaptable immediately for semi-nonparametric methodologies, for disentangling inter-correlation and random taste variation effects, i.e. mixing combination of R.U.M. distributions. Particularly, using linear combination of Nested Logit models, in the way described by Papola (2016), may represent an easier and a more effective way to enrich the target with respect to the Mixed CNL model, other than avoiding the computational and identification burdens of error component. An alternative specific variance Normal error component CoRUM is also investigated, for testing the capacity to catch the effects of heteroscedasticity. Finally a more general Mixed CoRUM model, allowing for both inter-correlation and random taste variation, has been tested, showing its superiority with respect to the equivalently specified Mixed Logit model. The methodology is tested on real data. The application on the real case is conducted on a mode-choice stated survey in a 6-alternatives choice context. The results show a good improvement, in terms of goodness of fit, with reference to the all models specified with random parameters (Logit, Nested Logit, Cross Nested Logit and error component Logit). Thus, the mixed CoRUM formulation allows for a not-negligible improvement of goodness of fit, with respect to Mixed MNL and NL, and helps to avoid a) proliferation of random parameters due to error component specification of the error structure b) confounding effects in estimating random taste variation c) convergence problems of CNL and error component formulations adding random parameters.

3.2 Mixed RUMs: notation and formulation

This section introduces the notation and the general formulation of Mixed RUMs, subsequently particularized to the case of the CoRUM in Section 3.3. Following the RUM theory, the perceived utility $U^{n,t}_j$ of alternative $j\in$M for individual n and choice situation t can be decomposed into a systematic utility $V^{n,t}_j$ and a zero-mean random residual $\varepsilon^{n,t}_j$, usually representing the unobservable component of the perceived utility. In turn, $V^{n,t}_j$ is generally given by a linear combination of k observable attributes $X^{n,t}_{jk}$ and individual-specific marginal utilities $\beta^{n,t}_{jk}$, yielding:

$$U_j^{n,t} = V_j^{n,t}(\beta_{j,k}^{n,t}, X_{j,k}^{n,t}) + \varepsilon_j^{n,t} \tag{3.1}$$

Usually (Ben-Akiva and Bolduc, 1996; Cascetta, 2009; Train, 2009), in literature the Equation (3.1) does not include apices n and t for the marginal utilities, and all the unobservable part of utility lies in $\varepsilon^{n,t}_j$. However, to better focus the two problems mentioned in Section 3.1, i.e. the correlations and the taste heterogeneity effects on the choice probabilities, a general expression of the perceived utility $U^{n,t}_j$ for a mixed RUM can be expressed as:

$$U_j^{n,t} = \sum_k \cdot \beta_{kj}^{\,n,t} X_{kj}^{\,n,t} + \omega_j^{n,t} + \gamma_j^{n,t} = \sum_k (\overline{\beta_{kj}} + \lambda_{kj}^{n,t}) \cdot X_{kj}^{\,n,t} + \sum_{i\in C^n} \eta_{jm}^{n,t} \cdot y_{jm}^{t} + \gamma_j^{n,t} \tag{3.2}$$

wherein each marginal utility $\beta_{kj}^{n,t}$ is decomposed into a fixed term $\overline{\beta}_{kj}$ and a random term $\lambda_{kj}^{n,t}$ with covariance matrix Σ_λ, $\omega_j^{n,t}$ is a zero-mean error term with covariance matrix Σ_ω, and $\gamma^{n,t}_j$ is a random noise with covariance matrix Σ_γ; all random terms may follow any distributions, with $\omega_j^{n,t}$ and $\gamma^{n,t}_j$ not identically distributed. Practically, $\omega_j^{n,t}$ can be expressed also as in the right-hand side of equation (3.2), being y^t_{jm} be a 0/1 binary variable indicating absence/presence of correlation between $U_j^{n,t}$ and $U_m^{n,t}$ and $\eta_{jm}^{n,t}$ a mono-variate random term with variance $\sigma_{\eta^{n,t}_{jm}}$.

Thus, letting Σ_η be the (diagonal) covariance matrix of order m collecting all $\sigma_{\eta^{n,t}_{jm}}$'s and **Y** be the square matrix of order m collecting all y^t_{jm}'s, it occurs $\mathbf{\Sigma_\omega = \Sigma_\eta \cdot Y}$. As a special case, if $\omega_j^{n,t}$ are collectively $\boldsymbol{MVN}\sim(\mathbf{0},\mathbf{\Sigma}_\omega)$, the second sum can be expressed in a factor-analytic fashion through the Choleski factorization of the covariance matrix Σ_ω of the ω_j^n 's (Ben-Akiva and Bolduc, 1996; Walker et al., 2007).

Notably, the contribution of the first term in the right-hand side of equation (3.2) to the overall model variance can be expressed straightforwardly as $\mathbf{\Sigma}_{\boldsymbol{\beta}} = \mathbf{X}^{n,t}\cdot\mathbf{\Sigma}_{\boldsymbol{\lambda}}\cdot \mathbf{X}^{n,t\,T}$ (Hausman and Wise, 1978; Daganzo, 1979), being $\mathbf{X}^{n,t}$ the vector of all attributes of the systematic utility for the individual n and the choice scenario t. Overall, recalling the linearity of the covariance operator, the individual variance-covariance matrix reproduced by the model (3.2) can be written as:

$$\mathbf{\Sigma}^{n,t}{}_{\mathbf{U}} = \mathbf{X}^{n,t}\cdot\mathbf{\Sigma}_{\boldsymbol{\beta}}\cdot\mathbf{X}^{n,tT} + \mathbf{\Sigma}_{\eta}\cdot\mathbf{Y} + \mathbf{\Sigma}_{\boldsymbol{\gamma}} \tag{3.3}$$

Notably, the first contribution to the overall variance is individual-specific and choice scenario-specific, being a function of explanatory individual attributes, whilst the others are constant at least across individuals. Considering the all the playing random noises constant among choice situations, the same applies to t.

The probability statement of the model (3.2) takes the form:

$$p^{MRUM,n,t}(j) = \int_{\boldsymbol{\beta}}\int_{\boldsymbol{\omega}} p^{RUM,n,t}(j/\boldsymbol{\beta},\boldsymbol{\omega})\cdot f(\boldsymbol{\beta},\boldsymbol{\omega},\Psi)d\boldsymbol{\beta}d\boldsymbol{\omega} \tag{3.4}$$

wherein $\mathbf{\Psi}$ represents the vector of parameters of the joint distribution f($\boldsymbol{\beta}$, $\boldsymbol{\omega}$, $\mathbf{\Psi}$). Usually, $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ are uncorrelated, thus f($\boldsymbol{\beta}$, $\boldsymbol{\omega}$, $\mathbf{\Psi}$)=$f_{\boldsymbol{\beta}}(\boldsymbol{\beta},\mathbf{\Phi})\cdot f_{\boldsymbol{\omega}}(\boldsymbol{\omega}, \mathbf{\Theta})$ being $\mathbf{\Phi}$ and $\mathbf{\Theta}$ the vectors of parameters for $f_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ and $f_{\boldsymbol{\omega}}(\boldsymbol{\omega})$ respectively.

Expressions (3.1) and (3.2) can be particularized to obtain all models recalled in Section 3.1. For instance, the Mixed Logit is a special instance of (3.2) with γ_j^n identically and independently EV-I (Gumbel) distributed, yielding Multinomial Logit probabilities as special case of p^{RUM} in (3.4). Also, following Train (2009), two noteworthy limiting formulations are the pure random coefficient and the pure error component, assuming all $\eta^{n,t}_{ji}$'s or all $\lambda^{n,t}_{ji}$'s respectively to be deterministic.

Another special case is the alternative-specific variance Normal error component model, obtained by assuming in (3.2) all $\eta^{n,t}_{ii}$'s as mono-variate independent normal variables $\boldsymbol{N}\sim(\mathbf{0}, \sigma_{\eta_{ii}})$ and null η^n_{ji} $\forall$ $i\neq j$, that is $f_{\boldsymbol{\omega}}(\boldsymbol{\omega}, \mathbf{\Theta})$ is the product of mono-variate normal cdf's. In this case, letting $\boldsymbol{\sigma}_{\eta_{ii}}$ be the vector collecting all $\sigma_{\eta_{ii}}$'s, it occurs $\mathbf{\Sigma}_{\boldsymbol{\eta}}=diag(\boldsymbol{\sigma}_{\eta_{ii}})$ and (3.3) becomes:

$$\mathbf{\Sigma}^{n,t}{}_{\mathbf{U}} = \mathbf{X}^{n,t}\cdot\mathbf{\Sigma}_{\boldsymbol{\beta}}\cdot\mathbf{X}^{n,tT} + diag(\boldsymbol{\sigma}_{\eta_{ii}}) + \mathbf{\Sigma}_{\boldsymbol{\gamma}} \tag{3.5}$$

The probability statement (3.4) does not have usually closed-form primitives, thus simulation is needed, for instance by means of the following estimator:

$$p_{SIM}^{MRUM,n,t}(j) = \frac{1}{R}\sum_{r=1}^{R} p_r^{RUM,n,t}(j/\boldsymbol{\beta}^{r,n,t},\boldsymbol{\omega}^{r,n,t}) \tag{3.6}$$

wherein $\boldsymbol{\beta}^{r,n,t}$ and $\boldsymbol{\omega}^{r,n,t}$ represent the r^{th} vectors of draws from the joint distribution f($\boldsymbol{\beta}$, $\boldsymbol{\omega}$), with $r\in 1\ldots R$, varying, in general, across individuals and choice tasks. Guidance on simulation and on how to draw from multivariate distributions for mixed models estimation are reported, amongst others, by (Train, 2009; Daly et al., 2012). Consistently, also model estimation should be based on simulated log-likelihood (SLL) estimators, whose formulation depends upon the nature of the available estimation dataset. In the case of cross-sectional data, it occurs:

$$SLL = \sum_{n=1}^{N}\sum_{t=1}^{T^n}\ln\left[p_r^{MRUM,n,t}(j)\right] = \sum_{n=1}^{N}\sum_{t=1}^{T^n}\ln\left[\frac{1}{R}\sum_{r=1}^{R}p_r^{RUM,n,t}(j/\boldsymbol{\beta}^{r,n,t},\boldsymbol{\omega}^{r,n,t})\right] \tag{3.7}$$

In the case of panel data (Revelt and Train, 1998), the SLL estimator refers the probabilities of observing the sequence of individual's choices $y^{MRUM,n}$ and it takes the form:

$$SLL = \sum_{n=1}^{N} \ln(y^{MRUM,n}) = \sum_{n=1}^{N} \ln\left[\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{\mathrm{T}^n} p_r^{RUM,n,t}\left(j/\boldsymbol{\beta}^{r,n},\boldsymbol{\omega}^{r,n}\right)\right] \tag{3.8}$$

being T^n the number of choice situations t faced by individual n.

3.3 The Combination of RUMs (CoRUM): a generalization for taste heterogeneity

Under the framework (3.2), the assumption (2.115) can be generalized as a function of the perceived utilities, to take into account all the unobservable components within them. The CoRUM perceived utilities cdf can be thus expressed as a linear combination of a set C of n_{MC} absolutely continuous cdf's of m perceived utilities with the same weights of (2.115) and the constraints (2.116) and (2.117). Assuming the vector notations $\mathbf{U}=(U^{n,t}{}_1,\ldots, U^{n,t}{}_m)$ and $\boldsymbol{\beta}=(\beta^{n,t}{}_{1,1}, \ldots,\beta^{n,t}{}_{m,q})$, where q is the generic number of coefficients within the utility of the alternative m, the (3.9) can be written as:

$$F(\mathbf{U}) = \sum_{c\in \mathrm{C}} w^c \cdot F^c(\mathbf{U}) = \sum_{c\in \mathrm{C}} w^c \cdot F^c(\boldsymbol{\varepsilon},\boldsymbol{\beta}) \tag{3.9}$$

The CoRUM main assumption (2.115) can be viewed, under this framework, as a particular case of (3.9), wherein $\beta_{kj}^{\ n,t} = \overline{\beta_{kj}} \quad \forall n,t$.

Since the interest of the chapter is disentangling the effects of the random taste heterogeneity from the effects of the inter-correlations, the assumption that $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are independent is made, and (3.9), by definition, becomes:

$$F(\mathbf{U}) = \sum_{c\in \mathrm{C}} w^c \cdot F^c(\boldsymbol{\varepsilon}) \cdot F^c(\boldsymbol{\beta}) \tag{3.10}$$

The assumptions (3.10) and (3.2) lead to the following general expression of the choice probabilities:

$$p^{n,t}(j) = \sum_{c\in C} w^c \int_{\boldsymbol{\beta}}\int_{\boldsymbol{\omega}} p^{c,n,t}(j/\boldsymbol{\beta},\boldsymbol{\omega}) \cdot f^c(\boldsymbol{\beta},\boldsymbol{\Phi}^c) \cdot f^c(\boldsymbol{\omega},\boldsymbol{\Theta}^c)\, d\boldsymbol{\beta}\, d\boldsymbol{\omega} \tag{3.11}$$

Proof. Indicating with $\mathbf{P}$ the set of parameters describing the density function of $\boldsymbol{\varepsilon}$, the assumption (3.11) on the cdf's implies the well-known property (see Erto, 2004 for example) that, for each cdf component c:

$$f^c(\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\Psi}^c)\, d\boldsymbol{\beta}\, d\boldsymbol{\varepsilon} = f^c(\boldsymbol{\beta},\boldsymbol{\Phi}^c)\, d\boldsymbol{\beta} \cdot f^c(\boldsymbol{\varepsilon},\mathbf{P}^c) \cdot d\boldsymbol{\varepsilon} \tag{3.12}$$

That, under the assumption (3.2), indicating with $\boldsymbol{\Gamma}$ the set of parameters describing the density function of $\boldsymbol{\gamma}$, becomes:

$$f^c(\boldsymbol{\beta},\boldsymbol{\varepsilon},\boldsymbol{\Psi}^c)\, d\boldsymbol{\beta}\, d\boldsymbol{\varepsilon} = f^c(\boldsymbol{\beta},\boldsymbol{\Phi}^c)\, d\boldsymbol{\beta} \cdot f^c(\boldsymbol{\omega},\boldsymbol{\Theta}^c)\, d\boldsymbol{\omega} \cdot f^c(\boldsymbol{\gamma},\boldsymbol{\Gamma}^c)\, d\boldsymbol{\gamma} \tag{3.13}$$

Using the *convenient error partitioning procedure* (Train, 1995; Train, 2009), it is easy to obtain a choice probability expression equivalent to (3.4) for each component c. In fact, a function of the parameters $h(\boldsymbol{\beta},\boldsymbol{\varepsilon},\mathbf{X})$, and an indicator $\mathrm{I}\left[h(\boldsymbol{\beta},\boldsymbol{\varepsilon},\mathbf{X}) = j\right]$ that take the value of 1 when $h(\boldsymbol{\beta},\boldsymbol{\varepsilon},\mathbf{X}) = j$, 0 otherwise, can be assumed. Thus, the following choice probability expression can be obtained:

$$
\begin{aligned}
p^{c,n,t}(j) &= \int_{\mathbf{U}} \mathrm{I}\left[h(\boldsymbol{\beta},\boldsymbol{\varepsilon},\mathbf{X}) = j\right]\cdot f^{c}(\mathbf{U},\mathrm{T}^{c})d\mathbf{U} = \\
&= \int_{\boldsymbol{\beta}}\int_{\boldsymbol{\varepsilon}} \mathrm{I}\left[h(\boldsymbol{\beta},\boldsymbol{\varepsilon},\mathbf{X}) = j\right]\cdot f^{c}(\boldsymbol{\beta},\boldsymbol{\Phi}^{c})\cdot f^{c}(\boldsymbol{\varepsilon},\boldsymbol{\Theta}^{c})d\boldsymbol{\beta}d\boldsymbol{\varepsilon} = \\
&= \int_{\boldsymbol{\beta}}\left[\int_{\boldsymbol{\varepsilon}} \mathrm{I}\left[h(\boldsymbol{\beta},\boldsymbol{\varepsilon},\mathbf{X}) = j\right]\cdot f(\boldsymbol{\varepsilon},\boldsymbol{\Theta}^{c})d\boldsymbol{\varepsilon}\right]\cdot f^{c}(\boldsymbol{\beta},\boldsymbol{\Phi}^{c})d\boldsymbol{\beta} = \\
&= \int_{\boldsymbol{\beta}}\left[\int_{\boldsymbol{\omega}}\left(\int_{\boldsymbol{\gamma}} \mathrm{I}\left[h_1(\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\gamma},\mathbf{X}) = j\right]\cdot f^{c}(\boldsymbol{\gamma},\boldsymbol{\Gamma}^{c})d\boldsymbol{\gamma}\right)\cdot f^{c}(\boldsymbol{\omega},\boldsymbol{\Theta}^{c})d\boldsymbol{\omega}\right]\cdot f^{c}(\boldsymbol{\beta},\boldsymbol{\Phi}^{c})d\boldsymbol{\beta}
\end{aligned}
\tag{3.14}
$$

The integral within round brackets can be termed $p^{c,n,t}[j/\boldsymbol{\beta},\boldsymbol{\omega}]$, that is a conditional choice probability of j, given the values $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$.
Thus, according to Papola (2016), given the linearity of the derivative operator and the (3.13), the assumption (3.9), leads to:

$$
\begin{aligned}
p^{n,t}(j) &= \int_{\boldsymbol{\beta}}\left[\int_{\boldsymbol{\omega}}\left(\int_{\boldsymbol{\gamma}} \mathrm{I}\left[h_1(\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\gamma},\mathbf{X}) = j\right]\cdot \sum_{c\in C} w^{c}\cdot f^{c}(\boldsymbol{\gamma},\boldsymbol{\Gamma}^{c})d\boldsymbol{\gamma}\right)\cdot f^{c}(\boldsymbol{\omega},\boldsymbol{\Theta}^{c})d\boldsymbol{\omega}\right]\cdot f^{c}(\boldsymbol{\beta},\boldsymbol{\Phi}^{c})d\boldsymbol{\beta} = \\
&= \int_{\boldsymbol{\beta}}\int_{\boldsymbol{\omega}}\left(\sum_{c\in C} w^{c}\cdot p^{c,n,t}[j/\boldsymbol{\beta},\boldsymbol{\omega}]\right)\cdot f^{c}(\boldsymbol{\beta},\boldsymbol{\Phi}^{c})\cdot f^{c}(\boldsymbol{\omega},\boldsymbol{\Theta}^{c})d\boldsymbol{\beta}d\boldsymbol{\omega}
\end{aligned}
\tag{3.15}
$$

Or, since the linearity of the integral, to the equivalent expression (3.11), and the proof is given.

Some restrictive cases of (3.15) are already present in literature. First, each model can be obtained by (3.15) when, as boundary case, only one component is assumed. In this case, the unique weight takes the value of 1 and the general Mixed RUM formulation (3.4) occurs, which is able to accommodate any random utility model (McFadden and Train, 2000). Second, other examples of models obtained by combination of density functions can be obtained by (3.15). For example, a latent class Mixed logit (Bujosa et al., 2010; Greene and Hensher, 2012) occurs when $\boldsymbol{\omega}$ vanishes and a simple conditional MNL formulation for $p^{c,n,t}[j/\boldsymbol{\beta}]$, i.e. a GEV type-I assumption on the $\boldsymbol{\gamma}$, is assumed. In that case, *c* assumes the meaning of class, with reference to the distribution of the $\boldsymbol{\beta}$. Differently from the latter, (3.15) allows considering different density functions also for the $\boldsymbol{\varepsilon}$.

The reader can note that a particular case of (3.15) occurs when $f^{c}(\boldsymbol{\beta}) = f(\boldsymbol{\beta})$ and $f^{c}(\boldsymbol{\beta}) = f(\boldsymbol{\beta})\ \ \forall c \in C$. In this case, (3.15) expresses the choice probability as a Mixed RUM with a kernel given by a combination of RUMs, as in (2.118). The (3.11), instead, expresses it as a combination of Mixed RUMs. Thus, the assumption (3.9) allows to build a model that can accommodate the taste variation and a flexible substitution pattern, by putting a closed form expression (2.118) within the general formulation (3.4) for Mixed RUMs, as in (3.15), or, equivalently, by combining Mixed RUMs as in (3.11). Assuming the first interpretation, in the following, the (3.15) will be termed *Mixed CoRUM*. However, it could equivalently be termed *CoMixed*RUM.
The equations (2.119) and (2.120), according to the considerations exposed in Section 3.2, take the more general form:

$$
Var^{n,t}[U_j^{n,t}] = \sum_{c\in \mathrm{C}} w^{c}\cdot Var^{c,n,t}[U_j^{n,t}] \tag{3.16}
$$

$$Cov^{n,t}[U_j^{n,t}, U_m^{n,t}] = \sum_{c \in C} w^c \cdot Cov^{c,n,t}[U_j^{n,t}, U_m^{n,t}] \tag{3.17}$$

However, given the generality of the (3.4) per se, it makes sense to analyse some particular specifications of the general (3.11). The advantage of (3.11) lies in its possibility to accommodate both flexible substitution patterns and taste heterogeneity, by assuming more restrictive assumptions on the distribution of $\boldsymbol{\omega}$, thus decreasing the number of random terms, and a closed form expression for $p^{c,n,t}[j/\boldsymbol{\beta}]$.

In this regard, this chapter proposes the analysis of some particular specifications for (3.11). When assuming (2.122) for the random terms $\boldsymbol{\varepsilon}$, under the framework (3.2) for the $\mathbf{U}$, (3.11) becomes:

$$p^{n,t}(j) = \sum_{c \in C} w^c \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\omega}} \frac{e^{(V^{n,t}{}_j(\boldsymbol{\beta}) + \omega_j^{n,t})/\delta_{k(j)}^c} \cdot \left[\sum_{a \in k(j)^c} e^{(V^{n,t}{}_a(\boldsymbol{\beta}) + \omega_a^{n,t})/\delta_{k(j)}^c} \right]^{\delta_{k(j)}^c - 1}}{\sum_{k^c} \left[\sum_{a' \in k^c} e^{(V^{n,t}{}_{a'}(\boldsymbol{\beta}) + \omega_{a'}^{n,t})/\delta_k^c} \right]^{\delta_k^c}} \cdot f(\boldsymbol{\beta}, \boldsymbol{\Phi}) \cdot f(\boldsymbol{\omega}, \boldsymbol{\Theta}) d\boldsymbol{\beta} d\boldsymbol{\omega} \tag{3.18}$$

That is a *combination of Mixed Nested Logit* models or, equivalently, a *Mixed CoNL* model.

When deleting $\boldsymbol{\omega}$ from (3.2), the formulation (3.18) becomes a pure random coefficient CoNL. The CoNL kernel allows to accommodate very flexible substitution patterns, as shown in Papola (2016), without need of introducing random terms due to error components specification. However, the (3.18) can be used also when $\boldsymbol{\omega}$ are present, but assuming, some restrictive distribution for f($\boldsymbol{\omega}$,$\boldsymbol{\Theta}$). Consistently, a Mixed CoNL with random coefficient and alternative specific variance error component, i.e. a model (3.18) obtained by assuming independent Normally distributed error terms ω_i, yields the following particularization of (3.5):

$$\boldsymbol{\Sigma}^{n,t}{}_{\mathbf{U}} = \mathbf{X}^{n,t} \cdot \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \cdot \mathbf{X}^{n,tT} + diag(\boldsymbol{\sigma}_{\eta_{ii}}) + \pi^2 / 6 \cdot \mathbf{P}_{\boldsymbol{\gamma}} \tag{3.19}$$

Interestingly, all the terms within (3.19) have a closed form expression, being $\mathbf{P}_\gamma$ the correlation matrix of the kernel CoNL model, whose elements are given by (2.126).

All considerations available on simulation, estimation and application of Mixed Logit and Mixed GEV models apply straightforwardly also to the Mixed CoNL, and any CoRUM obtained as combinations of Mixed GEV's. According to (Walker et al., 2007), an error component Logit in contexts with more than m=2 alternatives, has (m-1) alternative specific variances $\sigma_{\eta_{ii}}$ that can be identified. In a Mixed CoRUM with error component structure, analogously to the error component Logit, only one scale parameter is added to the perceived utility definition, i.e. the scale introduced by GEV variance parameter θ, that can be easily normalized to $6/\pi^2$. Thus, the same conclusions about the identification properties of error component Logit can be extended to all error component GEV formulations, and so, to all CoRUM formulations combining GEV cdf's, apart from the normalization rules for the specific kernel model.

Furthermore, the shape of the distribution of the $\boldsymbol{\beta}$ reproduced by (3.18) can be made more flexible, by adopting one of the sieve estimators proposed in literature (see Vij and Krueger, 2017 for a comprehensive review).

In this chapter, the performances of the Mixed CoRUM, particularly the Mixed CoNL model (3.18), are evaluated and contrasted with competing relevant models on a real dataset.

However, the framework (3.11) appears to be very general, and successive works can be dedicated to test other more flexible specifications of it. For example it could be interestingly investigated by contrasting it with other well-known semi-nonparametric formulations for the $\boldsymbol{\beta}$ (see Train, 2008 and Fosgerau and Mabit, 2013). In fact, since the linearity of the expression in the $f^c(\boldsymbol{\beta},\boldsymbol{\Phi}^c)$, it may be more easy to estimate. Another way to deepen the potential of (3.11) is considering nonparametric distributions. By assuming a discrete distribution for $f(\boldsymbol{\beta},\boldsymbol{\Phi})$ and a deleting $\boldsymbol{\omega}$ from (3.2), in fact, the integral in (3.11) is substituted by a sum over the hypothesized set of mass points. When the assumption $f^c(\boldsymbol{\beta}) = f(\boldsymbol{\beta}) \quad \forall c \in C$ is made, the model evolves in latent class with CoNL kernel or, equivalently, a combination of Latent class NL. The latter could be another interesting formulation to analyse, since the possibility to avoid making a restrictive parametric assumption on the shape of the distribution $f(\boldsymbol{\beta})$.
The performance of the proposed Mixed CoRUM and Mixed CoNL models is evaluated and contrasted with competing relevant models on real data in the following sub-section.

3.4 Application on real data

3.4.1 Stated preference survey

In March 2008, a stated preference (SP) survey was carried out on a sample of travellers on the multi-modal connection Milan-Naples (Italy). The reference universe is made up of all the users who travelled on the connection under study with High Speed (*HS*) trains (1st and 2nd class), the alternative high speed trains *Italo Nuovo Trasporto Viaggiatori* (*NTV*, 1st and 2nd class), by car on the motorway and airplane. The dataset considered in this application consists of 211 respondents with 8 choice scenarios for each one, for a total of 1688 observations.

3.4.2 Utilities specification

The utilities specification for the six alternatives take into account several socio-demographic and level of service attributes, summarized inTable 3.1.
Concerning travel time and access/exit time specific parameters for the three main transport modes have been considered (car, air, train). About monetary cost, instead, a generic parameter has been specified in each systematic utility, considering the presence of more than one person within the utility of the car, for computing a travel cost per person. The same applies to time between two trips, considering a generic parameter in each utility. Other services have been considered, as the presence of the restaurant, the Internet, the payment modality (in particular, the possibility of paying without cash), especially for long trips (>400 km). As socio-demographic variables the sex, the professional condition and the graduated condition are considered. The sex is considered only in the car for measuring the impact in terms of willingness to drive and to move with car. The professional condition is considered, instead, in the first class of the train modes. The latter is a proxy of the income and can influence the willingness to choose the first class. The graduated condition has been considered in the train choice. Finally, a dummy variable for the NTV knowledge has been considered in the systematic utility of the NTV options. Totally, 15 utility parameters are estimated.

Attributes	Car	Air	HS1	HS2	NTV1	NTV2
Travel time	✓	✓	✓	✓	✓	✓
Monetary cost	✓	✓	✓	✓	✓	✓
Number of person on board	✓					
Access/exit time		✓	✓	✓	✓	✓
Access/exit cost			✓	✓	✓	✓
Time between two trips <60 minutes		✓	✓	✓	✓	✓
Time between two trips >60 minutes		✓	✓	✓	✓	✓
Restaurant on board for long trips (>400 km)						✓
Payment modality					✓	✓
All services (Internet, restaurant,payment modality) on long trips (>400 km)						✓
Male	✓					
Professional condition			✓		✓	
Degree			✓	✓	✓	✓
NTV knowledge					✓	✓

Table 3.1: Model estimation on SP survey for six alternatives mode choice Naples-Milan - Attributes for the six alternatives choice scenario.

3.4.3 Models error structure

In the comparison, the following models are considered: Multinomial Logit, Nested Logit, Cross Nested Logit, CoRUM, Mixed Logit, Mixed Nested Logit, Mixed Cross Nested Logit and Mixed CoRUM.
The Nested Logit model is tested with four different specifications of the error structure. The first one hypothesizes a correlation between the two classes of High speed and NTV modes. The second one hypothesizes a correlation between the two modes in the first class and the same in the second class, while another nest is hypothesized for capturing the effects of being no-train alternatives. The third one hypothesizes a constant correlation between all the alternatives that are no car (i.e. grouping all the inland transport modes) and the fourth applies the same with a no air constant correlation (i.e. grouping all collective transport modes). The NL specification are shown in Figure 3.1. These Nested Logit specifications add at the most three nesting parameters to estimate (NL 2), but the correlations are hypothesized to be constant between all the alternatives within the same nest.

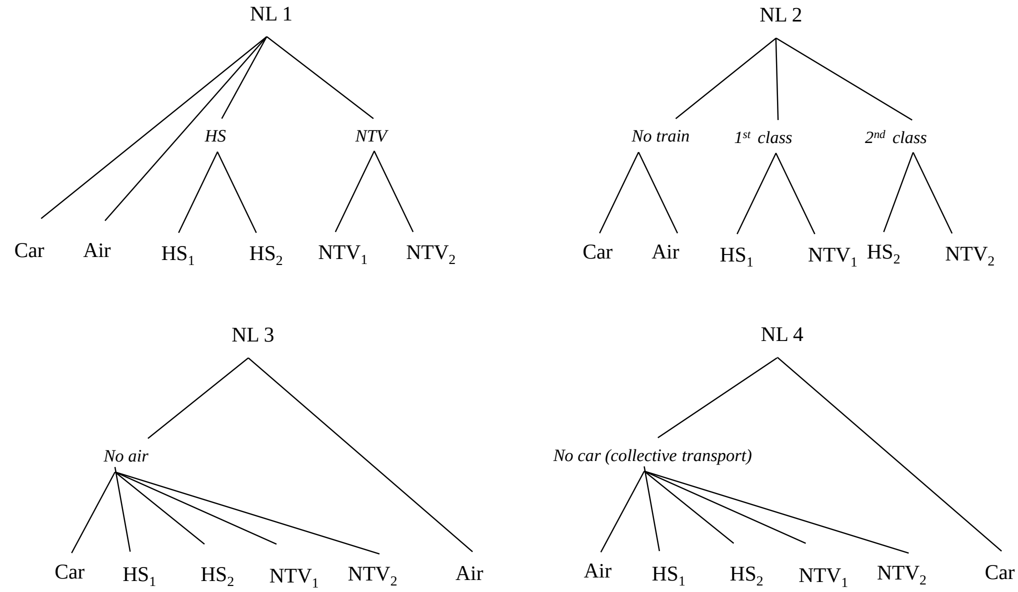


Figure 3.1: Model estimation on SP survey for six alternatives mode choice Naples-Milan - Nested Logit specifications for error structure.

The Cross Nested Logit relaxes the last assumption, giving the possibility to model a very flexible correlation matrix. Four interesting specifications are tested for the CNL, namely: the Pair Combinatorial specification, the full-nests specification, the union between the first two and, finally, a specification that brings together all the four NL specifications depicted in Figure 3.1. The Pair Combinatorial specification introduces 15 nests, with each alternative belonging to 5 nests and, thus, introducing 5 inclusion parameters (α_{jk}) to be estimated. Totally, the PCL needs 24 alpha's and 15 δ's to be estimated, but allows for a flexible correlation matrix representation. The second specification, i.e. the full nests one, is more economic, and allows for an high degree of flexibility in reproducing covariances, introducing only 6 alpha's and 2 δ's to be estimated. The third one represents a specification with a very labour intensive estimation, introducing a very high number of structural parameters to be estimated, because of each alternative belong to 7 different nests. Totally, we have 36 alpha's and 17 δ's to be estimated. Finally, the fourth specification is consistent with all the Nested Logit specification presented, allowing a not general correlation matrix, because of each alternative is not correlated with all the others. However, the relevant correlations are introduced with the estimation of 12 alpha's and 7 δ's to be estimated. The four specifications are depicted in Figure 3.2.

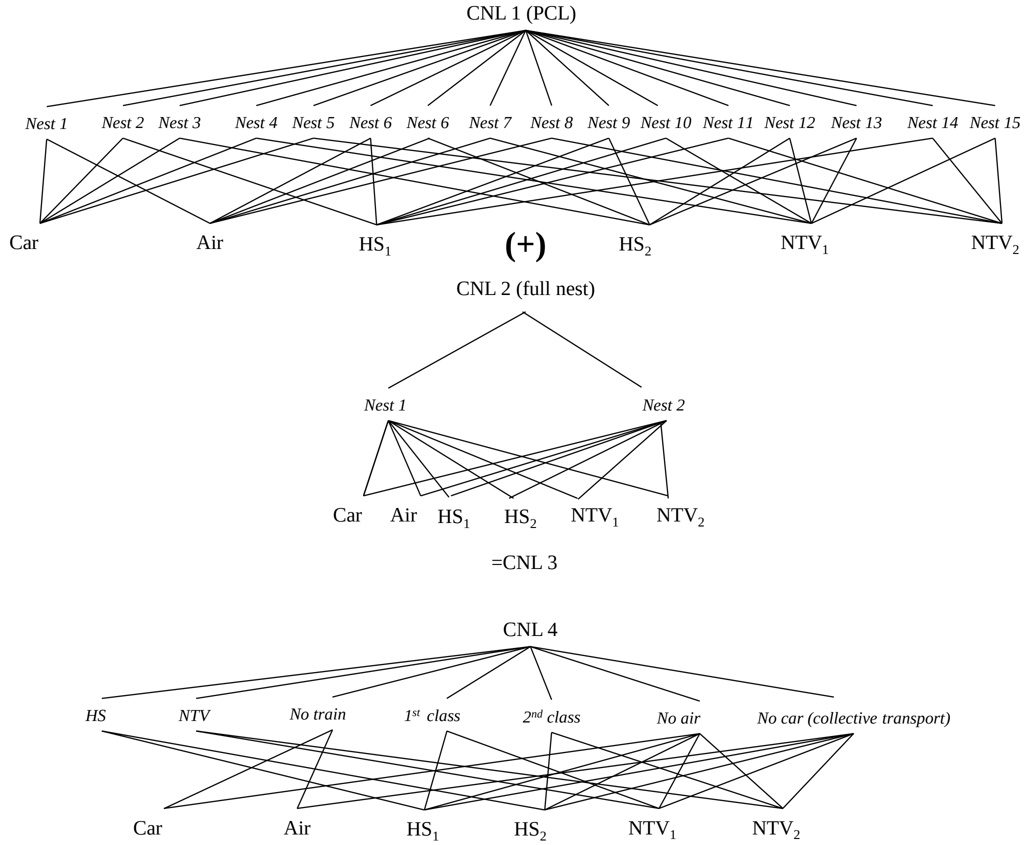


Figure 3.2: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Cross Nested Logit specifications for error structure.

The CoRUM model, as the CNL, allows for reproducing a flexible correlation scenario. It has been specified as CoNL in six different ways, as different combinations of the Nested Logit's depicted in Figure 3.1. The specifications and the parameters estimate are summarized as follows (in parenthesis the number of parameters):

- NL 1 + NL 2 (5 δ's and 1 CoNL weight);
- NL 2 + NL 4 (4 δ's and 1 CoNL weight);
- NL 1 + NL 4 (3 δ's and 1 CoNL weight);
- NL 1 + NL 2 + NL 3 (6 δ's and 2 CoNL weights);
- NL 1 + NL 3 + NL 4 (3 δ's and 2 CoNL weights);
- NL 1 + NL 2 + NL 3 + NL 4 (7 δ's and 3 CoNL weights);

The random coefficient models has been estimated with reference to all the error structure specifications described above. Particularly, each models has been estimated hypothesizing as random different combinations of parameters:

(1) Only generic cost parameter;
(2) Generic cost parameter and the specific travel time parameters (4 random coefficients);
(3) Generic cost, specific travel time and specific access/exit time parameters (6 random coefficients);

The shape distribution assumption represents a very sensitive operation, so seven different distributions have been tested, to try the one that improves at most the goodness of fit. They have been tested both in preferences space and in willingness to pay space. Such distributions are: Normal, logNormal, truncated Normal (only negative values), Uniform, Triangular, non-truncated Sb-Johnson and Rayleigh. The multi-dimensional integrals simulation have been performed with a variable number of draws per individual, starting from a very small number

(5/10) until 1000, to assess also the stability of the results. For the sake of simplicity only the Normal estimation results will be shown, representing the best ones in term of fitting measures. Further insights and details will be discussed in the next sub-section.

Finally, the models have been tested adding also an error component specification for random residuals. Particularly, the Mixed Logit and Mixed CoNL has been tested with Normal alternative specific variance error components. According to Walker et al. (2007), only m-1=5 parameters can be estimated in this case. This specification is perfectly equivalent to a pure random coefficient specification with alternative specific constant random with null mean. Thus, 11 random parameters need to be estimated for it.

Finally, the Mixed Logit is tested with an error component specification that is not alternative specific, but sharing Normal error components among the alternatives that are hypothesized as correlated. The joint random coefficient – error component specification theoretically represents the best that one can choose to approximate any RUM and for catching any correlation scenario. However, it is here analysed to better understand the gap between the mathematical potential of this theoretical framework and the estimation and simulation issues that such specification requires, because of the proliferation of random terms involved.

3.4.4 Estimation results

The estimation is performed with the aid of different software. All the results that will be discussed are the ones obtained with a flexible estimation code developed in Matlab R2018a. For validation purposes, all the existent closed form models and the Mixed Logit model have been estimated also using BIOGEME (Bierlaire, 2003), R and Matlab itself with other existent codes (Train, 2007). The maximum likelihood and the simulated maximum likelihood estimators are primarily used for comparison. The ρ^2, adjusted ρ^2, Akaike Information Criterion and Bayes Information Criterion are also used to assess the goodness of fit. Opportune likelihood ratio tests between a specification and another one taken as a benchmark are also shown, for evaluating the significance of introducing some parameters in the specification. Basically, likelihood ratio test are conducted to compare the maximum likelihood for a model and the model itself when some parameters is added. The same is performed between a closed form model with the corresponding mixed model, for evaluating the significance of introducing random parameters. In the results, the benchmark model will be reported in parenthesis. Finally, the estimation time is reported for each estimation, for consistence, using the computation times of our own code, being it the only one implementing all the models that are contrasted. All the mixed models estimation times refer a 100 Normal draws integral simulation for each random parameters and each individual.

The results are shown within the tables reported in Appendix 3/A, indicating for each parameter, the estimated value and the corresponding t-stat in parenthesis. The notation RC and EC mean, respectively, random coefficient and error component. Regarding CoNL model, the weights written in italic represents the weight obtained by the application of the constraint (3.10).

The first comment to the results refers the incredible difference that, in this estimation exercise, comes out from the mixed and the closed-form models. Although the utility specification takes into account both socio-demographic and level of service attributes, a lot seems to be unobservable to the analyst, so explicitly considering for randomness in tastes has an incredible adding value to the goodness of fit (ρ^2 exceeding 0.6 or even 0.7).

Another interesting aspect relates the balance of powers amongst the closed form models and amongst the mixed models. While the Cross Nested Logit performs better than the CoNL that, in turn, performs better than all the simple Nested Logit's, this trend is inverse when the taste variation is introduced into the models. In fact, the Mixed CoNL (see sixth specification) performs better than Mixed CNL with the same number of random parameters. Actually, all Mixed CNL estimations seem to fail in finding optimum, confirming what argued in Hess et al. (2005*a*). It means firstly that, notwithstanding the CNL can be easily estimated in a few seconds, even with a very demanding specification as the CNL 1 or CNL 3, the Mixed CNL suffers from several problems optimum research with each Matlab algorithm (inter-point, trust-region reflective, sqp, sqp-legacy, active-set). Only Mixed CNL with 2nd specification, i.e. the full nest one, converge to an optimum that is better than the equivalent Mixed Logit formulation one, but worse than the Mixed NL here tested (the 2nd NL specification). Second, a Mixed CoNL formulation does not suffer from these problems. Thus, the Mixed CoNL with 6 random parameters and exhibiting a log-likelihood of -827.56 outperforms the same specification for Mixed Logit (-930.71), Mixed NL(-867.9) and Mixed CNL (-918.37).

Furthermore, the error components specifications have been tested for Mixed CoNL and Mixed Logit. The Mixed CoNL with only alternative specific variance (no random parameters for marginal utilities) performs better than Mixed Logit with joint random coefficients and alternative specific variance specification (-793.78 vs -798.62). The CoNL specification number 4 has also been tested with both the 6 random coefficient and the 5 alternative specific variance error components, providing the best log-likelihood value (-779.52) of all estimated models. This result is to be compared with the error component Logit shown in Table 3.4 (fifth column) and with the joint error component random coefficient Logit shown in Table 3.4 (sixth column). The first Mixed Logit attempts to catch all the correlation effects with a Normal error component specification, going worse than the Mixed CoNL with only specific variance error component (-793.78 vs -795.49). The second specification, that is more general, even fails in finding optima, because of the excessive number of random parameters (23), and its log-likelihood is -932.36, that is worse than the Mixed Logit with only 6 random parameters.

The adjusted values of the ρ^2, AIC, BIC and the likelihood ratio tests performed between the Mixed CoNL and the Mixed Logit confirm the convenience of using the Mixed CoNL with reference to the Mixed Logit.

Thus, it can be concluded that although the Mixed Logit with a joint error component random coefficient structure can theoretically approximate any RUM, with a Mixed CoNL it is easier to find optima and it can outperform all relevant random utility models with random parameters.

3.5 Conclusions and future research

This chapter investigates the potential of the CoRUM model (Papola, 2016) for taste heterogeneity, particularly as mixtures of combination of RUMs. Although the Mixed Logit theoretically allows to approximate any RUM (McFadden and Train, 2000), it is often used only with random coefficient specification, potentially generating confounding effects that may bias the estimated parameters (Hess and Polak, 2004). Furthermore, the error component Logit notably suffers from several identification issues, as analysed in Walker et al. (2007). Thus, disentangling correlation and taste variation effects with a less expensive model is an open topic. The combination of RUMs, particularly the combination of Nested Logit's, has been analysed as a solution to this issues. The results show that 1) random coefficient CoNL systematically outperforms random component Logit, Nested Logit and Cross Nested Logit, having the last one several problems in optimum research 2) the Mixed CoNL avoids random terms proliferation due to error component specification, allowing for catching the covariance effects with its own structure 3) Mixed CoNL outperforms Mixed Cross Nested Logit also when the equivalent Cross Nested Logit outperforms the CoNL 4) the CoNL with random coefficient and specific alternative variance error component outperforms the Logit with random coefficients and a flexible error component structure, also because the latter has several problems in finding optimum.

One future step is surely represented by testing more flexible combination of mixed RUMs. Furthermore, the applications of a nonparametric or a semi-nonparametric approach for taste heterogeneity seems to be natural and not so expensive with reference to the obtained results, allowing making less restrictive assumptions on shape distribution of the random terms. Finally, it would be interesting to evaluate the gains of the formulation proposed in terms of posterior analysis with reference to the Mixed Logit. The latter is an interesting topic, given the increasing availability of individual data and the possibility to use the past choices in a Bayesian fashion, to make better individual forecasts.

Appendix 3/A: Estimation results on real data

	MNL 1		NL1		NL2		NL3		NL4	
Parameters	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*
β_{cm}	-0,032	*-11,5*	-0,025	*-8,8*	-0,026	*-9,6*	-0,018	*-8,4*	-0,032	*-7,5*
$\beta_{t,Car}$	-0,029	*-17,2*	-0,028	*-18,0*	-0,028	*-15,2*	-0,018	*-11,8*	-0,029	*-13,9*
$\beta_{t,Air}$	-0,019	*-4,1*	-0,020	*-4,6*	-0,024	*-5,1*	-0,020	*-5,6*	-0,019	*-4,0*
$\beta_{t,Rail}$	-0,015	*-8,9*	-0,015	*-9,6*	-0,016	*-9,8*	-0,012	*-10,0*	-0,015	*-7,2*
β_{tAEAir}	-0,013	*-4,5*	-0,015	*-5,2*	-0,013	*-4,5*	-0,012	*-5,1*	-0,013	*-4,4*
$\beta_{tAE,Rail}$	-0,028	*-10,3*	-0,028	*-10,6*	-0,029	*-10,6*	-0,019	*-9,6*	-0,028	*-8,1*
$\beta_{half\text{-}time<60}$	-0,036	*-11,6*	-0,033	*-10,4*	-0,024	*-5,4*	-0,013	*-5,5*	-0,036	*-10,0*
$\beta_{half\text{-}time>60}$	-0,002	*-1,7*	-0,001	*-1,7*	-0,001	*-1,9*	0,000	*-0,4*	-0,002	*-1,7*
β_{degree}	0,231	*1,9*	0,206	*1,7*	0,126	*1,0*	0,168	*2,6*	0,231	*1,8*
$\beta_{no\text{-}cash}$	-0,174	*-1,6*	0,025	*0,2*	0,088	*1,4*	-0,049	*-1,0*	-0,174	*-1,5*
$\beta_{NTV\text{-}known}$	-0,106	*-0,8*	0,059	*0,5*	-0,263	*-3,9*	-0,125	*-2,2*	-0,106	*-0,8*
$\beta_{restaurant}$	-1,325	*-3,0*	-0,633	*-2,8*	-1,034	*-2,3*	-0,578	*-2,9*	-1,325	*-2,8*
β_{male}	0,509	*3,1*	0,586	*3,6*	0,769	*4,4*	0,203	*2,7*	0,509	*3,1*
$\beta_{high\text{-}professional\text{-}condition}$	0,421	*4,0*	0,378	*4,9*	0,489	*3,6*	0,254	*4,8*	0,421	*3,8*
$\beta_{all\text{-}services}$	1,816	*3,8*	1,061	*3,9*	1,520	*3,1*	0,753	*3,3*	1,816	*3,5*
δ_{class1}	-	-	-	-	0,30	*-12,9*	-	-	-	-
δ_{class2}	-	-	-	-	0,69	*-1,5*	-	-	-	-
δ_{HS}	-	-	-	-	-	-	-	-	-	-
δ_{NTV}	-	-	0,393	*-10,5*	-	-	-	-	-	-
$\delta_{no\text{-}Air}$	-	-	-	-	-	-	0,426	*-11,6*	-	-
$\delta_{pub.transport(No\text{-}Car)}$	-	-	-	-	-	-	-	-	1	*0,0*
VTTs Car [euro/h]	54,13		67,06		64,18		58,84		54,13	
VTTs Air [euro/h]	35,30		48,37		56,27		64,61		35,30	

VTTs Rail [euro/h]	28,56	36,91	36,40	41,07	28,56
VTTs AE Air[euro/h]	24,97	35,89	30,65	40,48	24,97
VTTs AE Rail [euro/h]	52,59	67,24	66,30	63,64	52,59
maxLL	**-1988,97**	**-1964,60**	**-1964,58**	**-1971,59**	**-1988,97**
LL(0)	-2827,65	-2827,65	-2827,65	-2827,65	-2827,65
n_{par}	15	17	18	16	16
ρ2	0,2966	0,3052	0,3052	0,3027	0,2966
ρ2 adj.	0,2913	0,2992	0,2989	0,2971	0,2909
$t_{calibrazione}$*[sec]*	1,9	1,7	1,7	1,3	1,2
$t_{calibrazione}$*[min]*	0,03	0,03	0,03	0,02	0,02
LR test (MNL 1)	-3978	49	49	35	0
AIC	4008	3963	3965	3975	4010
BIC	4767	32625	34313	30951	30986

Table 3.2: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Multinomial Logit and Nested Logit estimation results.

Parameters	**CoNL 1=NL1+NL2**		**CoNL 2=NL1+NL3**		**CoNL 3=NL2+NL3**		**CoNL 4=NL1+NL2+NL3**		**CoNL 5=NL1+NL3+NL4**		**CoNL 6=NL1+NL3+NL4**	
	val.	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***
β_{cm}	-0,024	-5,9	-0,010	-5,4	-0,012	-6,5	-0,012	-6,5	-0,004	-4,7	-0,004	-3,4
$\beta_{t,Car}$	-0,028	-18,1	-0,018	-12,3	-0,018	-12,9	-0,019	-12,7	-0,013	-12,5	-0,013	-48,6
$\beta_{t,Air}$	-0,022	-4,4	-0,022	-6,3	-0,023	-6,7	-0,023	-6,7	-0,017	-7,2	-0,017	-5,0
$\beta_{t,Rail}$	-0,016	-9,9	-0,014	-11,3	-0,014	-11,3	-0,014	-11,2	-0,009	-12,1	-0,009	-17,1
β_{tAEAir}	-0,015	-5,7	-0,015	-6,1	-0,014	-5,8	-0,014	-5,8	-0,013	-7,3	-0,013	-6,6
$\beta_{tAE,Rail}$	-0,029	-11,4	-0,018	-9,0	-0,019	-9,5	-0,019	-9,3	-0,014	-7,9	-0,014	-4,4
$\beta_{half\text{-}time<60}$	-0,029	-2,6	-0,009	-5,1	-0,006	-6,3	-0,007	-6,2	-0,004	-5,5	-0,004	-1,5
$\beta_{half\text{-}time>60}$	-0,002	-1,9	0,000	-0,7	0,000	-0,3	0,000	-0,2	0,000	0,1	0,000	0,0
β_{degree}	0,159	1,3	0,073	1,2	0,093	1,5	0,090	1,4	0,095	1,2	0,094	0,4
$\beta_{no\text{-}cash}$	0,024	0,3	0,024	0,7	0,043	2,3	0,047	2,4	0,013	1,1	0,015	2,0
$\beta_{NTV\text{-}known}$	0,011	-0,4	-0,111	-2,8	-0,120	-5,4	-0,121	-5,2	-0,067	-4,2	-0,071	-1,6
$\beta_{restaurant}$	-0,722	-4,8	-0,252	-2,7	-0,278	-2,9	-0,263	-3,6	-0,134	-2,9	-0,150	-12,8
β_{male}	0,685	5,5	0,238	3,7	0,322	4,4	0,326	4,4	0,171	3,7	0,173	1,7
$\beta_{high\text{-}professional\text{-}condition}$	0,482	1,9	0,143	3,3	0,152	3,8	0,149	4,1	0,046	2,9	0,049	1,5
$\beta_{all\text{-}services}$	1,193	13,3	0,415	3,5	0,435	4,0	0,423	4,7	0,200	3,7	0,222	3,7
w1	0,557	6,6	0,246	5,4	0,320	7,6	0,046	1,0	0,000	0,0	0,000	0,0
w2	*0,443*	-	*0,754*	-	-	-	0,287	5,2	0,704	15,3	0,035	0,1
w3	-	-	-	-	-	-	*0,667*	-	*0,296*	-	0,693	2,5
w4	-	-	-	-	-	-	-	-	-	-	*0,271*	-
δ_{class1}	0,173	-11,1	-	-	0,049	-46,7	0,046	-42,3	-	-	0,025	-2,7
δ_{class2}	0,999	0,0	-	-	0,056	-33,6	0,045	-39,1	-	-	0,267	-1,0
δ_{HS}	-	-	-	-	-	-	-	-	-	-		
δ_{NTV}	0,355	-16,3	0,077	-15,3	-	-	0,025	-21,8	0,421	-1,6	0,424	-0,6
$\delta_{no\text{-}Air}$	-	-	0,219	-21,9	*0,217*	-25,0	0,218	-24,7	0,073	-47,3	0,076	-11,1
$\delta_{pub.transport(No\text{-}Car)}$	-	-	-	-	-	-	-	-	0,135	-20,5	0,143	-10,1
VTTs Car [euro/h]	68,25		103,62		93,93		94,64		214,91		206,79	

	CoNL 1=NL1+NL2		CoNL 2=NL1+NL3		CoNL 3=NL2+NL3		CoNL 4=NL1+NL2+NL3		CoNL 5=NL1+NL3+NL4		CoNL 6=NL1+NL3+NL4	
Parameters	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***
VTTs Air [euro/h]	53,24		125,98		117,90		118,90		273,52		259,46	
VTTs Rail [euro/h]	38,54		80,33		69,38		70,21		150,12		144,26	
VTTs AE Air[euro/h]	36,09		86,15		71,52		71,55		209,34		201,51	
VTTs AE Rail [euro/h]	70,20		100,76		96,66		97,22		229,70		222,20	
maxLL	**-1959,02**		**-1958,41**		**-1941,51**		**-1941,1**		**-1937,48**		**-1937,24**	
LL(0)	-2827,65		-2827,65		-2827,65		-2827,65		-2827,65		-2827,65	
n_{par}	21		20		19		22		21		21	
ρ2	0,3072		0,3074		0,3134		0,3135		0,3148		0,3149	
$\rho 2_{adj.}$	0,2998		0,3003		0,3067		0,3057		0,3074		0,3075	
$t_{calibrazione}$*[sec]*	4,2		5,2		3,2		8		14,3		14,3	
$t_{calibrazione}$*[min]*	0,07		0,09		0,05		0,13		0,24		0,24	
LR test (MNL 1)	60		61		95		96		103		103	
AIC	3.960		3.957		3.921		3.926		3.917		3.916	
BIC	39.366		37.677		35.955		41.018		39.323		39.322	

Table 3.3: Model estimation on SP survey for six alternatives mode choice Naples-Milan – CoNL estimation results.

Parameters	CNL 1 (PCL) val.	CNL 1 (PCL) *t-test*	CNL2 (full-nest) val.	CNL2 (full-nest) *t-test*	CNL 3=CNL1 + CNL2 val.	CNL 3=CNL1 + CNL2 *t-test*	CNL 4 (=CoNL 4 NL) val.	CNL 4 (=CoNL 4 NL) *t-test*
β_{cm}	-0,020	*-10,8*	-0,018	*-7,3*	-0,012	*-8,0*	-0,023	*-26,8*
$\beta_{t,Car}$	-0,018	*-29,1*	-0,019	*-11,1*	-0,013	*-38,6*	-0,021	*-14,8*
$\beta_{t,Air}$	-0,014	*-3,4*	-0,016	*-4,0*	-0,006	*-6,9*	-0,022	*-6,8*
$\beta_{t,Rail}$	-0,012	*-11,3*	-0,012	*-27,1*	-0,009	*-18,1*	-0,014	*-12,3*
β_{tAEAir}	-0,011	*-5,3*	-0,010	*-10,5*	-0,011	*-6,0*	-0,011	*-6,6*
$\beta_{tAE,Rail}$	-0,021	*-23,5*	-0,021	*-9,9*	-0,014	*-13,8*	-0,023	*-12,0*
$\beta_{half\text{-}time<60}$	-0,022	*-6,6*	-0,015	*-4,6*	-0,014	*-19,6*	-0,010	*-73,2*
$\beta_{half\text{-}time>60}$	-0,001	*-2,0*	-0,001	*-2,0*	0,000	*-2,1*	-0,001	*-2,0*
β_{degree}	0,276	*6,6*	0,046	*0,6*	0,178	*6,9*	-0,048	*-0,8*
$\beta_{no\text{-}cash}$	-0,082	*-1,3*	0,186	*3,9*	0,137	*2,9*	0,172	*3,9*
$\beta_{NTV\text{-}known}$	0,075	*1,2*	-0,149	*-2,4*	-0,168	*-6,9*	-0,073	*-2,4*
$\beta_{restaurant}$	-0,483	*-4,9*	-0,977	*-2,9*	-0,502	*-3,1*	-0,448	*-3,7*
β_{male}	0,329	*3,3*	0,650	*6,5*	0,469	*9,4*	0,594	*4,8*
$\beta_{high\text{-}professional\text{-}condition}$	0,356	*5,2*	0,347	*3,7*	0,239	*8,0*	0,518	*4,5*
$\beta_{all\text{-}services}$	1,079	*6,6*	1,468	*3,8*	1,122	*6,2*	0,843	*10,7*
$\alpha_{Car\text{-}1}$	0,000	*0,0*	0,697	*4,7*	0,025	*0,8*	1,000	*1,7*
$\alpha_{Car\text{-}2}$	0,214	*4,8*	0,303	*6,6*	0,264	*7,5*	-	-
$\alpha_{Car\text{-}3}$	0,208	*3,9*	-	-	0,232	*4,4*	-	-
$\alpha_{Car\text{-}4}$	0,325	*2,3*	-	-	0,141	*3,2*	-	-
$\alpha_{Car\text{-}5}$	0,253	*2,0*	-	-	0,047	*1,5*	-	-
$\alpha_{Car\text{-}6}$			-	-	0,095	*9,5*	-	-
$\alpha_{Car\text{-}7}$			-	-	0,197	*4,5*	-	-
$\alpha_{Air\text{-}1}$	0,070	*0,0*	0,901	*2,1*	0,005	*0,5*	1,000	*2,0*
$\alpha_{Air\text{-}2}$	0,070	*0,0*	0,099	*7,5*	0,028	*0,2*	-	-
$\alpha_{Air\text{-}3}$	0,070	*0,0*	-	-	0,030	*0,2*	-	-

Parameters	CNL 1 (PCL)		CNL2 (full-nest)		CNL 3=CNL1 + CNL2		CNL 4 (=CoNL 4 NL)	
	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*
α_{Air-4}	0,138	*0,4*	-	-	0,028	*4,1*	-	-
α_{Air-5}	0,652	*2,0*	-	-	0,439	*3,4*	-	-
α_{Air-6}			-	-	0,041	*8,6*	-	-
α_{Air-7}			-	-	0,429	*14,7*	-	-
α_{AV1-1}	0,645	*15,4*	0,000	*0,0*	0,207	*8,5*	0,853	*5,0*
α_{AV1-2}	0,000	*0,0*	1,000	*14,5*	0,001	*0,2*	0,000	*0,0*
α_{AV1-3}	0,000	*0,0*	-	-	0,001	*0,2*	0,147	*1,5*
α_{AV1-4}	0,094	*1,0*	-	-	0,001	*1,4*	-	-
α_{AV1-5}	0,261	*3,0*	-	-	0,002	*0,1*	-	-
α_{AV1-6}			-	-	0,307	*4,2*	-	-
α_{AV1-7}			-	-	0,481	*3,9*	-	-
α_{AV2-1}	1,000	*18,8*	1,000	*3,6*	0,770	*4,1*	0,001	*0,0*
α_{AV2-2}	0,000	*0,0*	0,000	*0,0*	0,002	*0,2*	0,208	*2,2*
α_{AV2-3}	0,000	*0,0*	-	-	0,002	*1,1*	0,791	*2,8*
α_{AV2-4}	0,000	*0,0*	-	-	0,004	*0,1*	-	-
α_{AV2-5}	0,000	*0,0*	-	-	0,138	*3,4*	-	-
α_{AV2-6}			-	-	0,043	*0,0*	-	-
α_{AV2-7}			-	-	0,043	*0,0*	-	-
α_{NTV1-1}	0,429	*4,6*	0,126	*0,7*	0,170	*3,3*	0,549	*2,9*
α_{NTV1-2}	0,050	*0,5*	0,874	*10,8*	0,025	*1,2*	0,149	*2,2*
α_{NTV1-3}	0,104	*1,0*	-	-	0,003	*0,3*	0,150	*2,1*
α_{NTV1-4}	0,000	*0,0*	-	-	0,005	*0,1*	0,152	*1,5*
α_{NTV1-5}	0,418	*7,2*	-	-	0,040	*0,9*	-	-
α_{NTV1-6}			-	-	0,228	*5,2*	-	-
α_{NTV1-7}			-	-	0,529	*3,7*	-	-
α_{NTV2-1}	0,074	*2,0*	0,685	*2,5*	0,116	*1,7*	0,067	*0,6*
α_{NTV2-2}	0,205	*3,7*	0,315	*2,8*	0,103	*1,6*	0,109	*1,4*

Parameters	CNL 1 (PCL)		CNL2 (full-nest)		CNL 3=CNL1 + CNL2		CNL 4 (=CoNL 4 NL)	
	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*
α_{NTV2-3}	0,236	*3,1*	-	-	0,011	*0,2*	0,238	*1,9*
α_{NTV2-4}	0,000	*0,0*	-	-	0,151	*3,4*	0,586	*2,5*
α_{NTV2-5}	0,484	*9,0*	-	-	0,077	*1,5*	-	-
α_{NTV2-6}			-	-	0,079	*4,5*	-	-
α_{NTV2-7}			-	-	0,464	*3,7*	-	-
δ_1	0,504	*0,0*	0,811	*-1,4*	0,600	*-1,0*	0,420	*-9,1*
δ_2	0,025	*-35,9*	0,150	*-31,6*	0,027	*-19,4*	0,025	*-19,6*
δ_3	0,025	*-28,5*	-	-	0,025	*-25,4*	0,026	*-18,8*
δ_4	0,112	*-5,9*	-	-	0,030	*-44,8*	0,025	*-21,0*
δ_5	0,025	*-12,7*	-	-	0,030	*-17,5*	0,170	*-32,7*
δ_6	0,528	*0,0*	-	-	0,570	*-0,5*	-	
δ_7	0,516	*0,0*	-	-	0,516	*-1,2*	-	
δ_8	0,025	*-4,9*	-	-	0,072	*-3,1*	-	
δ_9	0,025	*-34,9*	-	-	0,051	*-2,6*	-	
δ_{10}	0,492	*0,0*	-	-	0,494	*-1,2*	-	
δ_{11}	0,025	*-19,1*	-	-	0,491	*-1,0*	-	
δ_{12}	0,025	*-57,1*	-	-	0,385	*-0,5*	-	
δ_{13}	0,204	*0,0*	-	-	0,139	*-1,6*	-	
δ_{14}	0,530	*0,0*	-	-	0,026	*-40,9*	-	
δ_{15}	0,228	*-9,2*	-	-	0,079	*-2,8*	-	
δ_{16}	-	-	-	-	0,026	*-74,2*	-	
δ_{17}	-	-	-	-	0,029	*-118,0*	-	
VTTs Car [euro/h]	55,57		62,92		66,77		53,90	
VTTs Air [euro/h]	41,67		54,39		32,44		57,61	
VTTs Rail [euro/h]	36,68		38,84		44,57		35,72	

Parameters	CNL 1 (PCL)		CNL2 (full-nest)		CNL 3=CNL1 + CNL2		CNL 4 (=CoNL 4 NL)	
	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*
VTTs AE Air[euro/h]	33,87		33,13		57,24		27,28	
VTTs AE Rail [euro/h]	62,92		70,39		71,11		60,14	
maxLL	**-1869,79**		**-1915,15**		**-1814,55**		**-1913,00**	
LL(0)	-2827,65		-2827,65		-2827,65		-2827,65	
n_{par}	54		23		68		35	
ρ2	0,3387		0,3227		0,3583		0,3235	
ρ2 adj.	0,3197		0,3146		0,3342		0,3111	
$t_{calibrazione}$*[sec]*	38		4		46		25	
$t_{calibrazione}$*[min]*	0,63		0,07		0,77		0,42	
LR test (MNL 1)	238		148		349		152	
AIC	3.848		3.876		3.765		3.896	
BIC	94.892		42.654		118.413		62.906	

Table 3.4: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Cross Nested Logit estimation results.

Parameters	MMNL 1		MMNL 2		MMNL 3		Random coefficient + alt.specific-variance EC		Mixed Logit pure EC		Mixed Logit joint EC-RC	
	val.	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***
β_{cm}	-0.246	*-9.9*	-0.201	*-11.4*	-0.228	*-10.6*	-0.236	*-8.8*	-0.129	*-8.8*	-0.184	*-8.1*
$\beta_{t,Car}$	-0.078	*-15.3*	-0.289	*-11.9*	-0.315	*-9.3*	-0.199	*-12.5*	-0.103	*-10.1*	-0.434	*-6.9*
$\beta_{t,Air}$	-0.061	*-6.2*	-0.394	*-5.3*	-0.202	*-6.6*	-0.098	*-5.1*	-0.071	*-5.7*	-0.221	*-4.8*
$\beta_{t,Rail}$	-0.047	*-10.3*	-0.242	*-10.1*	-0.179	*-9.6*	-0.138	*-8.6*	-0.049	*-5.9*	-0.235	*-6.6*
β_{tAEAir}	-0.072	*-10.1*	-0.150	*-11.4*	-0.455	*-8.3*	-0.228	*-8.7*	-0.083	*-5.8*	-0.452	*-6.9*
$\beta_{tAE,Rail}$	-0.075	*-11.8*	-0.202	*-9.7*	-0.240	*-6.1*	-0.216	*-5.3*	-0.181	*-7.9*	-0.329	*-6.4*
$\beta_{half\text{-}time<60}$	-0.042	*-8.3*	-0.061	*-7.1*	-0.057	*-7.7*	-0.049	*-4.4*	-0.052	*-4.6*	-0.055	*-7.0*
$\beta_{half\text{-}time>60}$	-0.001	*-1.0*	-0.002	*-0.1*	-0.0012	*-0.8*	-0.008	*-1.9*	-0.009	*-2.3*	0.001	*0.3*
β_{degree}	1.092	*3.7*	2.458	*-2.4*	-1.045	*1.3*	-0.485	*-0.5*	0.743	*1.2*	0.012	*0.0*
$\beta_{no\text{-}cash}$	-0.082	*-0.6*	0.553	*2.1*	0.531	*3.1*	0.822	*1.9*	-0.040	*-0.1*	-0.229	*-1.0*
$\beta_{NTV\text{-}known}$	-0.850	*-5.4*	-1.445	*-7.6*	-1.575	*-7.5*	-0.860	*-1.9*	-0.748	*-1.4*	-0.445	*-1.6*
$\beta_{restaurant}$	-1.619	*-2.7*	-1.915	*-1.9*	-1.780	*-3.2*	-0.486	*-0.5*	0.700	*0.7*	-0.018	*0.0*
β_{male}	0.091	*0.2*	6.154	*0.9*	4.406	*1.7*	0.303	*0.4*	1.360	*6.1*	0.012	*0.0*
$\beta_{high\text{-}professional\text{-}condition}$	0.415	*2.1*	2.498	*6.6*	2.949	*7.2*	1.786	*4.3*	2.350	*4.4*	0.205	*1.2*
$\beta_{all\text{-}services}$	2.139	*3.3*	2.952	*2.9*	2.791	*4.0*	0.794	*0.7*	1.057	*1.1*	0.054	*0.1*
$\sigma\beta_{cm}$	0.319	*-10.7*	-0.305	*-13.6*	0.267	*11.0*	0.232	*9.2*	-	-	-0.268	*-7.4*
$\sigma\beta_{t,Car}$	-	-	-0.019	*-11.8*	0.157	*-10.1*	-0.064	*-8.9*	-	-	0.539	*6.8*
$\sigma\beta_{t,Air}$	-	-	-0.301	*8.5*	-0.056	*6.3*	-0.049	*-7.4*	-	-	0.044	*4.7*
$\sigma\beta_{t,Rail}$	-	-	-0.184	*-12.3*	-0.275	*10.4*	0.104	*9.3*	-	-	0.158	*6.9*
$\sigma\beta_{tAEAir}$	-	-	-	-	-0.229	*-8.8*	-0.027	*-3.4*	-	-	0.075	*6.8*
$\sigma\beta_{tAE,Rail}$	-	-	-	-	-0.136	*-5.0*	-0.094	*-6.8*	-	-	-0.257	*-7.4*
$\sigma\eta_{Car}$							-0.721	*-0.6*	4.630	*29.7*	0.097	*0.1*
$\sigma\eta_{Air}$							0.848	*1.4*	0.566	*0.8*	0.108	*0.2*

	MMNL 1		MMNL 2		MMNL 3		Random coefficient + alt.specific-variance EC		Mixed Logit pure EC		Mixed Logit joint EC-RC	
Parameters	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*
$\sigma\eta_{HS1}$							-4.174	*-12.4*	1.642	*8.5*	0.406	*1.9*
$\sigma\eta_{HS2}$							2.820	*8.2*	3.374	*17.3*	0.442	*2.1*
$\sigma\eta_{NTV1}$							-0.888	*-2.6*	-0.250	*-0.7*	0.184	*0.2*
$\sigma\eta_{NTV2}$							-		2.116	*10.1*	0.248	*1.1*
$\sigma\eta_{Car-Air}$							-		4.659	*7.8*	0.075	*0.2*
$\sigma\eta_{Car-HS1}$							-		1.967	*6.4*	0.383	*2.0*
$\sigma\eta_{Car-HS2}$							-		3.065	*9.9*	0.208	*1.2*
$\sigma\eta_{Car-NTV1}$							-		1.268	*5.3*	0.106	*0.5*
$\sigma\eta_{Car-NTV2}$									3.901	*23.9*	0.268	*1.0*
$\sigma\eta_{HS1-HS2}$									2.208	*6.9*	0.416	*2.4*
$\sigma\eta_{HS1-NTV1}$									5.131	*12.3*	0.232	*0.9*
$\sigma\eta_{HS1-NTV2}$									1.312	*25.5*	0.280	*1.7*
$\sigma\eta_{HS2-NTV1}$									0.657	*3.4*	0.545	*4.6*
$\sigma\eta_{HS2-NTV2}$									7.275	*14.2*	0.245	*1.1*
$\sigma\eta_{NTV1-NTV2}$									1.524	*5.6*	0.333	*2.2*
maxSLL	**-1297.72**		**-946.59**		**-930.71**		**-798.62**		**-795.49**		**-932,36**	
LL(0)	-2827.65		-2827.65		-2827.65		-2827.65		-2827.65		-2827.65	
n_{par}	16		19		21		26		32		38	
$\rho 2$	0.5411		0.6652		0.6709		0.7176		0.7187		-328.7293	
$\rho 2_{adj.}$	0.5354		0.6585		0.6634		0.7084		0.7074		-328.7427	
$t_{calibrazione}$*[sec]*	93		149		217		473		762		1044	
$t_{calibrazione}$*[min]*	1.55		2.48		3.62		7.88		12.70		17.40	
LR test (MNL 1)	1382.5		2084.764		2116.522		2380.692		2386.954		-1860740.06	
LR test (MMNL 1)	-		702.3		734.0		998.2		1004.5		-1862122.6	
AIC	2,627		1,931		1,903		1,649		1,655		1,864,794	

Parameters	MMNL 1		MMNL 2		MMNL 3		Random coefficient + alt.specific-variance EC		Mixed Logit pure EC		Mixed Logit joint EC-RC	
	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*
BIC	29,603		33,965		37,309		45,485		55,607		1,928,862	

Table 3.5: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Mixed Logit estimation results.

Parameters	Mixed NL2		Mixed CoNL 6		Mixed CoNL 6 with alternative specific variance EC		Mixed CoNL 6 with joint RC and alternative specific variance EC	
	val.	*t-test*	val.	*t-test*	val.	*t-test*	val.	*t-test*
β_{cm}	-0,176	*-10,0*	-0,062	*-8,3*	-0,107	*-8,8*	-0,024	*-8,9*
$\beta_{t,Car}$	-0,208	*-11,0*	-0,239	*-53,2*	-0,204	*-5,5*	-0,092	*-22,0*
$\beta_{t,Air}$	-0,145	*-6,4*	-0,067	*-3,3*	-0,054	*-2,2*	-0,072	*-5,7*
$\beta_{t,Rail}$	-0,116	*-9,7*	-0,210	*-12,6*	-0,053	*-3,5*	-0,021	*-10,8*
β_{tAEAir}	-0,106	*-7,7*	-0,361	*-28,4*	0,092	*4,1*	-0,138	*-8,9*
$\beta_{tAE,Rail}$	-0,151	*-6,7*	-0,302	*-4,5*	0,114	*2,6*	-0,184	*-35,2*
$\beta_{half\text{-}time<60}$	-0,054	*-7,9*	-0,019	*-5,2*	-0,039	*-3,6*	-0,004	*-3,1*
$\beta_{half\text{-}time>60}$	-0,001	*-0,4*	0,002	*2,6*	-0,002	*-0,7*	-0,001	*-1,7*
β_{degree}	-0,726	*-0,8*	2,238	*4,5*	2,994	*2,4*	-0,193	*-1,1*
$\beta_{no\text{-}cash}$	0,322	*2,0*	0,000	*0,0*	3,006	*6,8*	0,247	*6,7*
$\beta_{NTV\text{-}known}$	-1,236	*-6,9*	-0,332	*-3,8*	-1,395	*-3,8*	-0,259	*-6,6*
$\beta_{restaurant}$	-1,115	*-2,0*	-1,093	*-4,4*	-2,510	*-7,5*	-0,294	*-3,5*
β_{male}	3,427	*3,1*	12,093	*22,5*	44,029	*4,9*	1,303	*9,2*

	Mixed NL2		Mixed CoNL 6		Mixed CoNL 6 with alternative specific variance EC		Mixed CoNL 6 with joint RC and alternative specific variance EC	
Parameters	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***
$\beta_{high\text{-}professional\text{-}condition}$	2,271	*9,7*	0,747	*7,6*	1,418	*8,3*	0,263	*5,2*
$\beta_{all\text{-}services}$	1,941	*3,1*	1,378	*6,6*	3,997	*8,1*	0,478	*4,7*
$\sigma\beta_{cm}$	0,202	*10,4*	0,080	*9,9*	-	-	-0,021	*-7,3*
$\sigma\beta_{t,Car}$	0,085	*9,9*	0,137	*13,2*	-	-	0,027	*27,1*
$\sigma\beta_{t,Air}$	0,139	*8,8*	0,268	*16,2*	-	-	0,002	*1,1*
$\sigma\beta_{t,Rail}$	0,105	*9,6*	0,214	*15,1*	-	-	0,024	*25,1*
$\sigma\beta_{tAEAir}$	0,113	*9,4*	0,376	*20,1*	-	-	-0,056	*-13,5*
$\sigma\beta_{tAE,Rail}$	0,170	*10,0*	0,367	*23,3*	-	-	-0,055	*-42,9*
$\sigma\eta_{Car}$	-	-	-	-	42,537	*5,1*	-0,459	*-6,0*
$\sigma\eta_{Air}$	-	-	-	-	18,335	*3,6*	-0,212	*-0,5*
$\sigma\eta_{HS1}$	-	-	-	-	5,407	*9,6*	0,433	*9,1*
$\sigma\eta_{HS2}$	-	-	-	-	-3,105	*-8,5*	0,279	*6,2*
$\sigma\eta_{NTV1}$	-	-	-	-	3,088	*8,9*	0,136	*3,8*
w1	-	-	0,000	*0,0*	0,000	*0,0*	0,000	*0,0*
w2	-	-	0,496	*10,7*	0,891	*12,7*	0,020	*0,9*
w3	-	-	0,504	*10,9*	0,109	*1,6*	0,980	*45,7*
w4	-	-	*0,000*	-	*0,000*	-	*0,000*	-
$\delta_{no\text{-}train}$	-	-	0,514	*0,0*	0,524	*0,0*	*0,517*	*0,0*
δ_{class1}	-	-	0,506	*0,0*	0,664	*0,0*	0,557	*0,0*
δ_{class2}	-	-	0,535	*0,0*	0,579	*0,0*	0,523	*0,0*
δ_{HS}	0,046	*-56,4*	0,025	*-22,8*	0,323	*-4,5*	0,025	*-3,4*
δ_{NTV}	0,343	*-7,6*	0,258	*-9,9*	0,025	*-43,2*	0,071	*-4,8*

Parameters	Mixed NL2 val.	Mixed NL2 *t-test*	Mixed CoNL 6 val.	Mixed CoNL 6 *t-test*	Mixed CoNL 6 with alternative specific variance EC val.	Mixed CoNL 6 with alternative specific variance EC *t-test*	Mixed CoNL 6 with joint RC and alternative specific variance EC val.	Mixed CoNL 6 with joint RC and alternative specific variance EC *t-test*
$\delta_{no\text{-}Air}$	-	-	0,025	*-34,8*	1,000	*0,0*	0,088	*-66,9*
$\delta_{pub.transport(No\text{-}Car)}$	-	-	0,389	*-9,5*	0,591	*-0,5*	0,231	*-6,1*
maxSLL	**-867,90**		**-827,56**		**-793,78**		**-779,52**	
LL(0)	-2827,65		-2827,65		-2827,65		-2827,65	
n_{par}	24		31		30		35	
$\rho 2$	0,6931		0,7073		0,7193		0,7243	
$\rho 2_{adj.}$	0,6846		0,6964		0,7087		0,7119	
$t_{calibrazione}$*[sec]*			2798		3498		3691	
$t_{calibrazione}$*[min]*	0,00		46,63		58,30		61,52	
LR test (NL2/CoNL4)	2193,344		2219,364		2286,928		2315,44	
LR test (MMNL 3)	125,6		206,3		-		-	
AIC	1.784		1.717		1.648		1.629	
BIC	42.248		53.983		52.228		60.639	

Table 3.6: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Mixed CoNL estimation results.

Parameters	MCNL 1 (MPCL) val.	MCNL 1 (MPCL) *t-test*	MCNL2 (full-nest) val.	MCNL2 (full-nest) *t-test*	MCNL 4 (= MCoNL 6) val.	MCNL 4 (= MCoNL 6) *t-test*
β_{cm}	-0,169	*-15,6*	-0,162	*-136,3*	-0,186	*-314,9*
$\beta_{t,Car}$	-0,112	*-14,3*	-0,129	*-349,0*	-0,055	*-75,6*
$\beta_{t,Air}$	-0,067	*-9,8*	-0,104	*-137,2*	-0,108	*-194,7*
$\beta_{t,Rail}$	-0,096	*-14,2*	-0,148	*-267,6*	-0,131	*-135,7*
β_{tAEAir}	-0,224	*-20,0*	-0,160	*-60,5*	-0,180	*-332,2*
$\beta_{tAE,Rail}$	-0,121	*-7,7*	-0,063	*35,6*	-0,074	*-120,2*
$\beta_{half\text{-}time<60}$	-0,058	*-11,0*	-0,052	*-55,2*	-0,058	*-268,3*
$\beta_{half\text{-}time>60}$	-0,002	*-1,0*	-0,001	*-1,0*	0,000	*0,3*
β_{degree}	0,071	*0,1*	-3,273	*-1983,8*	-1,571	*-5040,9*
$\beta_{no\text{-}cash}$	0,248	*1,7*	0,228	*317,8*	0,295	*258,0*
$\beta_{NTV\text{-}known}$	-1,009	*-6,3*	-1,068	*-1875,6*	-1,303	*-1406,0*
$\beta_{restaurant}$	-1,760	*-3,7*	-1,068	*-74,9*	-1,110	*-1490,3*
β_{male}	0,910	*1,6*	-1,363	*3905,7*	-0,058	*-66,0*
$\beta_{high\text{-}professional\text{-}condition}$	1,735	*6,5*	2,103	*2110,2*	2,154	*3570,9*
$\beta_{all\text{-}services}$	2,768	*5,0*	1,711	*719,7*	2,022	*1572,9*
$\sigma\beta_{cm}$	0,224	*16,6*	0,219	*-554,1*	0,201	*262,5*
$\sigma\beta_{t,Car}$	0,084	*9,6*	0,137	*47,3*	0,301	*596,0*
$\sigma\beta_{t,Air}$	0,010	*1,6*	0,014	*-106,7*	0,133	*182,4*
$\sigma\beta_{t,Rail}$	-0,026	*-9,7*	-0,140	*-231,7*	0,127	*281,5*

Parameters	MCNL 1 (MPCL) val.	MCNL 1 (MPCL) *t-test*	MCNL2 (full-nest) val.	MCNL2 (full-nest) *t-test*	MCNL 4 (= MCoNL 6) val.	MCNL 4 (= MCoNL 6) *t-test*
$\sigma\beta_{tAEAir}$	0,140	*10,6*	0,011	*41,9*	0,151	*88,5*
$\sigma\beta_{tAE,Rail}$	-0,098	*-3,5*	0,093	*200,6*	0,083	*337,8*
$\sigma\eta_{Car}$	-	-	-	-	-	-
$\sigma\eta_{Air}$	-	-	-	-	-	-
$\sigma\eta_{HS1}$	-	-	-	-	-	-
$\sigma\eta_{HS2}$	-	-	-	-	-	-
$\sigma\eta_{NTV1}$	-	-	-	-	-	-
α_{Car-1}	0,195	*0,2*	0,464	*160,4*	1,000	-
α_{Car-2}	0,198	*0,4*	0,536	-	-	-
α_{Car-3}	0,202	*0,0*	-	-	-	-
α_{Car-4}	0,245	*0,5*	-	-	-	-
α_{Car-5}	0,161	*0,4*	-	-	-	-
α_{Car-6}	-	-	-	-	-	-
α_{Car-7}	-	-	-	-	-	-
α_{Air-1}	0,225	*0,3*	0,519	*1157,4*	1,000	-
α_{Air-2}	0,154	*0,3*	0,481	-	-	-
α_{Air-3}	0,227	*15,7*	-	-	-	-
α_{Air-4}	0,199	*0,0*	-	-	-	-
α_{Air-5}	0,195	*0,3*	-	-	-	-
α_{Air-6}	-	-	-	-	-	-
α_{Air-7}	-	-	-	-	-	-
α_{AV1-1}	0,160	*0,3*	0,056	*31,6*	0,132	*435,9*

	MCNL 1 (MPCL)		MCNL2 (full-nest)		MCNL 4 (= MCoNL 6)	
Parameters	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***
α_{AV1-2}	0,120	*0,2*	0,944	-	0,605	*404,1*
α_{AV1-3}	0,407	*0,6*	-	-	0,175	*481,4*
α_{AV1-4}	0,151	*0,4*	-	-	0,088	-
α_{AV1-5}	0,162	*1,2*	-	-	-	-
α_{AV1-6}	-	-	-	-	-	-
α_{AV1-7}	-	-	-	-	-	-
α_{AV2-1}	0,149	*0,3*	0,102	*10,9*	0,140	*710,8*
α_{AV2-2}	0,146	*0,3*	0,898	-	0,638	*647,2*
α_{AV2-3}	0,435	*0,0*	-	-	0,170	*880,1*
α_{AV2-4}	0,155	*0,0*	-	-	0,052	-
α_{AV2-5}	0,114	*0,4*	-	-	-	-
α_{AV2-6}			-	-	-	-
α_{AV2-7}			-	-	-	-
α_{NTV1-1}	0,212	*0,2*	0,653	*927,6*	0,327	*331,0*
α_{NTV1-2}	0,166	*0,5*	0,347	-	0,086	*29,8*
α_{NTV1-3}	0,167	*2,5*	-	-	0,337	*451,2*
α_{NTV1-4}	0,235	*0,3*	-	-	0,250	-
α_{NTV1-5}	0,220	*0,3*	-	-	-	-
α_{NTV1-6}	-	-	-	-	-	-
α_{NTV1-7}	-	-	-	-	-	-
α_{NTV2-1}	0,222	*0,0*	0,998	*2758,0*	0,341	*310,6*
α_{NTV2-2}	0,173	*5,6*	0,002	-	0,028	*807,3*
α_{NTV2-3}	0,222	*0,5*	-	-	0,393	*527,3*

	MCNL 1 (MPCL)		**MCNL2 (full-nest)**		**MCNL 4 (= MCoNL 6)**	
Parameters	**val.**	***t-test***	**val.**	***t-test***	**val.**	***t-test***
α_{NTV2-4}	0,155	*0,2*	-	-	0,238	-
α_{NTV2-5}	0,227	*0,0*	-	-	-	-
α_{NTV2-6}	-	-	-	-	-	-
α_{NTV2-7}	-	-	-	-	-	-
δ_1	0,980	*0,0*	0,996	*-182,5*	0,986	*-12,9*
δ_2	0,994	*0,0*	0,306	*-774,5*	0,638	*-821,0*
δ_3	0,993	*-0,2*	-		0,442	*-2043,9*
δ_4	0,578	*-0,6*	-		0,988	*-76,8*
δ_5	0,978	*-0,2*	-		0,979	*-32,5*
δ_6	0,964	*-0,1*	-		0,985	*-18,0*
δ_7	0,994	*0,0*	-		-	-
δ_8	0,975	*-0,1*	-		-	-
δ_9	0,928	*-0,1*	-		-	-
δ_{10}	0,136	*-10,3*	-		-	-
δ_{11}	0,984	*-0,3*	-		-	-
δ_{12}	0,994	*-0,1*	-		-	-
δ_{13}	0,982	*-0,5*	-		-	-
δ_{14}	0,996	*0,0*	-		-	-
δ_{15}	0,993	*-0,1*	-		-	-
δ_{16}	-	-	-		-	-
δ_{17}	-	-	-		-	-
maxSLL	**-963,25**		**-918,37**		**-937,51**	

Parameters	MCNL 1 (MPCL) val.	t-test	MCNL2 (full-nest) val.	t-test	MCNL 4 (= MCoNL 6) val.	t-test
LL(0)	-2827,65		-2827,65		-2827,65	
n_{par}	60		29		39	
ρ2	0,6593		0,6752		0,6685	
ρ2 adj.	0,6381		0,6650		0,6547	
$t_{calibrazione}$*[sec]*	1801		550		1172	
$t_{calibrazione}$*[min]*	30,02		9,17		19,53	
LR test (CNL 1/CNL 2/CNL 4)	1904		1947		1951	
LR test (MMNL 1/ MMNL 2/ MMNL 3)	669		25		-14	
AIC	2.046		1.895		1.953	
BIC	103.206		50.789		67.707	

Table 3.7: Model estimation on SP survey for six alternatives mode choice Naples-Milan – Mixed CNL estimation results.

Chapter 4: Random utility models: regression vs forecasting

This chapter proposes a comparative analysis of the performance of various RUMs – namely Multinomial Logit, Nested Logit, Cross Nested Logit, FinMix and CoNL– estimated on a synthetic dataset with variable sample size and correlation patterns. This experimental framework allows comparing model estimates in a fair, controlled environment wherein all relevant characteristics (coefficients, attributes, covariances, likelihood, elasticities) of the "true" underlying model are known.
Models are validated especially by comparing true and estimated market share elasticities where the market share is the sum over all observations of the individual probabilities of a given alternative. Indeed, this indicator represents the real forecasting capability of a model, that is the main target for the analyst. Moreover, its true value can be computed when dealing with a synthetic database by evaluating the difference in the number of choices of a given alternative between future and current scenarios, due to a difference in some attribute's value.
Comparisons are carried out on several choice contexts, characterized by different correlation matrices and variable sample size.

4.1 Closed-form R.U.M. and forecasting

Random utility models (RUMs) are a powerful tool for reproducing individual choice probabilities in possibly complex choice contexts. A considerable variety of closed-form RUMs – i.e. whose probability statement is expressed by a closed-form analytical formulation – has been proposed, mostly belonging to the family of Generalised Extreme Value (GEV) models (Mc Fadden, 1978)
The simplest model of this family is the Multinomial Logit model (MNL; Luce, 1959), which is based on the I.I.A. (independence of irrelevant alternatives) property and allows to reproduce choice contexts where the utilities of the alternatives are not correlated.
The Nested Logit model (NL; Mc Fadden, 1978; Williams, 1977; Ben-Akiva and Lerman, 1977; Daly and Zachary, 1978) introduces correlations among alternatives belonging to the same nest, that is a block diagonal correlation structure. It has the advantage of a closed-form correlation expression, enabling a proper and conscious interpretation of model estimation results. However it has a limited correlation flexibility, i.e. capability of reproducing any correlation matrix.

Interesting, successive generalization of the NL model, overcoming NL limited correlation flexibility, are all the CNL logit family models, allowing each alternative to belong to more than one nest. Example of these kind are the Ordered GEV (Small, 1987), the Paired Combinatorial Logit (Chu, 1989; Koppelman and Wen, 2000) the Cross-Nested Logit (Vovsha, 1997), the Generalized Nested Logit (Wen and Koppelman, 2001), the Recursive Nested Extreme Value (Daly, 2001), the Network GEV (Daly and Bierlaire, 2006; Newman, 2008).
Unfortunately, all these models have not a closed-form correlation expression.
Other interesting NL generalization characterized by very flexible correlation matrix have been obtained as finite mixtures of NL models. Example of this kind are the FinMix (Swait, 2003) and the CoNL (Papola, 2016). The latter, in particular, has the advantage of a closed-form correlation expression.
Generally, when estimating models in complex choice contexts, that is contexts with complex expectations in terms of correlation matrix, more complex models perform better in terms of both in-sample and out-of-sample goodness of fit tests (rho square, adjusted rho square, etc.).
However, all these tests have a comparative value and are difficultly interpretable in terms of quality/price ratio by practitioners.
In this chapter, some greater comprehension of the real operational advantage provided by more complex models is provided, through their estimation with synthetic dataset. The latter, indeed, allows evaluating model performances in terms of more interesting and interpretable indicators, like the real model forecasting capability, which represents the real interest in model application.

4.2 Model's elasticities

As mentioned at the end of the previous section, the main goal when estimating a RUM is to maximize its forecasting capability, i.e. the capability of predicting correctly possible market share variations generated by variations in some of the relevant attributes. Hence, the forecasting capability of a RUM is well-represented by the model's elasticity, which actually measures the variation of the model's choice probabilities corresponding to a perturbation of a generic attribute's value. The elasticity is generally analysed at the individual level with the following formulations:

$$E_{kj}^{p(j)} = \frac{\dfrac{\Delta p(j)}{p(j)}}{\dfrac{\Delta \bar{X}_{kj}}{\bar{X}_{kj}}} \simeq \frac{\dfrac{dp(j)}{p(j)}}{\dfrac{d\bar{X}_{kj}}{\bar{X}_{kj}}} \tag{4.1}$$

$$E_{kh}^{p(j)} = \frac{\dfrac{\Delta p(j)}{p(j)}}{\dfrac{\Delta \bar{X}_{kh}}{\bar{X}_{kh}}} \simeq \frac{\dfrac{dp(j)}{p(j)}}{\dfrac{d\bar{X}_{kh}}{\bar{X}_{kh}}} \tag{4.2}$$

where X_{kj} represents the generic attribute used for defining the utility of alternative j, while $E_{kj}^{p[j]}$ and $E_{kh}^{p[j]}$ represent direct and cross elasticities, i.e. the percentage variations of the choice probability of alternative j corresponding to a percentage variation of an attribute present in the same alternative j or in another alternative h respectively.
Some more insights on this general concept are provided by Ben-Akiva and Richards(1975) Domencich and Mc Fadden(1975), Dunne (1984) and Ben-Akiva and Lerman (1977), while

specific application to the GEV models are provided by Wen and Koppelman (2001) and Train (2009).
Elasticities (4.1) and (4.2) can be computed also by substituting individual choice probabilities with market shares, i.e. the sum over all observations of the individual probabilities of a given alternative. The result is a "market share elasticity", i.e. the percentage variation of the total demand of a certain good in correspondence of the percentage variation of the value of a certain attribute, which, on the other hand, represents the main target for the analyst.
Importantly, a synthetic dataset - like that used in this analysis – allows the calculation of the "true" value of this elasticity by calculating the difference in the number of choices of a given alternative between future and current scenarios, due to a difference in some attribute's value. Hence, with a synthetic dataset, the real forecasting capability of a model can be evaluated, by comparing true and estimated values of this kind of elasticities.

4.3 Experimental analysis

4.3.1 Experiment setting

As mentioned in the previous sections, and following a well-established and common approach in the literature, the different RUMs mentioned in the first section were estimated using synthetic datasets, allowing estimation results to be compared in a fair and controlled experimental setting.
These datasets were generated analogously to what done in Papola (2016) as described below, in terms of observations, choice context and definition of:

- observable components of the utilities (systematic utilities);
- unobservable components of the utilities (random residuals);
- choice.

Observation and choice context: The synthetic datasets encompass a variable number of observations (200,1000, 5,000, 10,000, 100,000, 1,000,000) related to hypothetical three-alternatives and four-alternatives choice contexts; each observation is associated with the full set of alternatives.

Definition of the systematic utilities. The systematic utility of each alternative is given by a linear combination of two specific attributes. The parameters β's of these linear combinations do not vary across observations, while values of the attributes for each observation are generated through independent random draws from a Normal variable with mean and variance defined as in Table 4.1, corresponding to a coefficient of variation equal to 0.1.

Definition of the random residuals. Random residuals are assumed to follow a multi-variate normal distribution with zero-vector mean and predefined homoscedastic covariance matrix. Operationally, given a covariance matrix, random residuals for each alternative can be obtained through its Cholesky factorization:

$$\boldsymbol{\varepsilon} = \mathbf{F} \cdot \mathbf{z} \tag{4.3}$$

in which F is the lower triangular matrix defined by the Cholesky factorization on the variance-covariance matrix Σ.

Definition of the choice. The alternative chosen for each observation is the one with maximal utility.

Knowledge of the ground truth behind the estimation sample allows estimated factors to be contrasted with true ones. This applies to taste parameters β, to correlation matrices (whose estimate can be computed through model structural parameters) and, especially, to market share elasticities.
Specifically, as mentioned in the previous section, the true market share elasticities can be computed by calculating the percentage variation in the total number of choices of a certain alternative in correspondence of the percentage variation of the value of a certain attribute.
Concerning correlations, it must be highlighted that while for NL and CoNL the estimated correlations can be computed through the existing closed-form formulas, CNL and Fin-Mix correlations must be computed through integral calculation, as suggested for instance by Marzano et al. (2013) and Marzano (2014).
Several types of correlation contexts are assumed, with increasing number of non-zero correlations, hence requiring an increasing model flexibility in terms of reproducible correlation matrices.

4.3.2 Specification of the structure of the tested models

Compared models are MNL, NL, CNL, CoNL and Fin-Mix. For the sake of brevity, their structure - in terms of nesting structure - are depicted in Figure 4.1 and Figure 4.2. Both CoNL and FinMix were specified by mixing binary nests. Conversely, the CNL was specified by using "full nests" – i.e. nests including all the alternatives - as suggested by Marzano and Papola (2008).

4.3.3 Experimental results

The estimation of the different models in the different choice contexts, with variable sample size, were carried out with Matlab. For validation purpose some estimation experiments were carried out also with different software: R, MS Excel and BioGEME (Bierlaire, 2003). Main outputs are taste parameters, correlations and market share elasticities which have been contrasted with the corresponding true values.
For the sake of brevity, only estimation results concerning the market share elasticities are showed, with the aid of a set of synthetic plots. Indeed, as mentioned in the introduction, this indicator represents the real forecasting capability of a model and hence the main interest of the analyst when applying a model.
In this plots, in particular, a synthetic elasticity indicator (I_e) is reported, representing the mean square error between true and modelled market share elasticities (both direct and cross). In more detail, for any model - characterized by a specific colour - the elasticity performances are plotted as a function of the sample size, in a semi-logarithmic scale. The same kind of plot is presented for several correlation contexts, related to the "three-alternative" (A,B,C) and the "four-alternative" (A,B,C,D) choice contexts.
Conversely, the goodness-of-fit measures trend, with reference to the sample size, is shown, particularly the adjusted r2 and the ratio between the optimum log-likelihood value and the sample size (called normalized log-likelihood). The objective of the comparison is contrasting the errors of the analysed models in terms of forecasting and the goodness of the estimated models in terms of fitting.

A first comment refers to the great importance of the sample size which help significantly in reaching better performances whatever the model. The main differences are observed when passing form hundreds to thousands of observations.
A second main comment is the general capability of reproducing almost perfectly the true market elasticities in all correlation contexts, if using the "correct" model, i.e. a model with an underlying correlation pattern compatible with the correlation context assumed as true.

	Alternative 1				Alternative 2				Alternative 3				Alternative 4			
	Attributes		Parameters		Attributes		Parameters		Attributes		Parameters		Attributes		Parameters	
	X_1	X_2	β_1	β_2	X_3	X_4	β_3	β_4	X_5	X_6	β_5	β_6	X_7	X_8	β_7	β_8
Av.	8	2	1	6	5	2	2	5	4	2	3	4	2	5	5	2
St.dv.	0.8	0.2	-	-	0.5	0.2	-	-	0.4	0.2	-	-	0.2	0.5	-	-

Table 4.1: Experimental setting.

On the other hand, and even more importantly, the performance of the "wrong models" can be significantly worse with respect to those of the correct models.
The first experiment to be shown is the one of Figure 4.3 (three alternatives). This experiment is actually shown as check experiment, because its simplicity allows the reader to create a precise expectation on the results. This is a typical one-level Nested correlation scenario, wherein only one true value of the correlation is set to be different from 0. Particularly, the experiment wants to show a boundary case of nested correlation (almost total), so the true value of ρ_{AB} is fixed to *0.95* (the reader can refer to the well-known Daganzo and Sheffi network, cited in Chapter 2 and depicted in Figure 5.5). The NL is the natural candidate to reproduce such situation, while the MNL fails to reproduce it, due to the limitations of the already mentioned I.I.A. property. In fact, in this case, MNL goodness-of-fit measures are clearly worse than the other models and the error in reproducing the true elasticities is absolutely significant. Conversely, NL reproduce very well the true elasticities (perfectly for sample size greater than 5,000 observations). More complex models collapse to a NL and their performances practically coincide among them and with those of the NL.
The second and third correlation experiments (again three alternatives), are instead incompatible with a NL model. In this case, for big sizes of the sample, not only MNL but even NL error in reproducing the true elasticities is significant. But looking at the small sample sizes (200) there is the evidence of a contrasting behaviour. The models with better fitting, i.e. those with higher absolute value of the goodness of fit measures, perform worse in terms of forecasting, i.e. its mean square indicator is higher. See, for instance, the CNL, CoNL and FinMix in Figure 4.4 and, particularly, the FinMix in Figure 4.5.
This "small sample effect", probably due to overfitting problems, is clearly evident in all the successive figures proposed (see FinMix and CoNL in Figure 4.6 and Figure 4.7, all the complex models in Figure 4.8 and Figure 4.9, the CNL in Figure 4.10).In other words, to express an unbiased forecast, it is necessary to work with a big enough sample of data. Figure 4.11 and Figure 4.12 enhance the same effect on the 4 alternatives context, particularly with reference to the CNL performance.
Increasing to more than a thousand observations, the forecasting capability of the appropriate model, i.e. the models which structurally allows to reproduce the considered correlation scenario, becomes stable.

Thus, with appropriate sample size, more complex models generally perform significantly better, even if their more complex expression may generate some algorithmic problems in finding the optimal solution: see for instance the performances of the Fin-Mix in Figure 4.10.
In terms of estimation time, referring to the maximum sample size (10^6 observations), MNL and NL are generally very efficient with estimation times around few minutes. CNL and CoNL estimation times are a few dozen minutes, while Fin-Mix require generally a few hours.

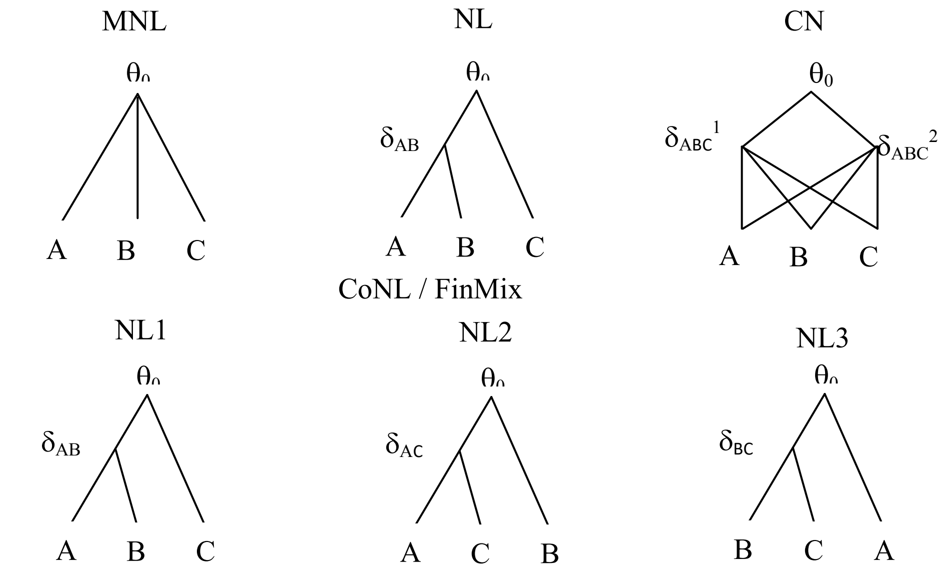


Figure 4.1: Model's specification for three alternatives-context.

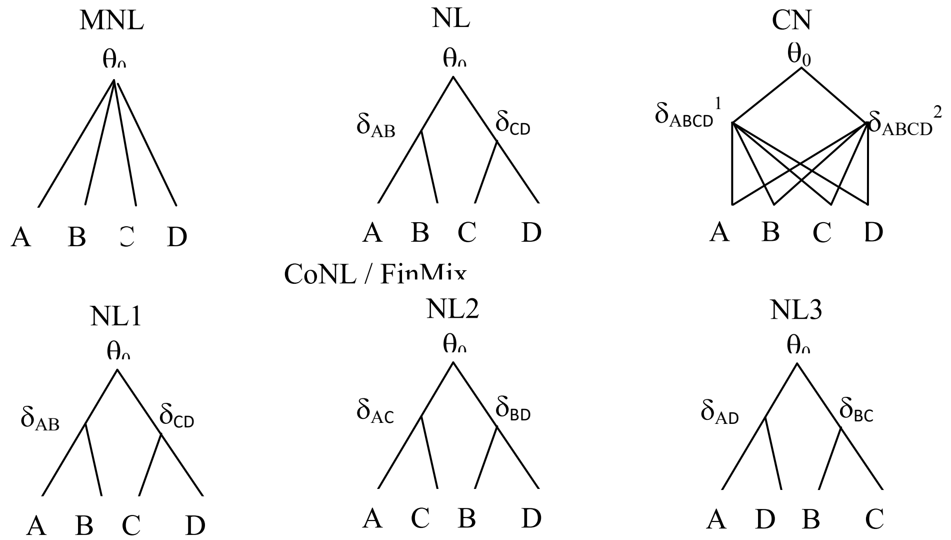


Figure 4.2: Model's specification for four alternatives-context.

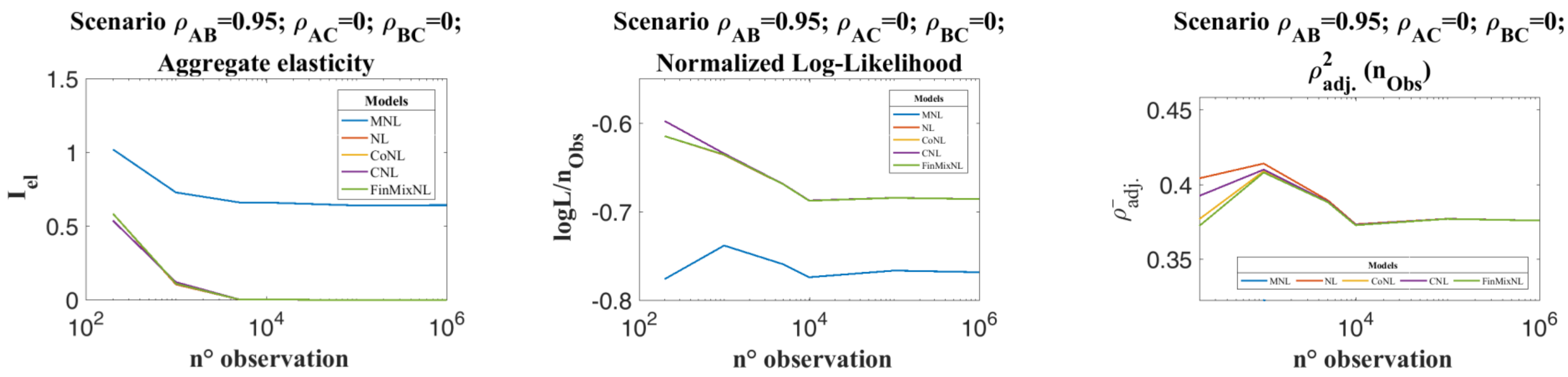


Figure 4.3: 0.95-0-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

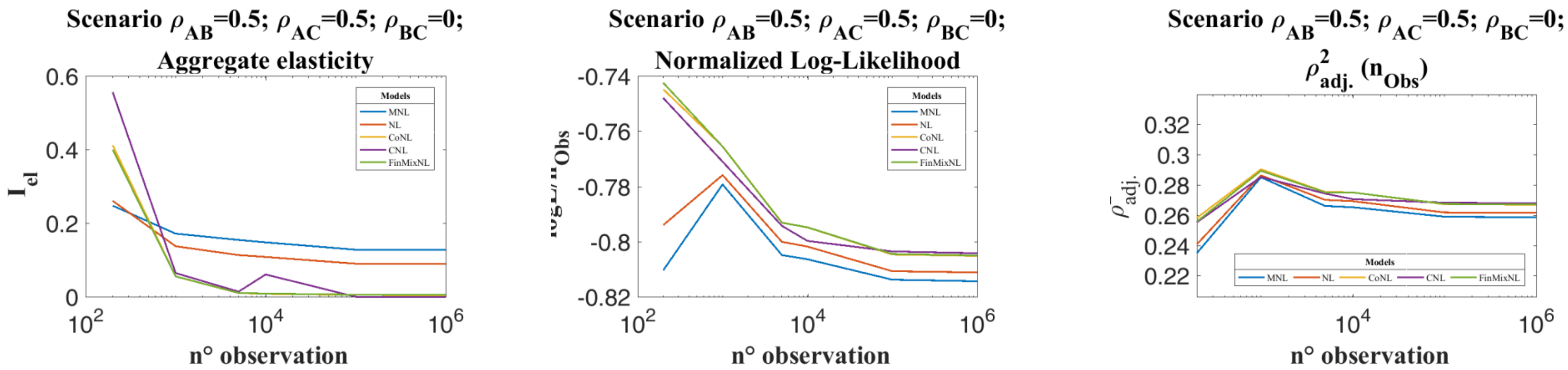


Figure 4.4. 0.5-0.5-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

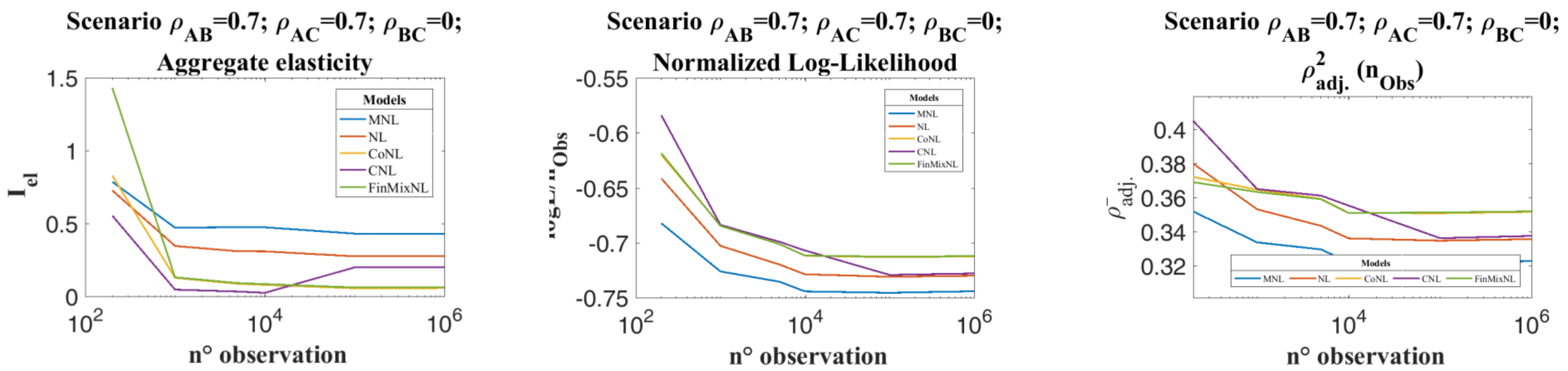


Figure 4.5: 0.7-0.7-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

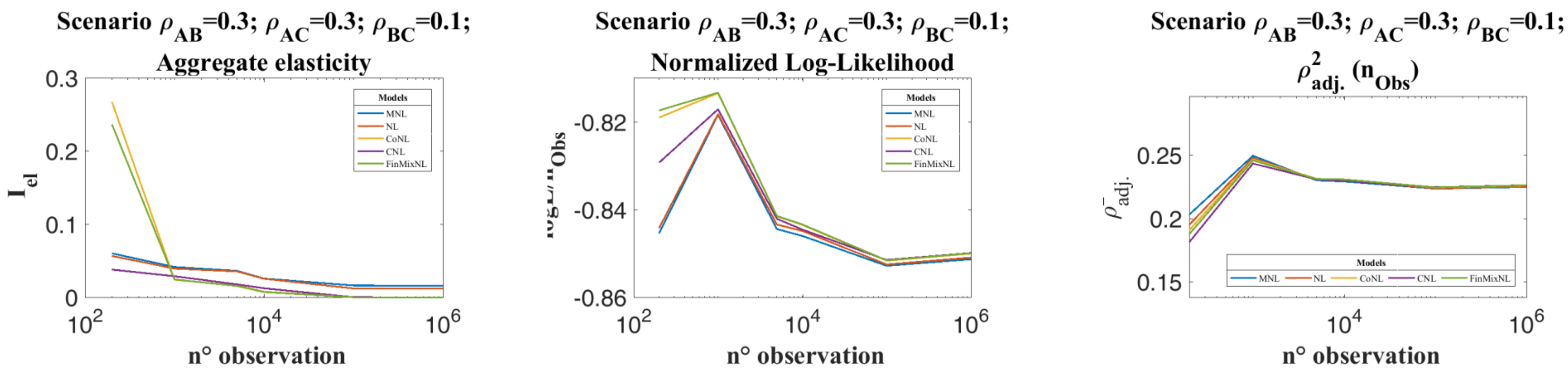


Figure 4.6: 0.3-0.3-0.1 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

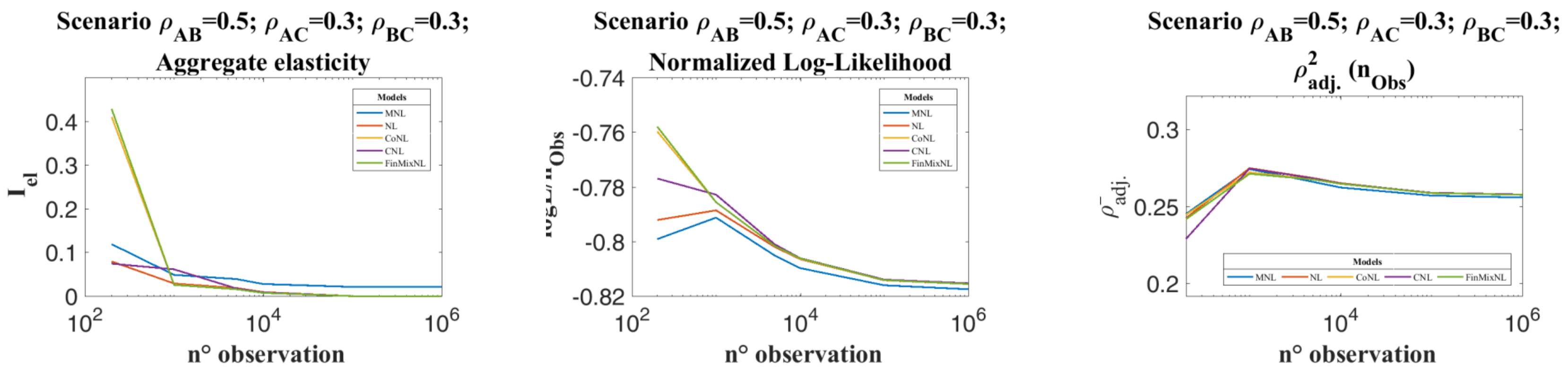


Figure 4.7: 0.5-03-0.3 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

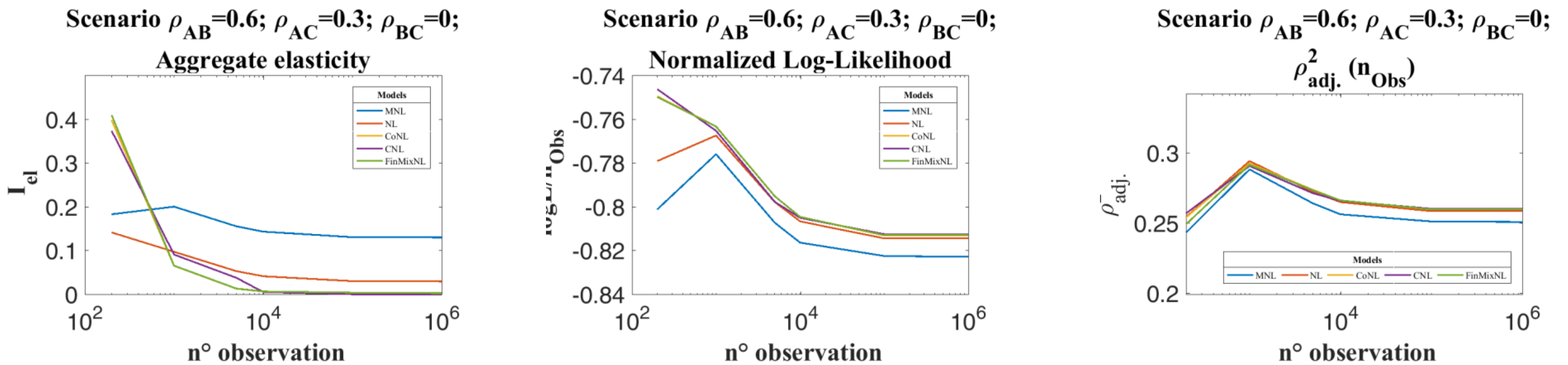


Figure 4.8: 0.6-0.3-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

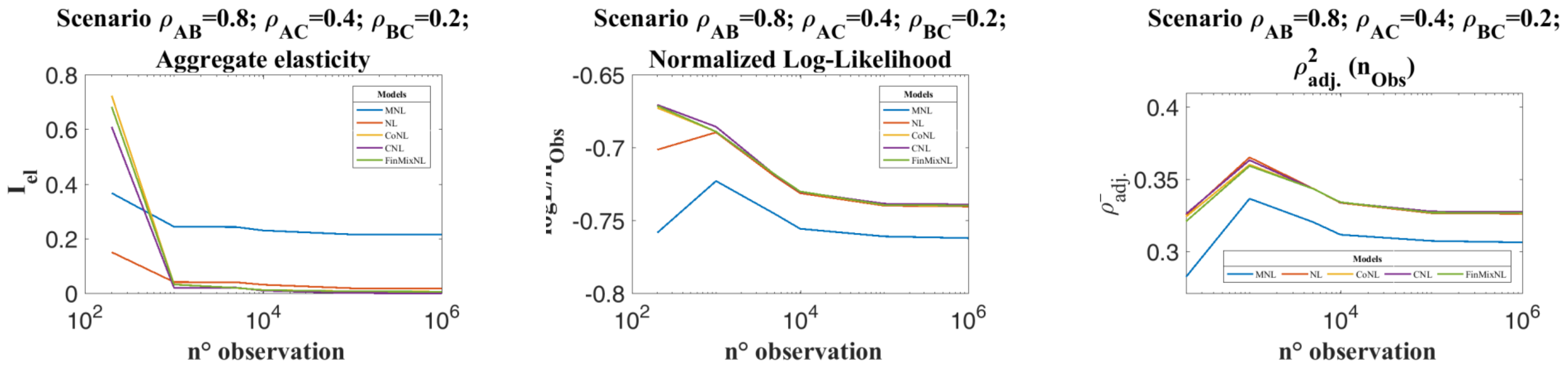


Figure 4.9: 0.8-0.4-0.2 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

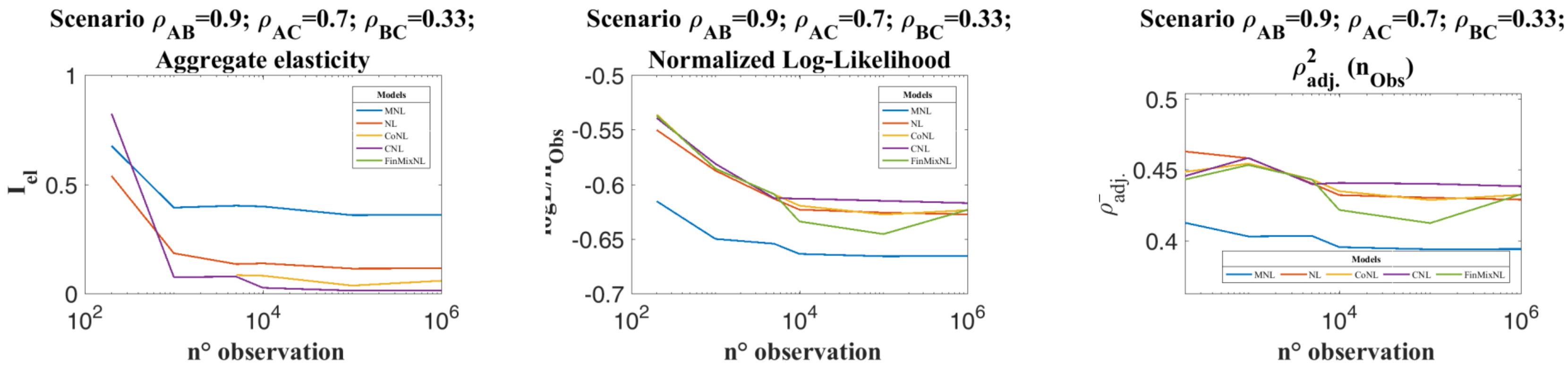


Figure 4.10: 0.9-0.7-0.33 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

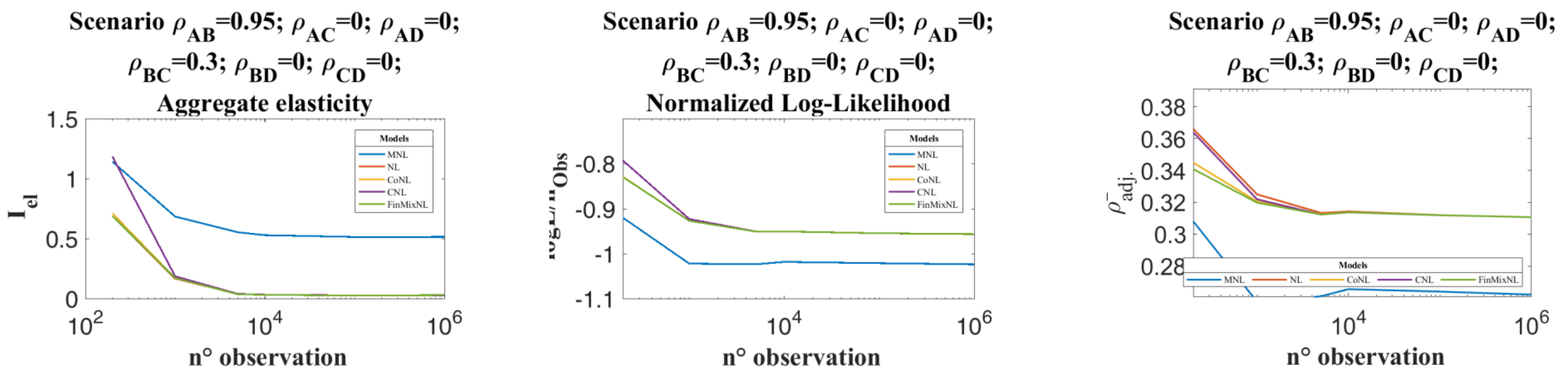


Figure 4.11: 0.95-0-0-0.3-0-0 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

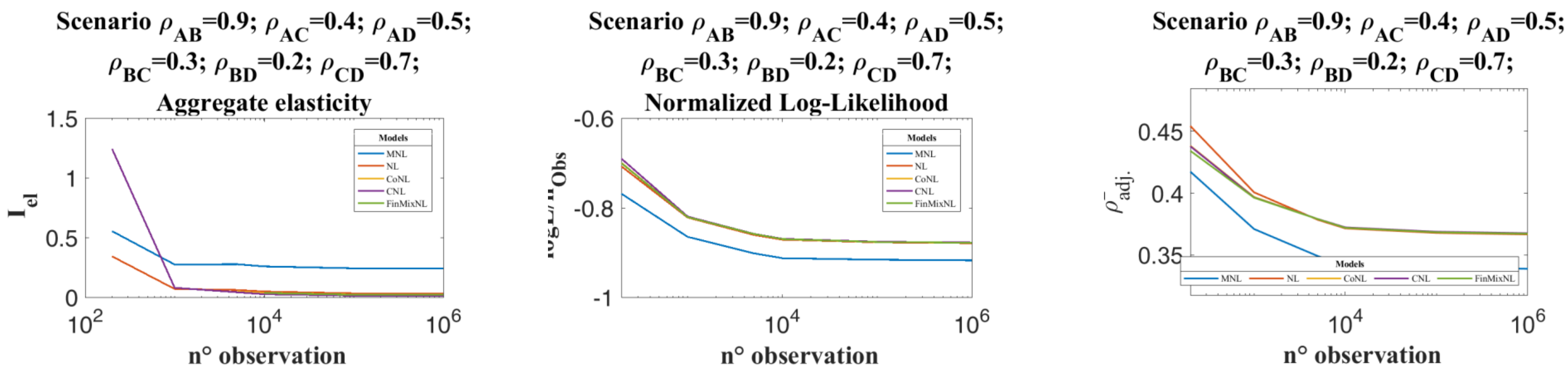


Figure 4.12: 0.9-0.4-0.5-0.3-0.2-0.7 correlation scenario – synthetic performance plots indicator for forecasting (column 1) and regression (column2 and 3).

4.4 Conclusions and future steps

Despite a high number of random utility models were proposed in the literature, very few of them are actually used by practitioners. In this chapter the real advantage of using a more complex model in a more complex choice context want to be acknowledged through estimation experiments on synthetic datasets. In order to do that, a new elasticity indicator is used, i.e. the market share elasticity. Indeed, from one hand, this indicator represents the main target for the analyst, since it indicates the percentage variation of the total demand of a certain good in correspondence of the percentage variation of the value of a certain attribute; from the other, its true value can be computed if dealing with a synthetic dataset, by calculating the percentage variation in the total number of choices of a certain alternative in correspondence of the percentage variation of the value of a certain attribute.

Specifically a synthetic elasticity indicator (I_e) has been used for analysis, representing the mean square error between true and modelled market share elasticities (both direct and cross). In other words, I_e represents a measure of the real forecasting capability of a model.

On the basis of this indicator, different models, namely MNL, NL, CNL, CoNL and FinMix, are compared on different choice contexts, characterized by different correlation matrices and variable sample size.

A first main comment refers to the great importance of the sample size which help significantly in reaching better performances whatever the model. In fact, the evidences show how overfitting issues may generate very unbiased forecasting if the sample size is too small (few hundreds). Also in very simple cross-sectional experiments like these presented in the chapter, using a complex model able to catch better fit the data, because of its structural capacity to reproduce complex correlation scenarios, can be a double edge weapon if the sample is small. To exploit the actual capacity of these models, namely the CNL, the CoNL and the FinMix, is absolutely necessary to estimate their parameters with several thousands of observations. If not, the risk of wrong policy choices can be dramatic.

The main differences are observed when passing form hundreds to thousands of observations.

A second main comment is the general capability of reproducing almost perfectly the true market elasticities in all correlation contexts, if using the “correct” model, i.e. a model with an underlying correlation pattern compatible with the correlation context assumed as true.

On the other hand, and even more importantly, the performance of the “wrong models” can be significantly worse with respect to those of the correct models. In other words, using the correct model, in each specific choice context, may improve significantly the analyst capability of make correct demand forecasting.

Hence, the use of complex models in complex choice contexts should be absolutely promoted by practitioners.

In terms of future research steps, this work need absolutely to be expanded, generalized and extended, for instance by analysing links and correlations among the different indicators: are the actual indicators (t-stats, rho square, etc.) able to correctly address us towards the best forecasting performances when specifying a model? Can we think about some more effective indicators?

Chapter 5: The CoNL route choice model

The chapter illustrates a route choice model based on the recently proposed Combination of Random Utility Models (CoRUM) model by Papola (2016). Specifically, the CoRUM specification with nested logit components, termed Combination of Nested Logits (CoNL), accounts for any correlation patterns amongst alternatives while keeping a closed-form expression for both probabilities and correlations. Thus, the chapter illustrates a CoNL route choice model capable to target correlations between routes based on their topological overlapping, and characterized by a closed-form probability statement. The CoNL route choice model is operationalized by means of an algorithm providing automatically CoNL specification and corresponding route choice probabilities on a set of enumerated paths. Model performance is tested on various networks, with very satisfactory results.

5.1 Background and motivation

Route choice modelling is a key topic in transport engineering, with a well-established research stream spanning over more than thirty years of relevant literature, summarized in various state-of-the-art reviews, including Ramming (2001), Prato and Bekhor (2007), Prato (2009), Papola and Marzano (2013). Researchers and analysts acknowledge unanimously some distinct, unique challenging features characterizing route choice contexts, primarily along three main viewpoints.

The first deals with the behavioural framework underlying route choices by decision-makers. In general, the concept of route itself as elemental choice alternative can be questioned. Various alternative paradigms have been formulated in the attempt to explain how decision-makers actually perceive a route, for instance based on sequences of waypoints, or on destination-oriented macro-directions. In terms of behavioural choice mechanism, many approaches have been explored with success, for instance the application of prospect theory to model risk-seeking and risk-averse behaviour (e.g. Katsikopoulos et al. 2000; Avinieri and Prashker, 2004; de Palma et al., 2008; Gao et al., 2010; de Luca and Di Pace, 2015). Notwithstanding, the classical definition of route as an ordered sequence of links connecting an origin-destination o-d pair and the Random Utility Models (RUMs) framework still represent the most effective assumptions to operationalize a route choice model for large-scale transport applications.

The second deals with the considerable number of alternative routes usually available for each o-d pair. As a first consequence, assuming full knowledge/perception of the choice set by decision-makers is unrealistic. This problem is circumvented often by applying a route choice set generation method prior to the route choice model, that is selecting a subset of routes to choose from based on heuristic rules (Ben-Akiva et al., 1984; De La Barra et al., 1993; Azevedo et al., 1993; Bekhor et al., 2001; Van der Zijpp and Catalano, 2005; Prato and Bekhor, 2006; Bekhor et al., 2006). In addition, the implementation of RUM-based route choice models on real-size network requires developing algorithms and/or procedures capable to specify effectively and with limited computational burden the route choice model directly from the graph representing the underlying transport network.

The third deals with the presence of a complex correlation structure amongst perceived utilities of route choice alternatives, structurally determined by the topological overlapping of alternative routes in a transport network. Under the usual definition of route as an ordered sequence of links in the network, there is consensus in the literature towards considering the Daganzo and Sheffi (1977) assumption – that is, a correlation between pair of routes proportional to their topological overlapping, measured using a given link impedance – as a key reference. Along this line, Frejinger and Bierlaire (2007) proposed the so-called subnetwork approach, that is an application of the assumption by Daganzo and Sheffi (1977) only to a portion of the network given by primary, most likely perceived, roads.

Summarizing, notwithstanding many variations on the theme, a prevailing and widely adopted research track in route choice modelling is to specify RUM-based route choice models with routes defined as ordered sequences of links and underlying correlation consistent with the Daganzo and Sheffi (1977) assumption.

Within this track, many relevant contributions are available in the concerned literature. The simplest model was proposed by Dial (1971), who applied the Multinomial Logit model (MNL) to route choice with an elegant and computationally very effective algorithm to calculate route choice probabilities without explicit route enumeration. Unfortunately, the MNL model hypothesizes null correlation amongst perceived utilities of alternatives, because of its underlying assumptions. Thus, Daganzo and Sheffi (1977) operationalized the Multinomial Probit model (MNP) as a natural for embedding their assumption in a route choice model, thanks to the possibility offered by the MNP model to specify directly its correlation matrix. However, the MNP suffers from the absence of a closed-form probability statement, leading to computational issues related to the need to simulate choice probabilities, see e.g. Horowitz (1982), McFadden (1989), Bunch (1991), Geweeke (1991), Train (2009), Connors et al. (2014). The same also apply to Mixed Logit applications to route choice, e.g. Bekhor et al. (2002), Frejinger and Bierlaire (2007).

A natural alternative research direction aimed at developing closed-form route choice models leveraging the class of Generalised Extreme Value (GEV) models proposed by McFadden (1978). Many models have been proposed so far in this context, including the Link-Nested Logit (LNL) model by Vovsha and Bekhor (1998); the Paired-Combinatorial Logit (PCL) model by Prashker and Bekhor (1998); the Path Multilevel PML model by Papola and Marzano (2013). A noticeable variation on the theme is represented by the so-called recursive models (Fosgerau et al., 2013; Mai et al., 2015; Mai, 2016). However, there is no evidence in the literature on their capability to target Daganzo and Sheffi (1977) correlations. Alternatively, several researchers tried to introduce correction/penalty factors in the systematic utility of a MNL model to mimic the effect of correlations on route choice probabilities, for instance the C-Logit model by Cascetta et al (1996), and Russo and Vitetta (2003), and the Path-size model by Ben-Akiva and Ramming (1998), Ben-Akiva and Bierlaire (1999), Ramming (2001), Hoogendoorn-Lanser et al. (2005). These models exhibit limitations in capturing the proper effect of route correlations on choice probabilities, as addressed amongst others by Prashker and Bekhor (1998), Prashker and Bekhor (2004), Marzano (2005), Papola and Marzano (2013).

In the light of the above, an interesting opportunity for route choice modelling is offered by the Combination of RUMs (CoRUM) model proposed by Papola (2016), particularly its Nested

Logit component-based form, termed Combination of Nested Logit (CoNL). For the purposes of this chapter, the key feature of the CoRUM, and thus of the CoNL, is the availability of a closed-form statement for both choice probabilities and correlations. This allows handling effectively the relationship between the CoNL specification (model structure, parameters) and its underlying correlations, thus enabling the possibility of specifying a CoNL capable to target Daganzo and Sheffi (1977) correlations.
In a nutshell, the cumulative distribution function (cdf) of a CoRUM is defined as a convex combination of cdf's of other RUMs, termed mixing components of the CoRUM. All mixing components should embed by definition the same alternatives in the choice set, that will be the choice set also of the resulting CoRUM. Such specification, resembling a latent class model (Gopinath, 1995; Bhat, 1997; Swait, 1994; Greene, 2001; Greene and Hensher, 2003; Walker and Li, 2007; Vij et al., 2011), allows expressing probability statements, correlations, and elasticities of a CoRUM as a convex combination of the corresponding expressions of the mixing components. As a matter of fact, choosing Nested Logit (NL) models as mixing components, i.e. components with closed-form probability statements and correlations, yields a CoNL with corresponding closed-form expressions. In addition, as addressed by Papola (2016), the NL components can be specified so as to obtain a CoNL correlation matrix with maximal flexibility.
Primary target of this chapter is to explore the applicability of the CoNL model to the route choice context, proposing a new route choice model: (a) capable to target Daganzo and Sheffi (1977) correlations, (b) characterized by a closed-form expression of choice probabilities, and (c) operationalized by means of an algorithm providing automatically the specification and the route choice probabilities of a CoNL route choice model on a transport network. The structure of the chapter is the following: Section 5.2 recalls key features of the CoRUM and of the CoNL models, Section 5.3 describes the specification of the CoNL model for route choice modelling, Section 5.4 introduces a methodology and an algorithm for its operationalization on real-size networks, Section 5.5 provides tests on synthetic and real-size networks, Section 5.6 draws conclusions and research prospects.

5.2 The CoRUM model and its CoNL particularization

This section entirely reprises the Section 2.4, providing the notation for adapting it to the route choice. In the following, the apices n and t will be suppressed.
Following Papola (2016), the CoRUM is an additive Random Utility Model (RUM), whose underlying cumulative distribution function (cdf) is the linear combination of a set I of n_{MC} absolutely continuous cdf's of n random residuals $F^i(\varepsilon_1,..., \varepsilon_n)$ $i\in\mathrm{I}$ with n_{MC} non-negative weights $w^1,..., w^{nMC}$ summing up to 1, that is:

$$F(\varepsilon^{n,t}_{1},...,\varepsilon^{n,t}_{m}) = \sum_{c\in\mathrm{C}} w^c \cdot F^i(\varepsilon^{n,t}_{1},...,\varepsilon^{n,t}_{m}) \tag{2.115}$$

By definition, the properties of the CoRUM model (2.115) can be derived straightforwardly from well-known mathematical properties of the mixture of cdf's. In particular, the probability statement of a CoRUM is given by the mixture of the probability statements of the corresponding mixing components, with the same weights:

$$p(j) = \sum_{i\in\mathrm{I}} w^i \cdot p^i(j) \qquad \forall j \tag{2.118}$$

and its variances/covariances are linked to the corresponding variances/covariances of the mixing components via the following expressions:

$$Cov[\varepsilon_j \varepsilon_m] = \sum_{i \in \mathrm{I}} w^i \cdot Cov^i[\varepsilon_j \varepsilon_m] \quad \forall j,m \tag{2.120}$$

$$Var[\varepsilon_j] = \sum_{i \in \mathrm{I}} w^i \cdot Var^i[\varepsilon_j] \quad \forall j \tag{2.119}$$

Expressions (2.119)-(2.120), in the light of (2.118), suggests interpreting the CoRUM as a latent class model wherein each class represents a contribution to the overall covariance pattern of the CoRUM. Papola (2016) leveraged this framework, showing that a closed-form RUM with flexible and closed-form covariance patterns can be obtained by mixing two-level Nested Logit (NL) models. The resulting CoRUM model, thus termed CoNL (Combination of Nested Logit models), exhibits the following cumulative distribution function:

$$F(\varepsilon_1,...,\varepsilon_n) = \sum_{i \in \mathrm{I}} w^i \cdot F^i(\varepsilon_1,...,\varepsilon_n) = \sum_{i \in \mathrm{I}} w^i \cdot \exp - \sum_{k^i} \left(\sum_{a' \in k^i} e^{-\varepsilon_{a'}/\delta_k^i} \right)^{\delta_k^i} \tag{2.122}$$

being k^i the generic nest associated with the i-th mixing NL, a the generic alternative belonging to that nest, and δ_k^i the structural parameter associated to nest k^i. The corresponding CoNL probability statement is given by:

$$p(j) = \sum_{i \in \mathrm{I}} w^i \cdot \frac{e^{V_j/\delta_{k(j)}^i} \cdot \left(\sum_{a \in k^i} e^{V_a/\delta_{k(j)}^i} \right)^{\delta_{k(j)}^i - 1}}{\sum_{k'^i} \left(\sum_{a' \in k'^i} e^{V_{a'}/\delta_{k'}^i} \right)^{\delta_{k'}^i}} \quad \forall j \tag{2.123}$$

being V_a the systematic utility of alternative a and $k_i(j)$ the specific nest of the i-th component containing alternative j. The corresponding CoNL variances/covariances can be derived in turn from (2.119)-(2.120), yielding:

$$Cov[\varepsilon_j, \varepsilon_m] = \sum_{i \in \mathrm{I}} w^i \cdot Cov^i[\varepsilon_j, \varepsilon_m] = \pi^2/6 \cdot \sum_{i \in \mathrm{I}^{jm}} w^i \cdot [1 - (\delta_{k(j,m)}^i)^2] \quad \forall j,m \tag{2.125}$$

$$Var[\varepsilon_j] = \sum_{i \in \mathrm{I}} w^i \cdot Var^i[\varepsilon_j] = \sum_{i \in \mathrm{I}} w^i \cdot \pi^2/6 = \pi^2/6 \quad \forall j \tag{2.124}$$

wherein I^{jm} denotes the subset of all NL mixing components exhibiting a nest $k(j,m)$ including both alternatives j and m. The corresponding general expression of the CoNL correlations is:

$$\rho_{jm} = \frac{Cov[\varepsilon_j \varepsilon_m]}{(Var[\varepsilon_j])^{0.5} \cdot (Var[\varepsilon_m])^{0.5}} = \sum_{i \in \mathrm{I}^{jm}} w^i \cdot [1 - (\delta_{k(j,m)}^i)^2] \quad \forall j,m \tag{2.126}$$

Thanks to the above properties, the CoNL is a natural candidate to tackle the research question stated in the introduction, i.e. to propose an operational closed-form route choice model with flexible covariances. For this aim, it is important to highlight that expression (2.126) – providing the mathematical vehicle to link correlations with model structure/parameters – indicates that a given correlation pattern can be targeted with diverse CoNL specifications, by varying number and structure of the mixing components. Clearly, the choice of the most appropriate specification should enable an effective operationalization of the CoNL route choice model: this aspect will be addressed in detail in the next section.

5.3 A CoNL specification for route choice modelling

This section introduces the proposed CoNL specification for route choice modelling. For the sake of clarity, all concerned definitions and terms are illustrated with reference to o-d pair **1-4** in the network depicted in Figure 5.1, connected by eight routes labelled $k_1 \ldots k_8$ as reported in the bottom of Figure 5.1. Let G≡{L,N}be a graph representing a road network, being L a set of

n_l links and N a set of nodes. For any $l \in$L, let t(l) and h(l) denote respectively the tail and the head of link l. Let K_{od} be a set of acyclic routes connecting the pair of centroids o and d with o, $d\in$N. Each route $k\in K_{od}$ is associated with an ordered set of links $L_k \subseteq L$. Consistently, let $L_{od} \equiv \{\cup L_k \ \forall k \in K_{od}\}$ be the collection of all links of all routes within K_{od}. In turn, each link $l \in L_{od}$ is associated with a set $K_l \subseteq K_{od}$ of routes including l; it obviously occurs $K_{od} \equiv \{\cup K_l \ \forall l \in L_{od}\}$. The above definitions yield the following sets for the example in Figure 5.1[16]: $K_{\mathbf{1\text{-}4}} \equiv \{k_1, k_2, \ldots, k_8\}$; $L_{k_1} \equiv \{$*1-2*, *2-3*, *3-4*$\}$, $L_{k_2} \equiv \{$*1-2*, *2-3*, *3-7*, *7-8*, *8-4*$\}$, ...; $L_{1\text{-}4} \equiv \{$*1-2*, *1-5*, *5-6*, *2-6*, *6-2*, *2-3*, *6-7*, *3-7*, *7-3*, *3-4*, *7-8*, *8-4*$\}$; $K_{1\text{-}2} \equiv \{k_1, k_2, k_3, k_4\}$.

Let $c_l \ \forall l \in L$ be an additive link impedance and let $C_k = \Sigma_l \ c_l \ \forall l \in L_k$ the corresponding route impedance. In a RUM context, C_k might be just one of the attributes entering the systematic utility V_k of route k, whose perceived utility U_k is generally expressed as $U_k = V_k + \varepsilon_k$, being ε_k a random residual. Recalling the introduction, the dispersion matrix Σ_K of the ε_k's usually resorts to the specification by Daganzo and Sheffi (1977), who hypothesized the covariance between random residuals of a pair of routes to be proportional to their topological overlapping, that is:

$$\sigma_{jm} = Cov[\varepsilon_j, \varepsilon_m] = \xi \sum_{l \in L_j \cap L_m} c_l \qquad \forall j, m \neq j \in K_{od} \tag{5.1}$$

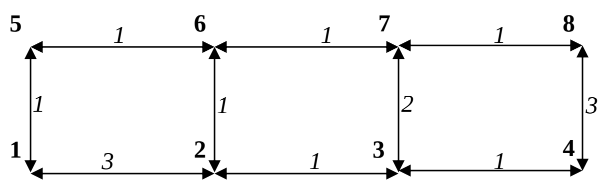
and the variance of a random residual of each route to be proportional to its impedance, that is:

$$\sigma_{mm} = Var[\varepsilon_m] = \xi \sum_{l \in L_m} c_l = \xi C_m \qquad \forall m \in K_{od} \tag{5.2}$$

which is a special case of (5.1) with j=m. In turn, the generic correlation corresponding to (5.10)-(5.2) is given by:

$$\rho_{jm} = \frac{\sigma_{jm}}{\sqrt{\sigma_{jj}\sigma_{mm}}} = \frac{1}{\sqrt{C_j C_m}} \cdot \sum_{l \in L_j \cap L_m} c_l \qquad \forall j, m \neq j \in K_{od} \tag{5.3}$$

Said that, the specification of a CoNL to target route correlations (5.3) is not unique, as recalled at the end of the previous section. In this respect, the following subsections illustrate in detail an effective CoNL specification for route choice, respectively in terms of model structure and model parameters.



[16] In the following, a specific *od* pair will be denoted in bold with centroids separated by hyphen, e.g. **1-2**, whilst a link will be denoted in italics with tail and head nodes separated by hyphen, e.g. *1-2*.

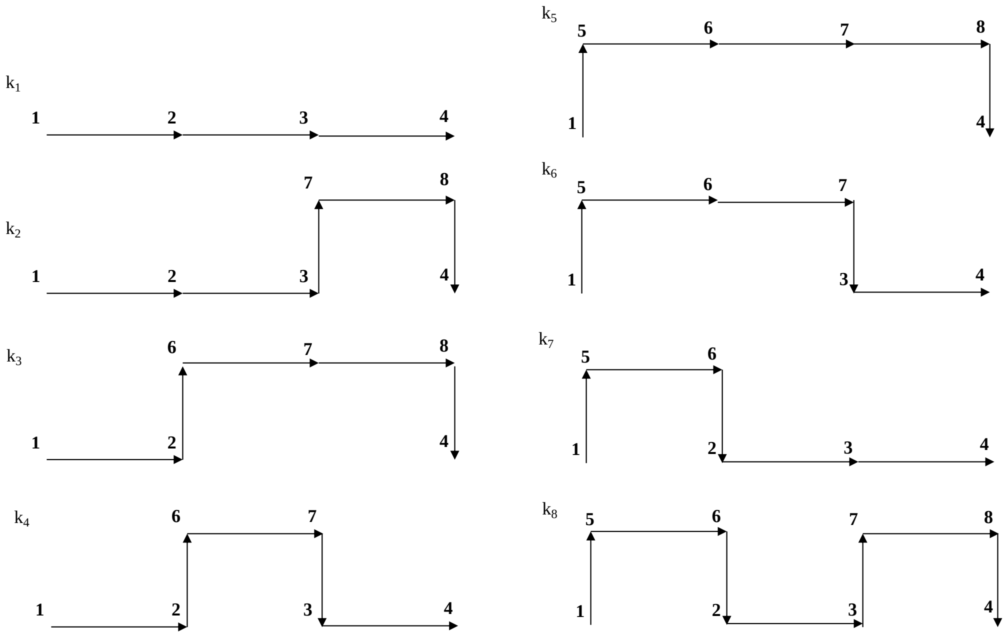


Figure 5.1: Topology of a toy network to showcase throughout the chapter the definitions of the proposed CoNL route choice model

5.3.1 CoNL route choice: model structure

A CoNL model structure consistent with the assumption by Daganzo and Sheffi (1977) is intuitively suggested by a comparison between (5.3) and (2.126). For this aim, let a link embedded in multiple routes for any given o-d pairs be termed a *shared* link, in contrast with links belonging to a single route. That said, given an o-d pair, (5.3) indicates that the specification of a CoNL route choice model should be built based on nests representing shared links of the network, each nest grouping together all routes sharing that link. Once identified the type of nests to include, the subsequent aspects to address are the definition of the number of CoNL mixing components and the consistent allocation of the nests across those components, consistent with theoretical and operational requirements.

In terms of theoretical requirements, it should be noticed first that any feasible CoNL specifications should be genuine, i.e. consistent with the properties defined by Papola (2016) and already recalled in Section 5.2. First, for each mixing component to be a NL model, the following should hold: for any pairs of routes j and m, nests representing links belonging to the set $L_j \cap L_m$ should be necessarily accommodated into different mixing components. Indeed, if this condition were not met, there would be at least a mixing component with routes j and m belonging two more than one nest, i.e. a CNL mixing component, in contrast with the definition of CoNL.

Furthermore, by definition, all alternatives (routes) should be included in each mixing component of the CoNL. To address this requirement, it can be necessary to allocate some nests, representing some of the links $l \in L_j \cap L_m$, across multiple mixing components, and to add some routes not belonging to any nests to each mixing component.

By way of example, Figure 5.2 illustrates the application of the proposed model structure to the o-d pair **1-4** in the network of Figure 5.1, whose 8 routes include 12 shared links. It is easy to recognize that 7 mixing components suffice to accommodate the above requirements. For instance, routes k_2 and k_3 yield $L_{k_2} \cap L_{k_3} \equiv \{$*1-2*, *7-8*, *8-4*$\}$, and the three nests corresponding to the shared links within $L_{k_2} \cap L_{k_3}$ – all embedding by construction both routes k_2 and k_3 – are included in different mixing components (#1, #5 and #7 respectively in Figure 5.2), to preserve the NL-genuineness of each mixing component. Moreover, all routes within $K_{1\text{-}4}$ are included in each mixing component: for this aim, nest representing link *3-4* is included in multiple mixing components (#5 and #6) and, for instance, routes k_1 and k_2 not belonging to any nests are added as singletons in mixing components #2 and #3.

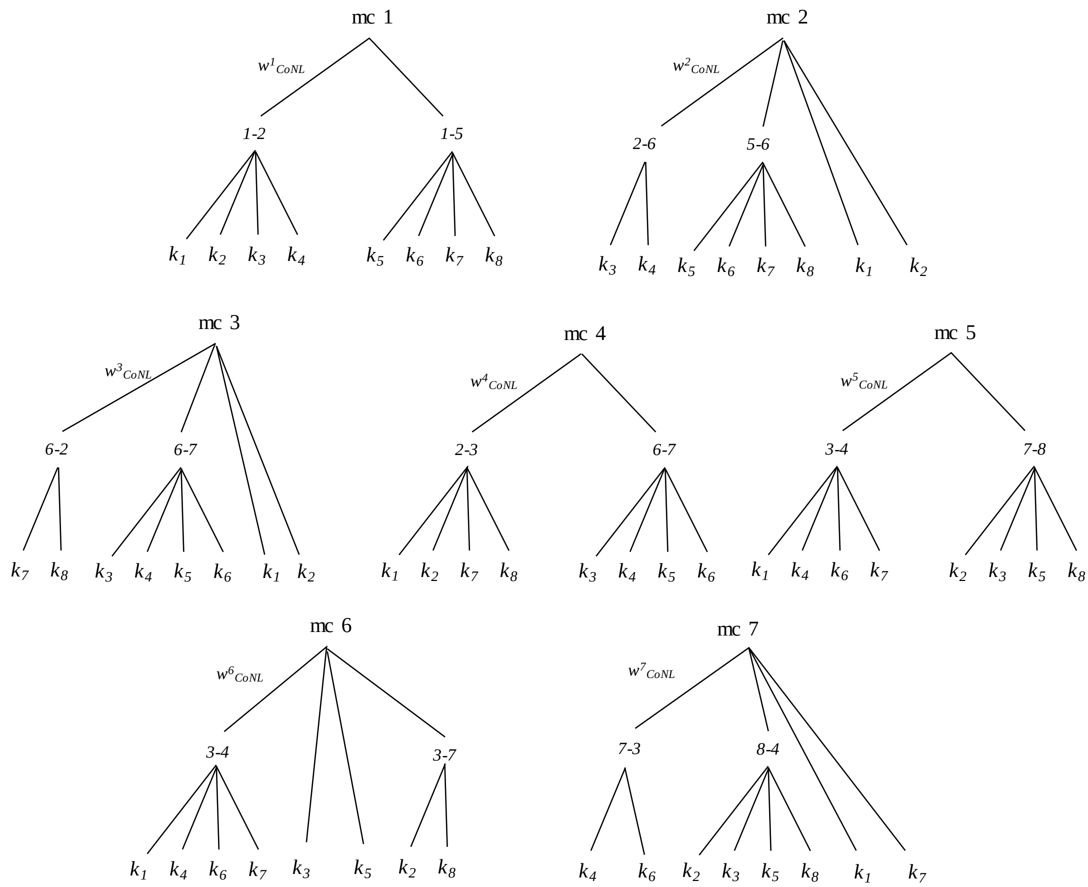


*Figure 5.2: Example of CoNL structure for the o-d pair **1-4** in the network in Figure 5.1.*

In terms of operational requirements, the application of the CoNL route choice model would imply the modeller to inspect manually the set of routes and identify explicitly all mixing components with the corresponding nesting structures to write explicitly the probability statement (2.123). This is infeasible in real-size networks; thus, a procedure should be implemented, capable to provide directly the mixing components and their nesting structure, without explicit inspection of the routes: that is, a CoNL specification allowing implicit probability statement. In addition, it is desirable to keep at minimum the number of mixing components of the CoNL. For this aim, a straightforward conceptual vehicle is to group together links of the network based on properly defined *network levels*, each corresponding to a CoNL mixing component. A recursive algorithm providing a genuine CoNL route choice specification with implicit probability statement, based on the concept of the network levels, will be illustrated in detail in Section 5.2.

Summarizing, for each o-d pair, the proposed CoNL route choice specification adopts routes as alternatives and shared links as elemental nests, grouped in mixing components representing levels of the transport network. The nest identified by a specific shared link might appear in various mixing components, to guarantee that all routes will appear in each mixing component (genuineness of the CoNL specification). In formal terms, I^l will denote the subset of mixing components which a link l belongs to, n_{MC} the number of mixing components in the CoNL route choice model, and Ω_i the subset of network links associated with the generic i-th mixing component.

Hence, being $I^{jm} \equiv \{\cup_l I^l,\ l \in L_j \cap L_m\}$, (2.126) can be rewritten as:

$$\rho_{jm} = \sum_{i \in I^{jm}} w^i \cdot [1 - (\delta^i_{k(j,m)})^2] = \sum_{l \in L_j \cap L_m} \sum_{i \in I^l} w^i \cdot (1 - {\delta^i_l}^2) \tag{5.4}$$

By way of example, routes k_1 and k_4 in Figure 5.1 yield $L_{k_1} \cap L_{k_4} \equiv \{1\text{-}2, 3\text{-}4\}$ and $I^{k_1 k_4} \equiv \{I^{12} + I^{34}\} \equiv \{1\} + \{5,6\} \equiv \{1,5,6\}$, so (5.4) becomes:

$$\rho_{k_1 k_4} = w^1 \cdot \left(1 - {\delta^1_{12}}^2\right) + w^5 \cdot \left(1 - {\delta^5_{34}}^2\right) + w^7 \cdot \left(1 - {\delta^6_{34}}^2\right) \tag{5.5}$$

Comparing (5.4) with (5.4), the following sufficient condition occurs on the generic link l shared between routes j and m for the CoNL route choice model to target Daganzo and Sheffi (1977) covariances:

$$\sum_{i \in I^l} w^i \cdot (1 - {\delta^{i}_l}^2) = \frac{c_l}{\sqrt{C_j C_m}} \qquad \forall l \in L_j \cap L_m \quad \forall j, m \neq j \in K_{od} \tag{5.6}$$

From a practical standpoint, (5.6) indicates that the contribution of the topological overlapping of link l on the correlation between routes j and m is split in the CoNL route choice model across the mixing components identified by the subset I^l. In turn, setting the parameters within (5.6) is the final key task to specify the CoNL route choice model, as illustrated in the next sub-section.

5.3.2 CoNL route choice: model parameters

In terms of model parameters, the first important consideration is that the nesting parameters within each mixing component should not be smaller than a lower bound δ_{min} close to zero, i.e. leading to mixing components represented by NL models with deterministic choice within nests. In fact, the CoNL model belongs to the family of random utility models, including also the Cross Nested Logit model, that provides the overall correlation between a pair of alternatives as the superposition of the partial correlations given by the nests the pair of alternatives belongs to. Usually, the contribution of each single nest to the overall correlation depends upon the nesting parameter of that nest and a weight component (e.g. the weight of the mixing component in the CoNL, or the membership degree in a CNL) of that nest within the overall model structure. In this respect, there is evidence in the literature that choosing nesting parameters close to zero deteriorates the performance of such models, as highlighted in the route choice context amongst other by Prashker and Bekhor (2004) and Marzano (2005). Furthermore, for the sake of the operationalization of the CoNL route choice model, it is intuitive to impose the same nesting parameter to the possibly multiple nests representing link l across mixing components, that is:

$$\delta^i_l = \delta_l \geq \delta_{\min} \qquad \forall i \in I^l, \forall l \in L_j \cap L_m \quad \forall j, m \neq j \in K_{od} \tag{5.7}$$

Consistently, (5.6) becomes:

$$(1 - {\delta_l}^2) \cdot \sum_{i \in I^l} w^i = \frac{c_l}{\sqrt{C_j C_m}} \qquad \forall l \in L_j \cap L_m \quad \forall j, m \neq j \in K_{od} \tag{5.8}$$

Again, with reference to the CoNL model presented in Figure 5.2, (5.5) and (5.8) become respectively:

$$\rho_{k_1k_4} = w^1 \cdot \left(1-{\delta_{12}}^2\right) + (w^5+w^7)\cdot\left(1-{\delta_{34}}^2\right)(w^5+w^7) \cdot \left(1-{\delta_{34}}^2\right) = \frac{c_{34}}{\sqrt{C_{k_1}C_{k_4}}} \tag{5.9}$$

Notably, the collection of expressions (5.8) leads to a system of n_{SL} equations, as many as the number of links shared by all pair of routes within K_{od}, that is $n_{SL}=\Sigma_{m,j\in K_{od}}|L_j \cap L_m|$. It is easy to recognize that such system has $n_{SL}+n_{MC}$ unknowns. Indeed, each shared link identifies a nest by definition and (5.7) ensures the total number of δ's to equal the total number of shared links. In addition, there are as many unknown weights as the number n_{MC} of mixing components of the CoNL. In the example of Figure 5.1 and Figure 5.2, all 12 links belonging to $L_{1\text{-}4}$ are shared links, thus the collection of expressions (5.8) yields a system of 12 equations in 12+7=19 unknowns.

This unbalance suggests fixing exogenously a value for each weight w_i $\forall i\in 1\ldots n_{MC}$ to balance unknowns and equations, and then calculating the corresponding values of the nesting parameters by solving the aforementioned system.

The value of each wi ∀i∈1… nMC can be determined by recalling that the weights of the mixing components enter with direct proportionality in equation (2.126), that defines the correlation between any pairs of alternatives. In addition, in accordance with the CoRUM definition in Section 5.2, weights are constrained to sum up to 1, thus they should be allocated appropriately across mixing components, consistent with the correlations each mixing component should reproduce. Since in the CoNL route choice model each mixing component i should reproduce multiple correlation contributions of type cil/Codmin, one for each shared link l allocated to that mixing component, each mixing component i can be considered in charge of targeting a correlation contribution equal to cilmean/Codmin, being cil,mean the average cost of the shared links belonging to that mixing component. As a consequence, the total unitary budget can be straightforwardly allocated to the different mixing components as follows:

$$w^i = \frac{c^i_{l,mean}}{\sum_{i'\in \mathrm{I}} c^{i'}_{l,mean}} \qquad \forall i \tag{5.10}$$

With reference to the nesting parameters, expressions (5.3) should be first rewritten by considering that the CoNL is a homoscedastic model, thus route variances can be conveniently levelled on the cost of the shortest route, yielding:

$$\sigma_{mm} = Var[\varepsilon_m] = \xi \sum_{l\in \mathrm{L}_m} c_l = \xi C_{od,\min} \qquad \forall m \in \mathrm{K}_{\mathrm{od}} \tag{5.11}$$

Thus, (5.8) becomes:

$$(1-\delta_l^2)\cdot\sum_{i\in \mathrm{I}^l} w^i = \frac{c_l}{C_{od,\min}} \qquad \forall l \in \mathrm{L}_j \cap \mathrm{L}_m \;\; \forall j, m \neq j \in \mathrm{K}_{\mathrm{od}} \tag{5.12}$$

and, recalling (5.11) and (5.7), it occurs:

$$\delta_l = \max\left\{\delta_{\min};\ \sqrt{1-\frac{c_l}{C_{od,\min}\cdot\sum_{i\in \mathrm{I}^l} w^i}}\right\} = \max\left\{\delta_{\min};\ \sqrt{1-\frac{c_l}{C_{od,\min}}\cdot\frac{\sum_{i\in \mathrm{I}} c^i_{l,mean}}{\sum_{i\in \mathrm{I}^l} c^i_{l,mean}}}\right\} \tag{5.13}$$

Clearly, (5.14) holds if the radicand is nonnegative, thus, recalling (5.7), expression (5.14) should be rewritten as follows:

$$\delta_l = \begin{cases} \max\left\{\delta_{\min}; \sqrt{1-\dfrac{c_l}{C_{od,\min}}\cdot\dfrac{\sum_{i\in \mathrm{I}} c^i_{l,mean}}{\sum_{i\in \mathrm{I}^l} c^i_{l,mean}}}\right\} & \text{if } \; 1-\dfrac{c_l}{C_{od,\min}}\cdot\dfrac{\sum_{i\in \mathrm{I}} c^i_{l,mean}}{\sum_{i\in \mathrm{I}^l} c^i_{l,mean}} > 0 \\ \delta_{\min} & \text{otherwise} \end{cases} \tag{5.14}$$

Overall, expressions (5.4), (5.11) and (5.15) allow particularizing the CoNL model to route choice contexts. The key aspect to address is how to operationalize this model, in particular through the proposition of a recursive algorithm providing a genuine CoNL route choice specification with implicit probability statement, as Section 5.4 illustrates in depth.

5.4 The CoNL route choice model with implicit probability statement

The key step for the operationalization of the CoNL route choice model proposed in Section 5.3 is the implementation of an algorithm providing a genuine CoNL route choice specification with implicit probability statement, based on the concept of the network levels. Such algorithm is based on two main steps: the former is the specification of the mixing components; the latter is the calculation of model parameters. Each step is illustrated separately in the next two sub-sections. Clearly, once known the mixing components (i.e. the model structure) and its parameters, calculating choice probabilities is straightforward via equation (2.123).

5.4.1 Specification of mixing components and model structure

The mixing components of the CoNL route choice model should be specified separately for each o-d pair and, as it will be clarified soon, attention should be restricted only to efficient routes of the network with respect to the origin[17] in the sense defined by Dial (1971), collected into a subset $\mathrm{K}_{od}^{eff\_o}$. Clearly, being a model with explicit route enumeration, a pre-processing providing such set of routes for each o-d pair should be performed. Following the notation introduced in Section 5.3, let $\mathrm{L}_{od}^{eff\_o}$ be the collection of all links of all routes within $\mathrm{K}_{od}^{eff\_o}$, and $\mathrm{K}_l^{eff\_o}$ the set of routes including link $l\in\mathrm{L}_{od}^{eff\_o}$. Furthermore, $h(l)$ and $t(l)$ denote respectively the head and the tail nodes of link l, and EFS(n) and EBS(n) denote respectively the efficient forward star and backward star of a node n. Specifically, the forward star of a node n is the collection of links exiting from n, i.e. whose tail is represented by n; conversely, the backward star of a node n is the collection of links entering n, i.e. whose head is represented by n.

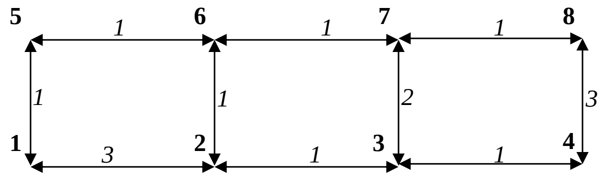
That said, the generic mixing component i should include a set of efficient links Ω_i, interpretable as a level of the network for that o-d pair, consistent with the requirements for a genuine CoNL specification, stated in Section 5.3.1. This means satisfying the following conditions:

C1. $\{\cup_l \mathrm{K}_l^{eff\_o} \; \forall l\in\Omega_i\} \equiv \mathrm{K}_{od}^{eff\_o} \quad \forall i\in\mathrm{I}$, that is each mixing component should include all efficient routes connecting that o-d pair as elemental alternatives;

[17] The algorithm would clearly work also with the Dial efficiency with respect to both origin and destination, being this a more restrictive assumption that further limits the number of considered route. In general, the Dial efficiency assumption introduces a restriction in the choice set; however, as recalled by many papers in the literature, the Dial efficiency is a well-established assumption, generally not excluding reasonable paths and hence not significantly affecting route choice probabilities.

C2. $K_{l_1}^{eff\_o} \cap K_{l_2}^{eff\_o} \equiv \varnothing \ \forall \ l_1, l_2 \neq l_1 \in \Omega_i$, that is each efficient route should belong to only one nest in each mixing component, for the latter to be a Nested Logit.

By way of example, Figure 5.3 illustrates the above sets with reference to the CoNL model structure depicted in Figure 5.2 and related to the network in Figure 5.1 for the o-d pair **1-4**, that is: $K_{1\text{-}4}^{\ eff\_1} \equiv \{k^{eff}_1, k^{eff}_2, k^{eff}_3, k^{eff}_4\}$ with $L^{eff}_{k_1} \equiv \{1\text{-}2, 2\text{-}3, 3\text{-}4\}$, $L^{eff}_{k_2} \equiv \{1\text{-}5, 5\text{-}6, 6\text{-}2, 2\text{-}3, 3\text{-}4\}$, $L^{eff}_{k_3} \equiv \{1\text{-}5, 5\text{-}6, 6\text{-}7, 7\text{-}3, 3\text{-}4\}$, $L^{eff}_{k_4} \equiv \{1\text{-}5, 5\text{-}6, 6\text{-}7, 7\text{-}8, 8\text{-}4\}$, and $L_{14}^{\ eff\_1} \equiv \{1\text{-}2, 1\text{-}5, 5\text{-}6, 6\text{-}2, 2\text{-}3, 6\text{-}7, 7\text{-}3, 3\text{-}4, 7\text{-}8, 8\text{-}4\}$.



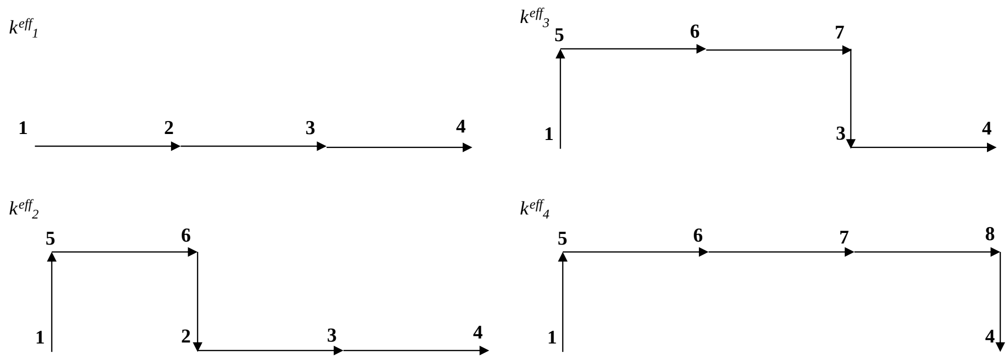


*Figure 5.3: Efficient choice sets with respect to the o-d pair **1-4** in the example of Figure 5.1.*

The following sub-sections illustrate a recursive algorithm for the specification of the mixing components and the model structure. Specifically, Section 5.4.1.1 proposes a double-step algorithm (i.e. requiring a forward and a backward network exploration for each o-d pair) and then Section 5.4.1.2 a simplified single-step algorithm.

5.4.1.1 Double-step algorithm

For each o-d pair, a collection of subsets Ω_i (i.e. of mixing components) satisfying these conditions can be built recursively, through a forward exploration of the network, starting from the origin *o*. The *i*-th iteration of this recursive algorithm, allowing calculation of Ω_i given availability of the set Ω_{i-1} identified in the previous iteration, includes the following three steps:

1. generating a set Γ_i of candidate links for inclusion into Ω_i, given by all efficient links within $L_{od}^{eff\_o}$ belonging to the efficient forward stars of the head of links within Ω_{i-1}, that is $\Gamma_i \equiv \{\cup_l \text{EFS}(\text{h}(l)) \ \forall l \in \Omega_{i-1}\} \cap L_{od}^{eff\_o}$. It is easy to verify that Γ_i satisfies C1.

   Proof. By construction, Γ_i includes all links belonging to all forward stars of links belonging to Ω_{i-1}. Thus Γ_i should satisfy C1, otherwise there would be at least a route not including any links belonging to the forward stars of links in Ω_{i-1}, in contrast with the fact that Ω_{i-1}, satisfies C1. ■

   On the contrary, Γ_i does not necessarily satisfy C2, because $\cup_l \text{EFS}(\text{h}(l)) \ \forall l \in \Omega_{i-1}$ might include multiple links of the same route.

2. meeting condition C2 by generating a subset $\Lambda_i \subseteq \Gamma_i$ obtained by elimination of "descendant links" from Γ_i. By definition, given two links l_1 and l_2, l_2 is a descendant of l_1 if l_1 and l_2 are connected in the efficient subnetwork, that is the cost of the shortest path from the head of link l_1 towards the tail of link l_2 is less than infinite. Equivalently, if l_2 is a descendant of l_1 then l_1 is an ascendant of l_2. In principle, a circular dependency might occur, i.e. having at the same time l_1 depending upon l_2 and vice versa. It is easy to prove that restricting the attention only to Dial efficient links with respect to the origin o is a sufficient condition to overcome this problem.

   Proof. Given two links a and b with $a,b \in L_{od}^{eff\_o}$, let b be a descendant of a, that is by definition:

   $$c_{min}{}^{eff\_o}{}_{h(a)t(b)} < \infty \rightarrow c_{min}{}^{eff\_o}{}_{oh(a)} < c_{min}{}^{eff\_o}{}_{ot(b)} \tag{5.15}$$

   being $c_{min}{}^{eff\_o}{}_{h(a)t(b)}$ the cost of the shortest route $k_{h(a)t(b)}$ between h(a) and t(b) on the network including only links within $L_{od}^{eff\_o}$. Equation (5.25) also implies:

   $$c_{min}{}^{eff\_o}{}_{ot(a)} < c_{min}{}^{eff\_o}{}_{oh(a)} < c_{min}{}^{eff\_o}{}_{ot(b)} < c_{min}{}^{eff\_o}{}_{oh(b)}. \tag{5.16}$$

   As a result, h(b) is farther from origin than t(a) in a Dial sense, thus no route can exist on $L_{od}^{eff\_o}$ connecting h(b) with t(a), i.e. if b is a descendant of a, a cannot be a descendant of b. ■

   It is easy to prove that Λ_i meets condition C2.

   Proof. The absence of descendant links within Λ_i ensures that for any links $l_1 \in \Lambda_i$ the corresponding collection of routes $K_{l_1}^{\mathbf{eff\_o}}$ will be disjoint from any other collections $K_{l_2}^{\mathbf{eff\_o}}$ $\forall\ l_2 \neq l_1 \in \Lambda_i$: if there were a common route between $K_{l_1}^{\mathbf{eff\_o}}$ and $K_{l_2}^{\mathbf{eff\_o}}$ there would be a descendant/ascendant relationship between l_1 and l_2, in contrast with the definition of Λ_i. ■

   Unfortunately, Λ_i lost compliance with condition C1, because of the removal of some links from Γ_i with their concerned routes.

3. restoring condition C1 on Λ_i, by augmenting Λ_i with all links belonging to both Ω_{i-1} and to all efficient backward stars EBS(t(l)) of the eliminated descendant links l at step #2. Formally, this means augmenting Λ_i with the set $A_i \equiv \{\cup\ \text{EBS}(t(l))\ \forall\ l \in E_i\} \cap \Omega_{i-1}$ being $E_i \equiv \Gamma_i - \Lambda_i$ the set of eliminated links at step #2. It is easy to prove that the resulting set $\Lambda_i \cup A_i$ meets condition C1.

   Proof. The elimination of a link l from Γ_i at step #2 implies actually the elimination of all links belonging to EFS(t(l)): indeed, if l is descendant of another link l', also all other efficient links exiting t(l) will be descendant of l' by definition. As a result, the elimination of l at step #2 has implied the elimination of all routes including any of the links belonging to EFS(t(l)). Since such routes are also all those including any links of EBS(t(l)), adding all such links will contribute to restore the full set of paths K_{od}. ■

   Also, it is easy to prove that $\Lambda_i \cup A_i$ meets also condition C2.

Proof. There is no relationship of descendance either between links within A_i, because by construction A_i is a subset of $\boldsymbol{\Omega}_{i-1}$ that satisfies C2 by definition, or between links within Λ_i, as a direct result of step #2. Thus, it should be proved that there is no relationship of dependence between members of A_i and Λ_i. On the one hand, links within $A_i \subseteq \boldsymbol{\Omega}_{i-1}$ cannot be descendant of links within $\Lambda_i \subseteq \Gamma_i$ by construction (step #1). On the other hand, links within Λ_i cannot be descendants of links within A_i. Indeed, if there were a link $l \in \Lambda_i$ depending upon a link $l' \in A_i$ by construction l should either belong to the forward star of l' or be even farther in Dial sense than a link of the forward star of l'. In both cases, l would thus belong to the set of eliminated links at step #2, i.e. $l \in E_i$, and this is absurd because l belongs by hypothesis also to Λ_i, being in fact $\Lambda_i \cap E_i \equiv \varnothing$. ■

Overall, this yield $\boldsymbol{\Omega}_i \equiv \Lambda_i \cup A_i$.

Obviously, the algorithm repeats recursively these steps, starting from $\boldsymbol{\Omega}_1 \equiv \{l \in EFS(o)\}$, and until $\Gamma_i \equiv \varnothing$. At a glance, it can be summarised as follows:

for each *od*
  $n_{MC}=1$
  $\boldsymbol{\Omega}_1 \equiv \{l \in EFS(o)\}$
  $\Gamma_2 \equiv \{\cup_l EFS(h(l))\ \forall l \in \boldsymbol{\Omega}_1\} \cap L_{od}^{eff\_o}$
  $i=2$
  do
    $\Lambda_i \equiv \Gamma_i$
    for each l_1, $l_2 \in \Gamma_i$
      if $cmin^{eff\_o}_{h(l1)t(l2)} < \infty$,
        $\Lambda_i \equiv \Lambda_i - \{l_2\}$ (step #2)
        $\Lambda_i \equiv \Lambda_i + EBS\{t(l_2)\} \cap \boldsymbol{\Omega}_{i-1}$ (step #3)
      end if
    next
    $\boldsymbol{\Omega}_i \equiv \Lambda_i$
    $n_{MC} = n_{MC} + 1$
    $i=i+1$
    $\Gamma_i = \{\cup_l EFS(h(l))\ \forall l \in \boldsymbol{\Omega}_{i-1}\} \cap L_{od}^{eff\_o}$ (step #1)
  until $\Gamma_i \equiv \varnothing$
next

The final output of this algorithm is the number of mixing components n_{MC} and the corresponding sets of links $\boldsymbol{\Omega}_i$ $i \in 1 \ldots n_{MC}$ to include in the i-th mixing component of the CoNL route choice specification. It is easy to recognize that the computational complexity of this algorithm is proportional to the number of o-d pairs. Notably, as special cases, it might occur $|K_l|=1\ \forall l \in \boldsymbol{\Omega}_i$, yielding mixing components characterized by a Multinomial Logit structure and, if the above occurs for all mixing components, the CoNL degenerates into a Multinomial Logit model. These special cases require *ad hoc* treatment, as illustrated in Section 5.4.2.

By way of example, the generic iteration of the proposed algorithm is illustrated in Table 5.1 for the choice set reported in Figure 5.3. Notably, iteration #1 starts from EFS(1), represented by $\Omega_1 \equiv \{1\text{-}2,\ 1\text{-}5\}$, whilst iteration #2 initializes the set $\Gamma_2 \equiv \{2\text{-}3,\ 5\text{-}6\}$. Notably, Γ_2 clearly satisfies the condition C1 but does not match condition C2, because the efficient route $k^{eff}{}_2$ (*156234*) should belong to both nests representing links *5-6* and *2-3*. This implies applying step #2 of the algorithm illustrated in Section 5.4.1.1, that is deleting link *2-3* because of its descendance from link *5-6*. Then, it is easy to recognize that the resulting set $\Lambda_2 \equiv \{5\text{-}6\}$ does not satisfy condition C1, thus augmenting Λ_2 with the set $\Omega_1 \cap \text{EBS}\{2\text{-}3\} \equiv \{1\text{-}2\}$ in accordance with step #3 in Section 5.4.1.1 yields the final set $\Omega_2 \equiv \{1\text{-}2, 5\text{-}6\}$, satisfying both conditions C1 and C2. The resulting CoNL specification for the o-d pair **1-4** is represented in Figure 5.4.

Iteration	Γ_i	Λ_i	$\Omega_{i-1} \cap$EBS(descendant links)	Ω_i
1	-	-	-	*1-2, 1-5*
2	*2-3, 5-6*	*5-6*	*1-2*	*1-2, 5-6*
3	*2-3, 6-2, 6-7*	*6-2, 6-7*	*1-2*	*1-2, 6-2, 6-7*
4	*2-3, 7-3, 7-8*	-	-	*2-3, 7-3, 7-8*
5	*3-4, 8-4*	-	-	*3-4, 8-4*
6	-	end		

Table 5.1:– Illustration of the iterations of the algorithm described in Section 5.1.1 for the o-d pair ***1-4*** *and the efficient routes depicted in Figure 5.3.*

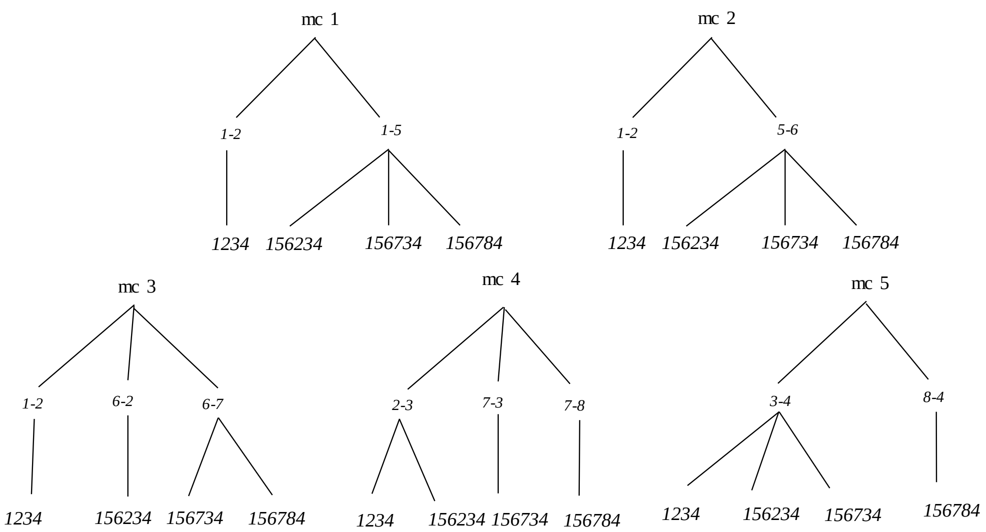


Figure 5.4: CoNL specification provided by the double-step algorithm for the o-d pair ***1-4*** *in the network in Figure 5.1.*

5.4.1.2 Single-step algorithm

From a practical standpoint, once determined the mixing components for each o-d pair, the model structure underlying the probability statement (2.123) should be built by exploring, for each mixing component *i*, the corresponding set of links Ω_i. Specifically, the nest in the CoNL route choice model structure corresponding to a link $l \in \Omega_i$ should be built by identifying all routes $k \in K_{od}{}^{eff\_o}$ such that $l \in k$. This suggests a slightly different version of the algorithm described so far, potentially leading to a computational time saving in some circumstances.

The conceptual vehicle for this simplification is a modification of the condition yielding Γ_i at step #1, consisting of removing the need to restrict attention only to links belonging to $L_{od}^{eff\_o}$, i.e. imposing the less restrictive condition $\Gamma_i \equiv \{\cup_l \text{ EFS}(h(l)) \ \forall l \in \Omega_{i-1}\}$ instead of $\Gamma_i \equiv \{\cup_l \text{ EFS}(h(l)) \ \forall l \in \Omega_{i-1}\} \cap L_{od}^{eff\_o}$. The practical consequence is that the algorithm keeps running until there are no further efficient links in the network to explore, i.e. given an origin o, all destinations d are processed at a glance, leading to a single set of levels Ω^*_i equal for all destinations. Intuitively, given two destinations d_1 and d_2 such that d_2 is farther than d_1 with respect to o in Dial sense, the set of mixing components associated with d_2 will include that associated with d_1. Thus, when considering, by way of example, the destination d_1, a generic Ω^*_i will likely include links not belonging to any of the efficient routes $\mathrm{K}_{\mathbf{od_1}}^{\mathbf{eff\_o}}$. However, the nests corresponding to those links can be easily discarded from the CoNL route choice model structure related to the pair o-d_1 when writing the probability statement, i.e. removing nests related to links $l \in \{\Omega^*_i - L_{od}^{eff\_o}\}$.

Clearly, this leads to a trade off in the computational time of the two versions of the algorithm. The double-step version of the algorithm illustrated in Section 5.4.1.1 runs as many times as the number of o-d pairs to get the CoNL mixing components by o-d pair, but is parsimonious in assigning routes to nests because, at o-d pair level, the number of mixing components will be usually low, thus reducing the computational time needed to check the membership of routes to nests. The simplified version is faster in determining network levels, because they are processed together for all destinations given a single origin but is much slower in checking the membership of routes to nests, because of the remarkably higher number of nests to process. As a result, the modeller will have to choose between the two based on the specific topological structure of the network under analysis. A practical example of both versions of the algorithm will be presented in Section 5.5 on a test network.

5.4.2 Calculation of model parameters

In terms of CoNL route choice model parameters, the identification of the weights of the CoNL model via expression (5.11) requires prior identification of $c^i_{l,mean}$ for each mixing component $i \in \mathrm{I}$. This is easy task, once identified the set of links Ω_i associated with each mixing component i, thanks to the algorithm illustrated in Section 5.4.1. In turn, it is easy to iterate this calculation across mixing components to get the denominator of (5.11), and then the corresponding weight for each mixing component.

Importantly, the average cost $c^i_{l,mean}$ in expression (5.11) is calculated by definition over the shared links within Ω_i. As a result, if the algorithm proposed in Section 5.4.1 leads to network levels Ω_i not including any shared links ($|K_l|=1 \ \forall l \in \Omega_i$), i.e. to MNL mixing components, such mixing components would be automatically assigned a null weight and thus discarded from the overall CoNL route choice structure. Consistently, a null denominator of expression (5.11) implies all mixing components of the CoNL route choice model to be in fact MNL components, leading to an overall MNL route choice model. Thus, checking for $\sum_{i' \in \mathrm{I}} c^{i'}_{l,mean} = 0$ allows identifying when MNL route choice probabilities should be calculated on the set of enumerated paths, otherwise CoNL route choice probabilities can be calculated via expression (2.123).

5.5 Experimental analysis

This section illustrates the performance of the proposed CoNL route choice model. Following a consolidated approach in the literature (Daganzo and Sheffi, 1977; Cascetta et al., 1996; Vovsha and Bekhor, 1998; Prashker and Bekhor, 1998; Papola and Marzano, 2013), the model is tested first on toy networks, wherein CoNL route choice probabilities can be contrasted with expected probabilities, and then on more realistic networks.
For the sake of completeness, the CoNL route choice model is contrasted on such networks with the Multinomial Logit Model (Luce, 1959), the C-Logit model (specified as in Cascetta et al., 1996, equation 3.3 with $\beta_0=\gamma=1$), the Path-Size Logit (specified as in Ben-Akiva and Bierlaire, 1999), the LNL with two different specifications of the model parameters (Vovsha and Bekhor, 1998; Prashker and Bekhor, 1998) and the Paired Combinatorial Logit (Chu, 1998; Prashker and Bekhor, 1998; Gliebe et al., 1999). All models are specified with a variance level consistent with a coefficient of variation cv=0.1 and/or cv=0.2, depending upon the network. Furthermore, the CoNL model is applied with different lower bounds for the nesting parameters in equation (5.24), i.e. different $\delta_{\min}$ values.
Importantly, in accordance with Section 5.3, the CoNL route choice model aims at targeting Daganzo and Sheffi (1977) correlations: in this respect, the Probit model (Daganzo and Sheffi, 1977) allows introducing explicitly such correlation pattern, and thus it is natural to adopt it as a reference for the performances of the CoNL and of the other tested route choice models. Consistently, for each tested model, the sum of square errors (SSE) from Probit route choice probabilities is reported as aggregate measure of distance.

5.5.1 Four links-three routes network

The first test is on the well-known four-links network proposed by Daganzo and Sheffi (1977), depicted in Figure 5.5, Link impedances are such that the impedance of the bypass route $1\text{-}2_a\text{-}3$ is $c+k$ whilst the two other routes have the same impedance c. The magnitude of the correlation between routes $1\text{-}2_a\text{-}3$ and $1\text{-}2_b\text{-}3$ can be modified by changing the parameters h and k.

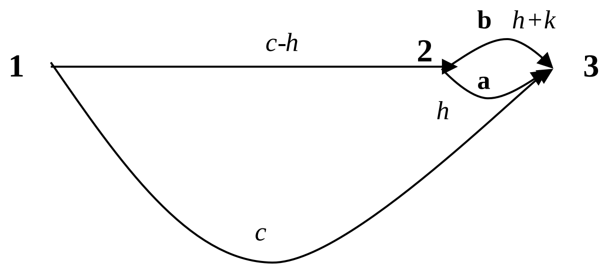
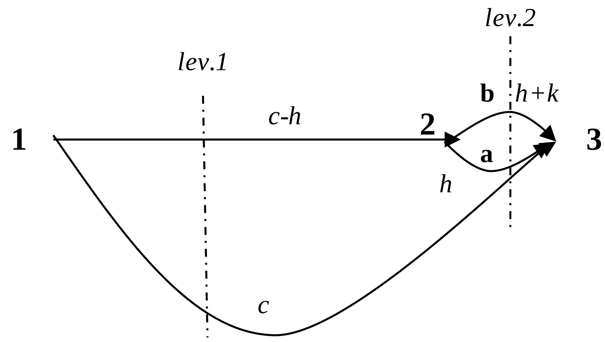


Figure 5.5: Daganzo and Sheffi (1977) test network: topology and link cost structure.

Restricting attention to origin node 1, the recursive algorithm described in Section 5.4.1 identifies a single network level towards destination 2 and two network levels towards destination 3, depicted in the top of Figure 5.6, together with the corresponding structure of the mixing components in the bottom of Figure 5.6. For the o-d pair **1-2**, there are no shared links, thus the algorithm reported in Section 5.4.1 yields an MNL model (with a single alternative represented by the sole path-link *12*) as special case. In the case of destination 3, the second level is an MNL level, thus it occurs $w^2_{\mathrm{CoNL}}=0$ consistent with Section 5.4.2. In turn, this implies $w^1_{\mathrm{CoNL}}=1$ and the CoNL in this case collapses to a Nested Logit model.



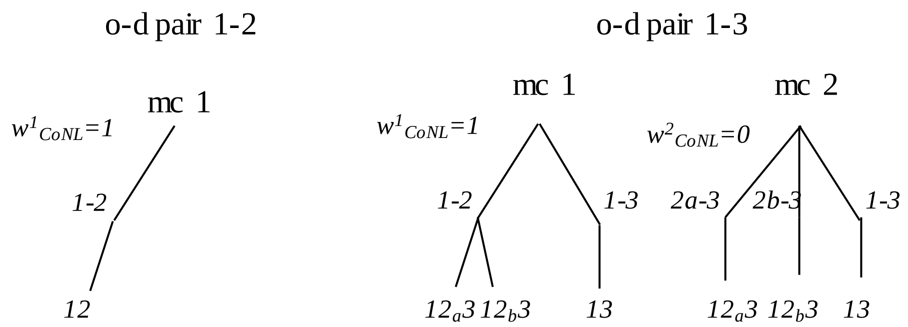


*Figure 5.6: Daganzo and Sheffi (1977) network: CoNL network levels for o-d pair **1-3** (top) and CoNL structure for destinations 2 and 3, given origin 1 (bottom).*

For the sake of completeness, the single-step illustrated in Section 5.4.1.2 has been applied as well, leading to the CoNL mixing components depicted in Figure 5.7. In this case, the structure of the network levels, and hence of the mixing components, is equal for both destinations *2* and *3*, and the allocation of routes across nests allows obtaining the same structure reported in Figure 5.6.

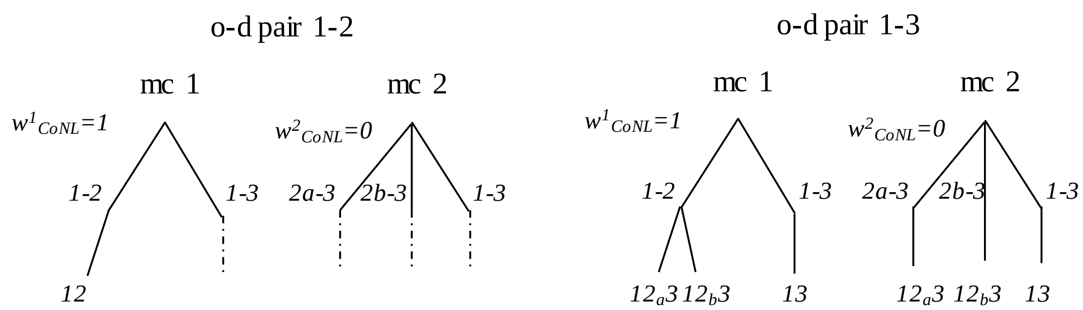


Figure 5.7: Daganzo and Sheffi (1977) network: network levels and mixing components resulting from the application of the single-step illustrated in Section 5.4.1.2.

Route choice probabilities for this network are reported in the following Table 5.2, by assuming $c \to 10$, $h \to 0$ and $k=1$.
Results in Table 5.2 evidence the incapability of modified Logit models to target expected route choice probabilities in the toy network of Figure 5.6, while all the other models perform quite well.

Efficient paths K^{EFF}_{1-3}	C_k	Probabilities (cv=0.1)										
		MNL	C-Logit	PS-Logit	PCL	LNL (δ=0)	LNL (δ ar. mean)	Probit	CoNL (δ_{min}= 0.1)	CoNL (δ_{min}= 0.2)	CoNL (δ_{min}= 0.3)	CoNL (δ_{min}= 0.4)
12ₐ3	10	0,439	0,310	0,303	0,450	0,494	0,494	0,496	0,500	0,499	0,494	0,484
12_b3	11	0,122	0,086	0,092	0,082	0,013	0,013	0,002	0,000	0,001	0,007	0,020
13	10	0,439	0,605	0,605	0,469	0,494	0,494	0,502	0,500	0,500	0,499	0,496
SSE Probit (*10³)=		**21,54**	**52,06**	**55,75**	**9,58**	**0,19**	**0,19**	**0,00**	**0,03**	**0,02**	**0,04**	**0,48**

Efficient paths K^{EFF}_{1-3}	C_k	Probabilities (cv=0.2)										
		MNL	C-Logit	PS-Logit	PCL	LNL (δ=0)	LNL (δ ar. mean)	Probit	CoNL (δ_{min}= 0.1)	CoNL (δ_{min}= 0.2)	CoNL (δ_{min}= 0.3)	CoNL (δ_{min}= 0.4)
12ₐ3	10	0,396	0,287	0,280	0,471	0,488	0,488	0,468	0,499	0,482	0,455	0,432
12_b3	11	0,208	0,151	0,161	0,087	0,024	0,024	0,034	0,001	0,020	0,054	0,087
13	10	0,396	0,561	0,559	0,443	0,488	0,488	0,498	0,500	0,498	0,492	0,482
SSE Probit (*10³)=		**46,09**	**50,21**	**55,03**	**5,92**	**0,63**	**0,63**	**0,00**	**2,10**	**0,43**	**0,60**	**4,37**

Table 5.2: Daganzo and Sheffi (1977) network: route choice probabilities for the o-d pair **1-2**, under the link cost configuration c→*10,* h→*0 and* k=*1*

5.5.2 Braess' network

The CoNL route choice model has been applied also to the well-known Braess' network, assuming the link costs reported in Figure 5.8 and hypothesizing a single origin (node *1*) and three destinations (nodes *2* to *4*). The corresponding structure of the mixing components for each o-d pair is reported in Figure 5.9. Interestingly, all key situations appear in this network: for o-d pairs **1-2** and **1-3** there are no shared links, thus in both cases the algorithm reported in Section 5.4.1 yields an MNL model as special case.

Instead, for o-d pair **1-4** there is just an MNL mixing component, discarded through a null weight in the overall CoNL structure.

Table 5.3 reports route choice probabilities on the o-d pair **1-4** for various models and for different values of the *a*/*b* ratio, thus assuming different correlations amongst the three paths.

Table 5.3 shows that all contrasted models, apart from MNL and PCL, perform quite well. The CoNL yields the same results for any $\delta_{min}\leq 0.33$ because, being the true value of both $\delta_{1\text{-}2}$ and $\delta_{3\text{-}4}$ equal to 0.33, the lower bound δ_{min} is attained only if $\delta_{min}>0.33$, e.g. with δ_{min}=0.4.

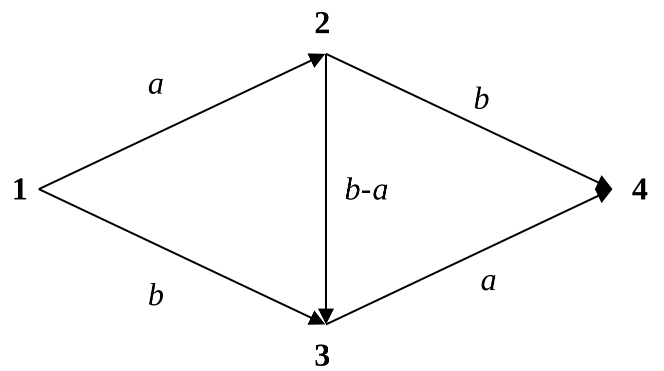


Figure 5.8: Braess' network: link costs

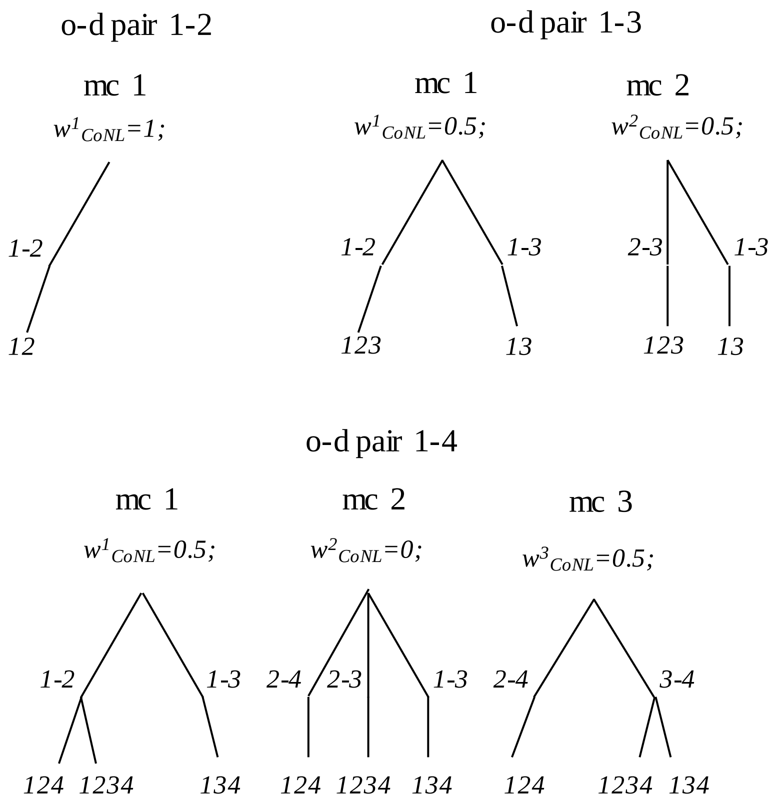


Figure 5.9: Braess' network: CoNL structure for o=1 and d=2,3,4 respectively.

Efficient paths K^{EFF}_{1-4}	C_k	Probabilities (cv=0.1) a/b=0.8 MNL	C-Logit	PS-Logit	PCL	LNL (δ=0)	LNL (δ ar. mean)	Probit	CoNL (δ_{min}= 0.1, 0.2, 0.3)	CoNL (δ_{min}= 0.4)
124	a+b	0,333	0,362	0,368	0,341	0,367	0,349	0,373	0,361	0,358
1234	a+b	0,333	0,277	0,263	0,301	0,265	0,302	0,261	0,279	0,284
134	a+b	0,333	0,362	0,368	0,341	0,367	0,349	0,373	0,361	0,358
SSE Probit (*10³)=		**8,38**	**0,50**	**0,05**	**3,63**	**0,08**	**2,87**	**0,00**	**0,62**	**1,01**

Efficient paths K^{EFF}_{1-4}	C_k	Probabilities (cv=0.2) a/b=0.8 MNL	C-Logit	PS-Logit	PCL	LNL (δ=0)	LNL (δ ar. mean)	Probit	CoNL (δ_{min}= 0.1, 0.2, 0.3)	CoNL (δ_{min}= 0.4)
124	a+b	0,333	0,362	0,368	0,341	0,367	0,349	0,368	0,361	0,358
1234	a+b	0,333	0,277	0,263	0,317	0,265	0,302	0,264	0,279	0,284
134	a+b	0,333	0,362	0,368	0,341	0,367	0,349	0,368	0,361	0,358
SSE Probit (*10³)=		**7,23**	**0,24**	**0,00**	**4,28**	**0,00**	**2,21**	**0,00**	**0,33**	**0,63**

Table 5.3: Braess' network: route choice probabilities for the o-d pair ***1-4****.*

5.5.3 Mesh network with long bypass

The toy network in Figure 5.10 provides an interesting test for the proposed CoNL route choice model: on the one hand, it allows illustrating the rationale underlying the steps of the algorithm introduced in Section 5.4.1 on a more realistic network; on the other, it offers a more challenging correlation pattern amongst routes. In particular, the network includes 12 nodes and 14 bi-directional links. Restricting attention to the o-d pair **1-12**, and recalling the assumption of Dial efficiency with respect to the origin, the choice-set $K_{\mathbf{1\text{-}12}}^{eff\_1}$ includes 18 efficient routes (see Table 5.4) and $L_{\mathbf{1\text{-}12}}^{eff\_1}$ includes only mono-directional links.
The topology of the network for the o-d pair **1-12** is designed to challenge the structure of the network levels (and thus of the mixing components) of the CoNL route choice model: link *3-5* introduces an asymmetry in the mesh, leading to non-trivial descendance/ascendance relationships between links (see Section 5.4.1), and link *1-9* represents a bypass to include in various mixing components to satisfy condition C1.
For the sake of clarity, the first iteration of the algorithm described in Section 5.4.1, 1 generates the initial set $\Omega_1 \equiv \{1\text{-}2, 1\text{-}4, 1\text{-}9\}$ representing the forward star of origin node *1*. Then, for any subsequent generic iterations i with $i \in 2\ldots5$, links *9-10* and *9-11* (members of the forward stars of link *1-9*) are included first in the set Γ_i (step #1), and then discarded because descendants of some other links belonging to Γ_i (step #2); finally, step #3 reintroduces link *1-9*, being ascendant of eliminated links and member of $\Omega_{i\text{-}1}$. The complete sets of network levels and the nesting structure of the corresponding mixing components identified by the algorithm are shown in Figure 5.11, wherein routes belonging to each nest are not represented for the sake of brevity.
Table 5.4 reports choice probabilities for all 18 efficient routes linking o-d pair **1-12**, applying the contrasted models.

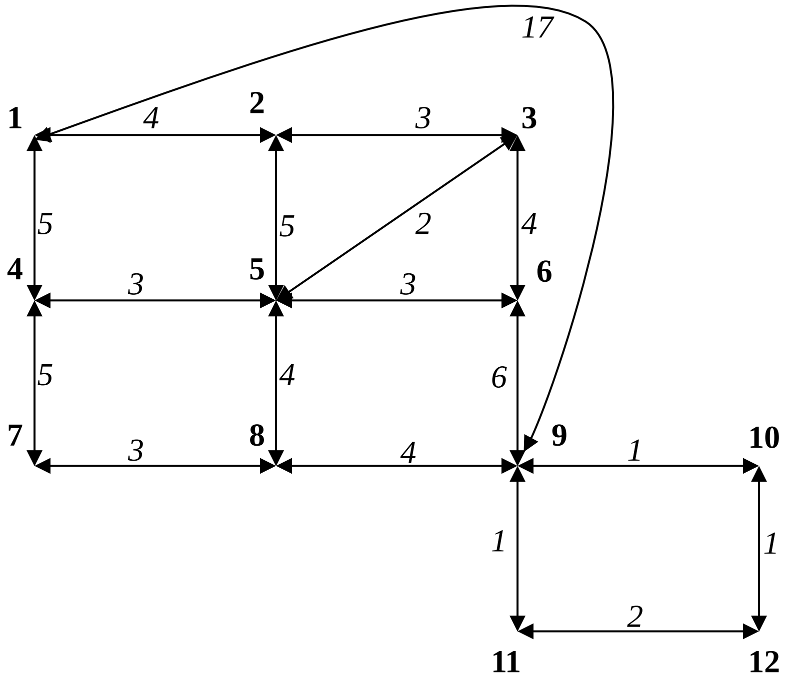


Figure 5.10: Mesh network: topology and link costs.

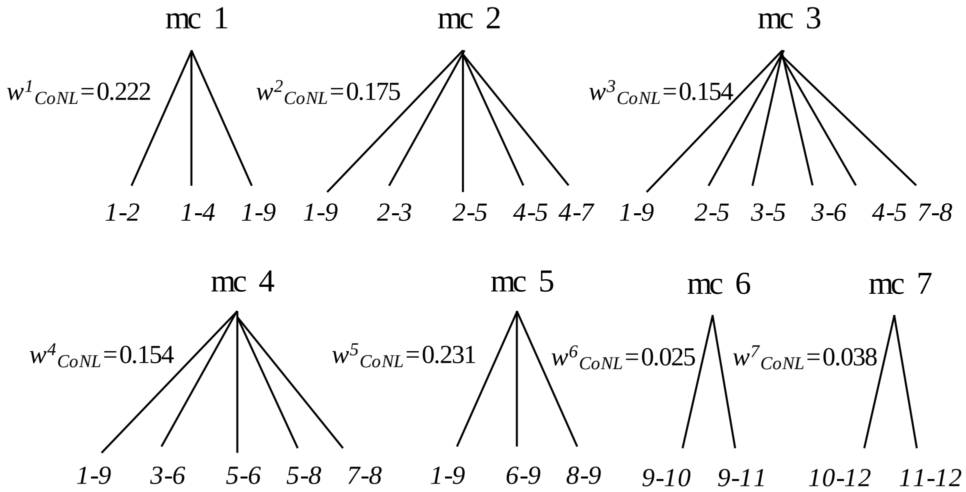


Figure 5.11: CoNL mixing components and corresponding nesting structure (elemental alternatives not illustrated for the sake of brevity) for the network of Figure 5.10, o-d pair ***1-12***

Interestingly, the CoNL route choice model outperforms all contrasted models in targeting Probit probabilities, for any value of $\delta_{\min}$ and of the coefficient of variation (*cv*). Furthermore, for *cv*=0.1, all other contrasted models perform significantly worse, with similar results, whilst for *cv*=0.2, both modified Logit models and PCL outperform the LNL, especially with null δ.

			Probabilities (cv=0.1)										
Efficient paths $K^{EFF}_{1\text{-}12}$	**ID**	**C_k**	**MNL**	**C-Logit**	**PS-Logit**	**PCL**	**LNL (δ=0)**	**LNL (δ ar.mean)**	**Probit**	**CoNL (δ_{min}=0.1)**	**CoNL (δ_{min}=0.2)**	**CoNL (δ_{min}=0.3)**	**CoNL (δ_{min}=0.4)**
1 2 3 5 6 9 10 12	**1**	20	0,000	0,032	0,029	0,035	0,010	0,039	0,018	0,014	0,015	0,019	0,023
1 2 3 5 6 9 11 12	**2**	21	0,041	0,016	0,014	0,011	0,000	0,017	0,004	0,002	0,002	0,003	0,005
1 2 3 5 8 9 10 12	**3**	20	0,057	0,033	0,029	0,036	0,010	0,041	0,021	0,013	0,013	0,016	0,022
1 2 3 5 8 9 11 12	**4**	21	0,041	0,016	0,015	0,011	0,000	0,018	0,006	0,001	0,001	0,002	0,004
1 2 3 6 9 10 12	**5**	19	0,080	0,068	0,077	0,094	0,149	0,090	0,114	0,119	0,118	0,115	0,109
1 2 3 6 9 11 12	**6**	20	0,057	0,033	0,039	0,036	0,003	0,038	0,021	0,019	0,020	0,022	0,026
1 2 5 6 9 10 12	**7**	19	0,080	0,065	0,059	0,093	0,102	0,086	0,085	0,101	0,100	0,098	0,093
1 2 5 6 9 11 12	**8**	20	0,057	0,032	0,030	0,035	0,003	0,036	0,016	0,016	0,016	0,019	0,022
1 2 5 8 9 10 12	**9**	19	0,080	0,067	0,061	0,093	0,106	0,089	0,100	0,102	0,100	0,096	0,092
1 2 5 8 9 11 12	**10**	20	0,057	0,033	0,030	0,036	0,003	0,038	0,019	0,014	0,014	0,016	0,020
1 4 5 6 9 10 12	**11**	19	0,080	0,072	0,063	0,093	0,099	0,086	0,100	0,102	0,104	0,102	0,098
1 4 5 6 9 11 12	**12**	20	0,057	0,035	0,031	0,035	0,003	0,036	0,018	0,018	0,020	0,023	0,026
1 4 5 8 9 10 12	**13**	19	0,080	0,075	0,064	0,093	0,104	0,089	0,099	0,103	0,103	0,101	0,097
1 4 5 8 9 11 12	**14**	20	0,057	0,036	0,032	0,036	0,003	0,038	0,017	0,017	0,018	0,020	0,024
1 4 7 8 9 10 12	**15**	19	0,080	0,093	0,111	0,093	0,158	0,095	0,131	0,126	0,125	0,121	0,115
1 4 7 8 9 11 12	**16**	20	0,057	0,045	0,055	0,037	0,003	0,040	0,026	0,022	0,023	0,026	0,031
1 9 10 12	**17**	19	0,080	0,171	0,174	0,095	0,240	0,114	0,172	0,202	0,197	0,180	0,160
1 9 11 12	**18**	20	0,057	0,077	0,085	0,039	0,003	0,008	0,033	0,011	0,011	0,021	0,031
SSE Probit (*10^3)=			**26,77**	**10,62**	**11,59**	**10,25**	**9,69**	**9,59**	**0,00**	**1,87**	**1,56**	**0,59**	**0,81**

			Probabilities (cv=0.2)										
Efficient paths $K^{EFF}_{1\text{-}12}$	**ID**	**C_k**	**MNL**	**C-Logit**	**PS-Logit**	**PCL**	**LNL (δ=0)**	**LNL (δ ar.mean)**	**Probit**	**CoNL (δ_{min}=0.1)**	**CoNL (δ_{min}=0.2)**	**CoNL (δ_{min}=0.3)**	**CoNL (δ_{min}=0.4)**
1 2 3 5 6 9 10 12	**1**	20	0,057	0,038	0,034	0,041	0,014	0,048	0,028	0,024	0,027	0,031	0,035
1 2 3 5 6 9 11 12	**2**	21	0,041	0,026	0,024	0,013	0,000	0,032	0,016	0,007	0,009	0,012	0,017
1 2 3 5 8 9 10 12	**3**	20	0,057	0,039	0,035	0,041	0,014	0,051	0,036	0,024	0,026	0,028	0,033
1 2 3 5 8 9 11 12	**4**	21	0,041	0,027	0,025	0,014	0,000	0,034	0,017	0,005	0,007	0,010	0,015
1 2 3 6 9 10 12	**5**	19	0,080	0,058	0,066	0,087	0,147	0,074	0,086	0,096	0,093	0,089	0,084
1 2 3 6 9 11 12	**6**	20	0,057	0,040	0,046	0,041	0,004	0,048	0,037	0,033	0,035	0,038	0,041
1 2 5 6 9 10 12	**7**	19	0,080	0,055	0,050	0,087	0,100	0,071	0,065	0,084	0,081	0,077	0,073
1 2 5 6 9 11 12	**8**	20	0,057	0,038	0,035	0,040	0,004	0,046	0,028	0,027	0,030	0,033	0,035
1 2 5 8 9 10 12	**9**	19	0,080	0,057	0,052	0,087	0,104	0,073	0,075	0,083	0,080	0,075	0,071
1 2 5 8 9 11 12	**10**	20	0,057	0,039	0,036	0,041	0,004	0,048	0,033	0,027	0,028	0,030	0,033
1 4 5 6 9 10 12	**11**	19	0,080	0,062	0,054	0,086	0,098	0,071	0,079	0,090	0,087	0,083	0,079
1 4 5 6 9 11 12	**12**	20	0,057	0,042	0,038	0,040	0,004	0,046	0,034	0,028	0,034	0,038	0,040
1 4 5 8 9 10 12	**13**	19	0,080	0,064	0,055	0,086	0,102	0,074	0,079	0,089	0,086	0,081	0,078
1 4 5 8 9 11 12	**14**	20	0,057	0,043	0,039	0,040	0,004	0,048	0,032	0,028	0,032	0,035	0,038
1 4 7 8 9 10 12	**15**	19	0,080	0,079	0,095	0,086	0,156	0,079	0,106	0,110	0,105	0,099	0,094
1 4 7 8 9 11 12	**16**	20	0,057	0,053	0,066	0,041	0,004	0,051	0,047	0,034	0,040	0,044	0,048
1 9 10 12	**17**	19	0,080	0,146	0,149	0,086	0,236	0,095	0,142	0,171	0,165	0,146	0,129
1 9 11 12	**18**	20	0,057	0,092	0,102	0,043	0,004	0,013	0,060	0,037	0,036	0,049	0,057
SSE Probit (*10^3)=			**10,47**	**4,20**	**5,03**	**5,21**	**28,18**	**7,75**	**0,00**	**2,78**	**1,90**	**0,58**	**0,65**

Table 5.4: Route choice probabilities for the o-d pair ***1-12****, on mesh network with long bypass in* Figure 5.10.

5.5.4 Grid network

For the sake of completeness, the contrasted route choice models are also tested on the grid network depicted in Figure 5.1, with focus on the o-d pair **1-4,** and with link costs $c_{1-2}=c_{8-4}=3$, $c_{7-3}=2$ and all remaining links. This yields a cost of 5 for routes #1 and #2, 6 for #3 and 7 for route #4: the corresponding route choice probabilities are reported in Table 5.5.

Efficient paths K^{eff}_{1-4}	ID	C_k	probabilities (cv=0.1) MNL	C-Logit	PS-Logit	PCL	LNL (δ=0)	LNL (δ ar.mean)	Probit	CoNL (δ_{min}= 0.1)	CoNL (δ_{min}= 0.2)	CoNL (δ_{min}= 0.3)	CoNL (δ_{min}= 0.4)
1 2 3 4	1	5	0,480	0,570	0,583	0,469	0,487	0,485	0,498	0,489	0,489	0,490	0,490
1 5 6 2 3 4	2	5	0,480	0,395	0,380	0,421	0,487	0,483	0,492	0,489	0,489	0,490	0,490
1 5 6 7 3 4	3	6	0,037	0,032	0,034	0,082	0,023	0,030	0,010	0,021	0,020	0,019	0,019
1 5 6 7 8 4	4	7	0,003	0,003	0,003	0,028	0,002	0,002	0,000	0,002	0,002	0,002	0,002
SSE Probit (*10^3)=			**1,20**	**15,28**	**20,44**	**11,84**	**0,32**	**0,67**	**0,00**	**0,22**	**0,19**	**0,17**	**0,15**

Efficient paths K^{eff}_{1-4}	ID	C_k	probabilities (cv=0.2) MNL	C-Logit	PS-Logit	PCL	LNL (δ=0)	LNL (δ ar.mean)	Probit	CoNL (δ_{min}= 0.1)	CoNL (δ_{min}= 0.2)	CoNL (δ_{min}= 0.3)	CoNL (δ_{min}= 0.4)
1 2 3 4	1	5	0,425	0,510	0,516	0,431	0,449	0,435	0,467	0,456	0,458	0,459	0,458
1 5 6 2 3 4	2	5	0,425	0,353	0,337	0,415	0,449	0,429	0,423	0,456	0,458	0,457	0,453
1 5 6 7 3 4	3	6	0,118	0,102	0,109	0,114	0,077	0,107	0,084	0,063	0,062	0,063	0,069
1 5 6 7 8 4	4	7	0,033	0,035	0,038	0,041	0,025	0,029	0,026	0,024	0,023	0,022	0,021
SSE Probit (*10^3)=			**3,02**	**7,04**	**10,52**	**2,53**	**1,08**	**1,65**	**0,00**	**1,69**	**1,82**	**1,68**	**1,26**

Table 5.5: Route choice probabilities for the network in *Figure 5.1*, o-d pair **1-4**, with the link cost configuration $c_{1-2}=c_{8-4}=3$, $c_{7-3}=2$ and *1* for all remaining links.

Even this very simple network is very illustrative of how models not embedding correlations in their generating function – such as the modified logit models - may provide results significantly worst even with respect to a simple MNL. The PCL appears also far from targeting Probit choice probabilities, especially in more deterministic contexts (e.g. cv=0.1), whilst both the CoNL and the LNL models generally perform very well for both values of the coefficient of variation.

5.5.5 Sioux-Falls network

For the sake of completeness, the CoNL route choice model has been applied also on the well-known Sioux-Falls network (Figure 5.12).

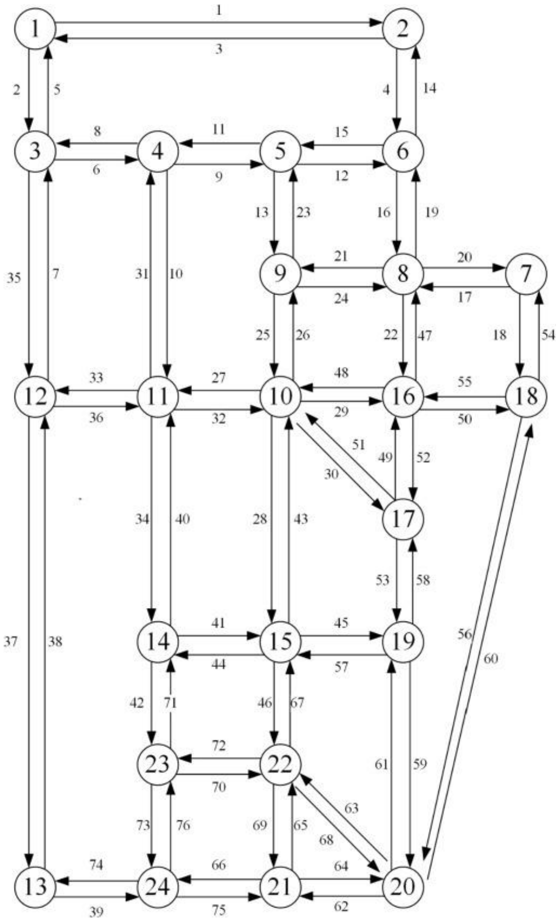


Figure 5.12: Sioux-Falls network, South Dakota.

Choice probabilities of the contrasted models are shown in Table 5.6 with reference to the o-d pair **1-15**, characterized by 16 efficient routes; the corresponding structure of the network levels, not reported for the sake of brevity, includes 10 levels. Also in this test, the CoNL route choice model is able to match very satisfactorily the Probit route choice probabilities, with an SSE outperforming all other models for any tested values of δ_{min}.

			Probabilities (cv=0.1)										
Efficient paths $K^{EFF}_{1\text{-}15}$	**ID**	C_k	**MNL**	**C-Logit**	**PS-Logit**	**PCL**	**LNL (δ=0)**	**LNL (δ ar.mean)**	**Probit**	**CoNL (δ_{min}=0.1)**	**CoNL (δ_{min}=0.2)**	**CoNL (δ_{min}=0.3)**	**CoNL (δ_{min}=0.4)**
1 2 6 8 9 10 15	**1**	32	0,001	0,002	0,001	0,000	0,001	0,001	0,000	0,000	0,000	0,000	0,000
1 2 6 8 9 10 17 19 15	**2**	39	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
1 2 6 8 16 17 19 15	**3**	25	0,061	0,091	0,064	0,055	0,107	0,070	0,100	0,100	0,097	0,094	0,090
1 3 4 5 6 8 9 10 15	**4**	35	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
1 3 4 5 6 8 16 17 19 15	**5**	28	0,011	0,010	0,010	0,006	0,003	0,010	0,001	0,003	0,003	0,003	0,004
1 3 4 5 9 10 15	**6**	24	0,106	0,090	0,079	0,105	0,124	0,109	0,125	0,100	0,103	0,108	0,111
1 3 4 5 9 10 17 19 15	**7**	31	0,002	0,002	0,002	0,001	0,001	0,001	0,000	0,000	0,000	0,000	0,000
1 3 4 11 10 15	**8**	25	0,061	0,051	0,041	0,051	0,011	0,052	0,030	0,030	0,030	0,032	0,036
1 3 4 11 10 17 19 15	**9**	32	0,001	0,001	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
1 3 4 11 14 15	**10**	23	0,185	0,177	0,166	0,201	0,226	0,190	0,211	0,222	0,218	0,212	0,205
1 3 12 11 10 15	**11**	25	0,061	0,051	0,041	0,051	0,011	0,051	0,032	0,030	0,030	0,031	0,035
1 3 12 11 10 17 19 15	**12**	32	0,001	0,001	0,001	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000
1 3 12 11 14 15	**13**	23	0,185	0,181	0,168	0,201	0,199	0,188	0,196	0,203	0,203	0,200	0,197
1 3 12 13 24 21 22 15	**14**	23	0,185	0,196	0,248	0,203	0,261	0,195	0,227	0,232	0,229	0,224	0,219
1 3 12 13 24 23 14 15	**15**	26	0,035	0,035	0,042	0,024	0,009	0,028	0,004	0,011	0,011	0,011	0,013
1 3 12 13 24 23 22 15	**16**	24	0,106	0,112	0,136	0,102	0,047	0,104	0,074	0,069	0,075	0,083	0,090
SSE Probit (*10³)=			**8,35**	**6,89**	**12,32**	**5,19**	**3,02**	**5,07**	**0,00**	**0,93**	**0,64**	**0,50**	**0,79**

			Probabilities (cv=0.2)										
Efficient paths $K^{EFF}_{1\text{-}15}$	**ID**	C_k	**MNL**	**C-Logit**	**PS-Logit**	**PCL**	**LNL (δ=0)**	**LNL (δ ar.mean)**	**Probit**	**CoNL (δ_{min}=0.1)**	**CoNL (δ_{min}=0.2)**	**CoNL (δ_{min}=0.3)**	**CoNL (δ_{min}=0.4)**
1 2 6 8 9 10 15	**1**	32	0,011	0,017	0,010	0,003	0,007	0,009	0,005	0,003	0,003	0,003	0,003
1 2 6 8 9 10 17 19 15	**2**	39	0,002	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
1 2 6 8 16 17 19 15	**3**	25	0,078	0,139	0,085	0,072	0,157	0,091	0,136	0,145	0,140	0,132	0,124
1 3 4 5 6 8 9 10 15	**4**	35	0,005	0,004	0,004	0,000	0,000	0,003	0,000	0,000	0,000	0,000	0,000
1 3 4 5 6 8 16 17 19 15	**5**	28	0,034	0,030	0,032	0,018	0,010	0,033	0,016	0,012	0,012	0,015	0,018
1 3 4 5 9 10 15	**6**	24	0,104	0,087	0,079	0,111	0,138	0,108	0,126	0,103	0,111	0,115	0,116
1 3 4 5 9 10 17 19 15	**7**	31	0,015	0,013	0,011	0,005	0,008	0,012	0,003	0,003	0,003	0,003	0,003
1 3 4 11 10 15	**8**	25	0,078	0,063	0,054	0,068	0,016	0,072	0,064	0,045	0,049	0,054	0,060
1 3 4 11 10 17 19 15	**9**	32	0,011	0,009	0,008	0,003	0,000	0,008	0,001	0,002	0,002	0,002	0,003
1 3 4 11 14 15	**10**	23	0,137	0,127	0,126	0,167	0,191	0,141	0,161	0,182	0,171	0,163	0,156
1 3 12 11 10 15	**11**	25	0,078	0,065	0,055	0,068	0,016	0,071	0,063	0,045	0,048	0,053	0,059
1 3 12 11 10 17 19 15	**12**	32	0,011	0,009	0,008	0,003	0,000	0,008	0,001	0,002	0,002	0,002	0,003
1 3 12 11 14 15	**13**	23	0,137	0,130	0,128	0,167	0,167	0,139	0,151	0,166	0,161	0,156	0,151
1 3 12 13 24 21 22 15	**14**	23	0,137	0,141	0,189	0,168	0,220	0,145	0,157	0,187	0,178	0,171	0,166
1 3 12 13 24 23 14 15	**15**	26	0,059	0,058	0,073	0,042	0,018	0,054	0,027	0,025	0,027	0,030	0,035
1 3 12 13 24 23 22 15	**16**	24	0,104	0,106	0,137	0,108	0,052	0,106	0,089	0,080	0,093	0,101	0,105
SSE Probit (*10³)=			**7,44**	**5,13**	**12,62**	**5,49**	**11,67**	**4,66**	**0,00**	**2,96**	**1,34**	**0,71**	**0,72**

*Table 5.6: Route choice probabilities for the Sioux-Falls network of Figure 5.12, o-d pair **1-15.***

5.5.6 Summary of experimental analysis

At a glance, the CoNL route choice model has been proved to be capable to perform very well in all tested networks, having provided almost always the best fit with respect to Probit choice probabilities and, more importantly, having never failed in targeting expected route choice probabilities. The same did not occur for the other tested models, including modified Logit models, the LNL with null δ's and the PCL whose performances do vary significantly across tested networks.

In terms of specification, fixing a lower bound δ_{min} for the nesting parameters helps stabilizing the performances of the CoNL, for the motivation reported in Section 5.3.2. Importantly, the experimental analysis indicates that relaxing such lower bound increases the variability of CoNL performances, however keeping the CoNL the best model in most of the tests. Overall, a rule-of-thumb is to set δ_{min} in between the interval [0.3-0.4], so as to stabilize the CoNL performances; alternatively, δ_{min} can be estimated of course based on disaggregated data.

It is also worth underlining that the CoNL capability to target Probit route choice probabilities better than other models is a straightforward consequence of its mathematical background, that allows specifying directly a CoNL model consistent with an underlying correlation structure and its corresponding correlation values. In fact, all LNL and PCL specifications proposed in the literature – see for instance Gliebe et al., (1999), Vovsha and Bekhor (1998) and Prashker and Bekhor (1998) – belong to the family of the Cross-Nested Logit model, characterized by a non-closed-form expression of its correlations and incapable to allow for a straightforward specification of the model to target a correlation matrix (Marzano and Papola, 2008; Marzano et al. 2013; Marzano, 2014). As a consequence, PCL and LNL route choice specifications are motivated only by empirically sound motivations. Conversely, the CoNL closed form correlation expression (2.126), clearly stating the relationship between model structure/parameters and corresponding underlying correlations, easily allows targeting Daganzo and Sheffi (1977) correlations through the route choice adaptation proposed in Section 5.3.

5.6 Conclusions and research prospects

The chapter has illustrated the application to route choice of the Combination of Nested Logit (CoNL) model, that is a particular specification of the Combination of RUMs (CoRUM) model proposed by Papola (2016), whose key feature is the availability of a closed-form statement for both choice probabilities and correlations. These features have allowed to propose a specific CoNL route choice adaptation, capable to target Daganzo and Sheffi (1977) correlations, differently from all the other closed form route choice models proposed so far – see for instance the PCL of Gliebe et al., (1999), and the LNL model of Vovsha and Bekhor (1998) and of Prashker and Bekhor (1998) – which should be regarded only as empirically sound specifications not motivated by a specific theoretical target. This more solid theoretical background allows the proposed model to outperform all the other closed form route choice models proposed so far. The proposed model has been also operationalized by means of an algorithm, providing CoNL route choice probabilities on a set of explicitly enumerated paths for a given o-d pair.

For each o-d pair, the proposed CoNL route choice specification adopts routes as alternatives and shared links as elemental nests, grouped in mixing components representing levels of the transport network. In particular, a network level represents a set of links such that the corresponding mixing component should satisfy two properties: (a) each mixing component should include all efficient routes connecting that o-d pair as elemental alternatives; (b) each efficient route should belong to only one nest in each mixing component, for the latter to be a Nested Logit. In this respect, an algorithm capable to detect network levels and to provide the corresponding mixing components on a set of enumerated paths without explicit inspection of the routes has been proposed, that is a CoNL specification allowing implicit probability statement.

The performance of the CoNL route choice model have been tested on various networks, contrasting its choice probabilities with other route choice models available in the literature, always leading to very satisfactory results. In particular, albeit the CoNL is a homoscedastic model - and hence structurally unable to model heteroscedasticities - its route choice probabilities are always very close to the Probit route choice probabilities, i.e. the benchmark model under the assumption of Daganzo and Sheffi (1977) route correlation pattern.

Two straightforward research prospects are envisaged. The former refers to the possibility of implementing a CoNL route choice model without explicit enumeration of the set of feasible paths for each o-d pair, leveraging the conceptual vehicle of the algorithm illustrated in Section 5.4.1. The latter aims at completing the assessment of the performance of the CoNL route choice model by means of a disaggregated estimation on individual route choice data.

Chapter 6: Some advance on CoNL route choice model

This chapter proposes some theoretical and practical advance on route choice modelling. The CoNL route choice model (Papola et al., 2018) has been presented in Chapter 5, to give an answer to some relevant issues in modelling choice behaviour when choosing paths within a network. Although its appealing property of closed-form correlation expression and its great robustness in reproducing target Multinomial Probit probabilities (see Section 5.5), it seems to have several margins of improvement. First, the model has been tested only with the explicit enumeration of paths and this could notoriously have a high computational burden in presence of large-size networks. Second, the CoNL has been tested only on efficient choice sets with respect to each origin, potentially leading to undesirable eliminations of paths from the choice-set. Third, the procedure previously proposed for computing the model parameters can be improved, to better catch the correlation effects on route choice probabilities. With this regards, an implicit enumeration algorithm is proposed, implementing the CoNL route choice model as presented in Chapter 5, i.e. with the only restriction due to Dial-efficiency. Furthermore, the CoNL route choice model has been tested with an acyclic choice set, without Dial-efficiency restriction, and slight changes in the model parameters computing are proposed. A study on the correlations reproduced by the main route choice models is shown, to investigate the real impact of capacity to reproduce overlapping effect (Daganzo and Sheffi, 1977) on choice probabilities. Several tests are conducted both on toy networks and on a real network, computing the likelihood of a real dataset of route choice observations, based on trajectory data of drivers moving within the Regione Campania network.

6.1 Route choice modelling issues – a brief recap

6.1.1 The computational problem in route choice: implicit enumeration algorithms

As recalled in Chapter 2, the first computational source of complexity in route choice modelling is represented by the size of the choice set. Assuming the classic path definition (i.e. a sequence of network links connecting an o-d pair), the problem on a real size network comes in a threefold way: 1) the number of links constituting a route 2) the number of acyclic path connecting an o-d pair 3) the number of o-d pairs that is generally necessary to guarantee a good zoning level of the area of study.

Thus, the explicit route enumeration may represent a very burdensome operation that one needs to sustain before working with a traffic assignment procedure. For solving this problem, under some (more or less) restrictive hypothesis, several implicit enumeration methods have been proposed in literature. The main assumption of all the classic implicit route choice models is working only with additive link costs. The reader can refer to Section 2.3 for a more comprehensive discussion on the main state of the art contribution.

The first stochastic network-loading (SNL) algorithm is the Dial algorithm (Dial, 1971), implementing a MNL model for route choice, via the computation of entities so-called Dial weights, allowing for link flows computation on the basis of a simple recursive network exploration. However, in addition to the well-known limitations of the MNL itself, the algorithm presents other two problems. First, the algorithm implicitly works on a selective choice set, given by the so-termed efficient paths with reference to a given origin o, i.e. it works only with links allowing to walk away from the origin (and/or to get closer to the destination). Second, in its most implemented version, given an origin o, it does not allow the variance parameter differentiation for the different destination d. The latter represents, in general, an unrealistic assumption, because of bigger is the route cost and higher would be the expected dispersion of the residuals.

For overcoming the MNL and Dial algorithm limitations, the MNP is commonly used, particularly with the Monte Carlo algorithm (Sheffi, 1985), drawing link costs from mono-variate density distributions and exploiting, at each drawing iteration, the application of a minimum cost path algorithm. Despite the desirable properties of the MNL, it suffers from the burdens due to simulation. Performing an SNL Probit requires a big number of draws for reaching stable values of the flows. The problem is worsening in a S.U.E. procedure, because of the repetition of the simulation needed for each fixed-point research iteration. Moreover, a computational advantage descending from Daganzo and Sheffi (1977) hypothesis on the dispersion of link impedances is represented by the dependence of the entire procedure from a unique proportionality parameter. But this means, as explained in Section 2.3.4.1, considering a coefficient of variation decreasing with link costs. This could represent another unrealistic hypothesis in a real world application. This is the reason why several authors proposed implicit enumeration algorithms that, actually, did not receive a strong interest in the practical applications. The Link Nested Logit (Vovsha and Bekhor, 1998) gives the possibility to avoid the enumeration of paths, but implementing a particular case of the CNL model (Vovsha, 1997), consisting of assuming a null value for all the nesting parameters. This assumption, as discussed in Papola et al.(2018), can be unpleasant, because of it enhances the effects of differences in systematic utility values on choice probabilities, since the choice within each nest tends to be too deterministic (see Section 2.3.3.1 for a discussion).

Another route choice solution is represented by the MNL with a deterministic correction for the overlapping link impedances. The first formulation that has been proposed in literature is the C-Logit model (Cascetta et al., 1996), wherein the systematic utilities are corrected by means of the so-called communality factor of the route (CF_k). Russo and Vitetta (2003) proposed an implicit enumeration paths procedure referring to one of the original formulation of the CF_k. Nevertheless, all the approximate MNL formulations with correction of utilities have shown their structural limitations (see Prashker and Bekhor, 1998; Prashker and Bekhor, 2004; Marzano, 2006, Frejinger and Bierlaire, 2007; Papola and Marzano, 2013; Papola et al., 2018). Papola and Marzano (2013) proposed an operationalization of the Network GEV model (Daly and Bierlaire, 2001) for route choice, with the possibility of implementing the model with the implicit enumeration of paths, by means of a single-step procedure on the efficient sub-network. Unfortunately, the underlying model does not allow to compute the covariances in a closed form. This means the model does not guarantee to catch the Daganzo and Sheffi (1977) covariances paradigm.

More recent developments on the subject are the Markovian Traffic Assignment procedures (Akamatsu, 1996, 1997; Baillon and Cominetti, 2008; Fosgerau et al., 2013, Mai et al., 2015; Mai, 2016). They are receiving a lot of attention in literature, because they allow assigning link flows on network with recursive algorithms, without need of restricting the set of available paths to the acyclic paths. This means considering an infinite choice set of cyclic paths, theoretically including paths with infinite cycles. It is a writer's opinion that, even though cyclic paths consideration can be a benefit in some appropriate circumstances (for instance, the searching for parking close to the destination or the pedestrian flows in some uncertainty circumstances), those methodologies have not yet demonstrated their effectiveness in reproducing real observed choice with reference to the classic approaches. The real (great!) danger is considering also very unrealistic cyclic paths, because no restriction is imposed and no differentiation is made between a path with a cycle close to the destination and another path passing more than once by the origin, or whatever intermediate node of a path connecting an o-d pair.
Thus, working with an implicit enumeration algorithm whose underlying choice model is theoretically robust, i.e. a model that structurally considers the covariances among unobservable factors in route choice, and that avoid simulations, represents, even now, an open research topic.

6.1.2 The overlapping problem in route choice

Albeit many definitions of routes exist, and diverse choice paradigms are available, the majority of contributions in route choice modelling field draws upon the classical definition of route as an ordered sequence of links connecting an origin-destination o-d pair and upon the Random Utility Models (RUMs) framework (Ben-Akiva and Lerman, 1985; Cascetta, 2009; Train, 2009). Within this framework, an unanimously acknowledged peculiarity of the route choice context is the presence of a complex correlation structure amongst perceived utilities of alternative routes – i.e. the so-called correlations among routes – structurally determined by the topological overlapping of alternative routes in a transport network. In this respect, Daganzo and Sheffi (1977) assumed the correlation between any pairs of routes to be proportional to their topological overlapping, measured using a given link impedance. Such assumption can be applied to the network as a whole or, alternatively, only to a portion of the network given by primary, most likely perceived, roads, as proposed by Frejinger and Bierlaire (2007). Daganzo and Sheffi (1977) identified the Multinomial Probit model (MNP) as a natural to operationalize their assumption into a route choice model, thanks to the possibility offered by the MNP model to specify directly its correlation matrix. As a result, the MNP is regarded as the reference RUM to account for the effects of correlations in route choice modelling.
Unfortunately, as recalled in Chapter 2, the MNP suffers from the absence of a closed-form probability statement, leading to computational issues related to the need to simulate choice probabilities. The same also applies to Mixed Logit applications to route choice, e.g. Bekhor et al. (2002), Frejinger and Bierlaire (2007). Thus, a challenging research question has been tackled by many researchers: can closed-form RUM-based route choice models be specified consistent with the Daganzo and Sheffi (1977) assumption?
The Generalized Extreme Value (GEV) modelling framework, comprehensively defined by the GEV model by McFadden (1978), represents the most straightforward mathematical vehicle to deal with this issue. The simplest GEV-based route choice model was proposed by Dial (1971), who applied the Multinomial Logit model (MNL) to route choice with an elegant and computationally very effective algorithm to calculate route choice probabilities without explicit

route enumeration. Unfortunately, the MNL model hypothesizes null correlation amongst perceived utilities of alternatives, because of its underlying distributional assumptions. Thus, several researchers introduced correction/penalty factors in the systematic utility of a MNL model, as a trick to mimic the effect of correlations on route choice probabilities: relevant examples in this respect are the above mentioned C-Logit model (Cascetta et al., 1996) and the Path-size model (Ben-Akiva and Ramming, 1998; Ben-Akiva and Bierlaire, 1999; Ramming, 2001; Hoogendoorn-Lanser et al., 2005). However, these models exhibit limitations in capturing the proper effect of route correlations on choice probabilities (see the references in given in Section 2.3.2.4 as conclusive discussion on the Logit based route choice models with deterministic utility correction).

Consequently, a theoretically sound research direction is aimed at applying "more complex" GEV models allowing for nonzero correlations amongst alternatives – such as the Cross-Nested Logit model (Vovsha, 1997) and the Network GEV model (Daly and Bierlaire, 2001*b*) – to route choice modelling. In this context, many models have been proposed so far, including the Link-Nested Logit (LNL) model by Vovsha and Bekhor (1998), the Paired-Combinatorial Logit (PCL) model by Prashker and Bekhor (1998), and the Path Multilevel PML (also known as Link Based Network GEV) model by Papola and Marzano (2013). Interestingly, there is no assessment in the literature on the capability of such models to target Daganzo and Sheffi (1977) correlations, being their performance rather evaluated just in terms of capability to target MNP route choice probabilities.

The GEV models is not the sole closed-form models to apply to route choice modelling. In Chapter 5 a new route choice model has been presented, i.e. the CoNL route choice model (Papola et al., 2018), that is a particular specification of the Combination of RUMs (CoRUM) model proposed by Papola (2016), whose key feature is the availability of a closed-form statement for both choice probabilities and correlations.

In particular, the proposed CoNL route choice model is characterized by a closed-form expression of choice probabilities, and operationalized by means of an algorithm providing CoNL route choice probabilities on a set of explicitly enumerated paths for a given o-d pair. The mathematical structure of the CoRUM allows handling effectively the relationship between the CoNL specification (model structure, parameters) and its underlying correlations, thus enabling the possibility of specifying a CoNL capable to target Daganzo and Sheffi (1977) correlations. In this respect, however, the CoNL model is still homoscedastic, and the impact of the operationalization of its parameters on the values of correlations values is not fully explored yet.

As a result, neither GEV-based route choice models (the LNL and the PCL, namely) nor the CoNL route choice model have been explored fully in their underlying correlations.

For GEV-based route choice models, this is actually a difficult task, because of the non-closed form of the expression of the correlations underlying any network-GEV models, as extensively studied by Abbé et al. (2007) and Papola and Marzano (2008). This implies numerical evaluation of double-integrals with strongly nonlinear integrands, a circumstance that likely prevented the aforementioned correlation assessment. In this respect, however, Marzano et al. (2013) for the CNL model and Marzano (2014) for any network GEV models, provided a simpler and more effective methodology for the calculation of GEV correlations, based on the numerical integration of a mono-dimensional integral, thus with parsimonious calculation times.

For the CoRUM (and CoNL) model, calculation of correlation is a straightforward task, thanks to the inherent model structure. However, the CoNL route choice model has been just released by Papola et al. (2018), thus a full assessment of the correlation values – being the correlation

pattern already consistent by definition with Daganzo and Sheffi (1977) – has not been exploited yet.
Given these premises, this chapter aims at filling this gap, providing a comprehensive analysis of the correlation values underlying LNL, PCL and CoNL route choice models. This analysis is also conducted in order to set some slight changes in computing CoNL route choice model structural parameters, investigating the impact of correlations and the procedure on choice probabilities values.

6.2 An implicit enumeration algorithm for CoNL route choice

6.2.1 CoNL recursive equations

Briefly recalling the notation of Chapter 5, we assume:

- $G \equiv \{L,N\}$ be a graph representing a road network;
- L a set of links l, whose size is n_L;
- N a set of nodes n, whose size is n_N;
- C a set of centroids, whose size is n_C;
- OD a set of o-d pair, whose size is n_C^2;
- $t(l)$, $h(l)$ respectively the tail and the head of link l;
- EFS(n) and EBS(n), respectively, the forward and the backward star of the node n;
- $c_l \; \forall l \in L$ the additive link impedance;
- K_{od} be a set of acyclic routes connecting the pair of centroids o and d with o, $d \in C$;
- $C_k = \Sigma_l \, c_l \; \forall l \in L_k$ the route cost;
- $L_k \subseteq L$ the ordered set of links associated with each route $k \in K_{od}$;
- $L_{od} \equiv \{\cup L_k \; \forall k \in K_{od}\}$ the collection of all links of all routes within K_{od};
- $K_l \subseteq K_{od} \equiv \{\cup K_l \; \forall l \in L_{od}\}$ the set of routes k including l;
- n_{MC} the number of mixing components in the CoNL route choice model;
- i the generic mixing component of the CoNL model;
- Ω_i the subset of network links associated with the generic i-th mixing component.
- I the set of all Ω_i;
- I^l the subset of mixing components which a link l belongs to;
- $I^{jm} \equiv \{\cup_l \, I^l, \; l \in L_j \cap L_m\}$ the set of sharing links of paths j and m;

The CoNL model (Papola, 2016) is based on the assumption (2.122) on random residuals, implying the (2.123) for choice probabilities. The latter can be written expressing the NL probabilities in using the Bayes theorem:

$$p(k) = \sum_{i \in I} w^i \cdot p^i(k/l^i) \cdot p^i(l^i) \qquad \forall k \in K_{od} \tag{6.1}$$

being k^i the generic nest associated with the i-th mixing NL. Note that the second term of the right side represents the probability of choosing path k within the nest l^i, while the third term represents the probability of nest l^i.
Following the same arguments of Dial (1971), the second term can be expressed as a sequence of conditional link choice probabilities. Indifferently, a conditional link choice probability can be expressed viewing at the choice as conditioned on coming from tail $t(l)$ or head $h(l)$, where l belongs to L_k:

$$p^i(k/l^i) = \prod_{l\in L_k} p(l/\mathrm{t}(l), l^i) = \\ = \prod_{l\in L_k} p^i(l/\mathrm{h}(l), l^i) \qquad \forall k \in \mathrm{K}_{\mathrm{od}} \tag{6.2}$$

With reference to the second expression on the right-side of (6.2), the generic link choice probability can be expressed as a Multinomial Logit probability among the links belonging to the EBS(h(*l*)), wherein it appears θ as the specific Gumbel variance parameter for that o-d pair, as:

$$\begin{aligned} p^i(k/l^i) &= \prod_{l\in L_k} \frac{\exp(V_{l/(\mathrm{h}(l))}/\theta)}{\sum_{l'\in \mathrm{EBS}(\mathrm{h}(l))} \exp(V_{l'/(\mathrm{h}(l))}/\theta)} = \\ &= \prod_{l\in L_k} \frac{\exp((-c_l + \theta\cdot Y_{\mathrm{t}(l)})/\theta)}{\sum_{l'\in \mathrm{EBS}(\mathrm{h}(l))} \exp((-c_{l'} + \theta\cdot Y_{\mathrm{t}(l')})/\theta)} = \\ &= \prod_{l\in L_k} \frac{\exp((-c_l + \theta\cdot \ln(\sum_{l'\in \mathrm{EBS}(\mathrm{t}(l))} \exp(V_{l'/(\mathrm{t}(l))}/\theta))/\theta)}{\sum_{l'\in \mathrm{EBS}(\mathrm{h}(l))} \exp((-c_{l'} + \theta\cdot \ln(\sum_{l'\in \mathrm{EBS}(\mathrm{t}(l))} \exp(V_{l'/(\mathrm{t}(l))}/\theta))/\theta)} = \\ &= \prod_{l\in L_k} \frac{\exp(-c_l/\theta)\cdot \sum_{l'\in \mathrm{EBS}(\mathrm{t}(l))} \exp(V_{l'/(\mathrm{t}(l))}/\theta)}{\sum_{l'\in \mathrm{EBS}(\mathrm{h}(l))} \exp(-c_{l'}/\theta)\cdot \sum_{l'\in \mathrm{EBS}(\mathrm{t}(l))} \exp(V_{l'/(\mathrm{t}(l))}/\theta)} \qquad \forall k \in \mathrm{K}_{\mathrm{od}} \end{aligned} \tag{6.3}$$

Evidently, each term of the product represent a Multinomial Logit probability on the link utilities and, above all, each link utility represents the utility of all the paths connecting h(*l*) to the considered destination *d*. Thus, (6.3) can be written as:

$$p^i(k/l^i) = \prod_{l\in L_k} \frac{\exp(-c_l/\theta)\cdot \sum_{k'\in \mathrm{K}_{(o-\mathrm{t}(l))}} \exp(V_{k'}/\theta)}{\sum_{l'\in \mathrm{EBS}(\mathrm{h}(l))} \exp(-c_{l'}/\theta)\cdot \sum_{k'\in \mathrm{K}_{(o-\mathrm{t}(l'))}} \exp(V_{k'}/\theta)} \qquad \forall k \in \mathrm{K}_{\mathrm{od}} \tag{6.4}$$

Dial(1971) expressed the last term of (6.3) as ratio between two entities, respectively called link weight and node weight. The conditional probability becomes:

$$p^i(k/l^i) = \prod_{l\in L_k, k\in \mathrm{K}_{l^i}} \frac{w_l}{W_{\mathrm{h}(l)}} \qquad \forall k \in \mathrm{K}_{\mathrm{od}} \tag{6.5}$$

being:

$$w_l = \exp(-c_l/\theta) \sum_{l'\in \mathrm{EBS}(\mathrm{t}(l))} \exp(V_{l'/(\mathrm{t}(l))}/\theta) = \\ = \exp((-c_l + Y_{t(l)})/\theta) \tag{6.6}$$

and:

$$W_{\mathrm{h}(l)} = \sum_{l'\in \mathrm{EBS}(\mathrm{h}(l))} \exp(-c_{l'}/\theta) \sum_{l'\in \mathrm{EBS}(\mathrm{t}(l))} \exp(V_{l'/(\mathrm{t}(l))}/\theta) = \\ = \sum_{l'\in \mathrm{EBS}(\mathrm{h}(l))} \exp((-c_{l'} + Y_{t(l')})/\theta) = \sum_{l'\in \mathrm{EBS}(\mathrm{h}(l))} w_{l'} \tag{6.7}$$

The computation of the link and node weights can be computed through a procedure that explores the nodes of the network, with a precise topological order (increasing minimum cost path for reaching the node), and the flows can be assigned without need of path enumeration.

In addition, the third term of (6.1) can be expressed as a function of Dial weights. Assuming the notation W_{n_o} indicating the Dial weight of node *n* with reference to the origin *o*, the nesting probability can be expressed as:

$$p^i(l^i) = \frac{\left[W_{\mathrm{t}(l)_o} \cdot \exp(-c_l / \theta_l) \cdot W_{\mathrm{h}(l)_d}\right]^{\delta_l}}{\sum_{l' \in \Omega_i} \left[W_{\mathrm{t}(l')_o} \cdot \exp(-c_{l'} / \theta_{l'}) \cdot W_{\mathrm{h}(l')_d}\right]^{\delta_{l'}}} \tag{6.8}$$

Proof: The NL nest probability can be expressed as a MNL choice probability among the nesting groups (Mc Fadden, 1978):

$$p^i(l^i) = \frac{\left[\exp(V_l / \theta\ )\right]}{\sum_{l' \in \Omega_i} \left[\exp(V_{l'} / \theta\ )\right]} \tag{6.9}$$

Being the nest represented by the link grouping all elemental alternatives/routes sharing the same link/nest, the utility of the link can be expressed as the utilities of a group including all routes sharing *l*. Thus, a generic link utility can be expressed as:

$$V_l = E\{\max_{k \supseteq l}[U_k]\} \tag{6.10}$$

According to the i.i.d. Gumbel assumption on random residuals, the expected value of perceived link utility is the logsum of the utilities U_k sharing the same link *l*:

$$\begin{aligned} V_l &= \theta_l \cdot \ln[\sum_{k' \supseteq l} \exp(V_{k'} / \theta\ )] = \\ &= \theta_l \cdot \ln[\sum_{k' \supseteq l} \exp(-\sum_{l' \in k'} c_{l'} / \theta)] = \\ &= \theta_l \cdot \ln[\sum_{k' \supseteq l} \prod_{l' \in k'} \exp(-c_{l'} / \theta)] \end{aligned} \tag{6.11}$$

Looking at the last term on the right side of (6.11) it can be noted that $\exp(-c_l / \theta)]$ is a common term to all the adding terms of the summation in *k*'. Thus, it can be put in evidence out of the sum, and the product made on all links *l*' that are different from *l*. This mean writing the (6.8) in the form:

$$V_l = -c_l + \theta_l \cdot \ln[\sum_{k' \supseteq l} \exp(-\sum_{l' \in k', l' \neq l} c_{l'} / \theta)] \tag{6.12}$$

Defining k'$_{o\text{-}t(l')}$ as the path connecting the origin *o* with the tail of the generic link *l*' and, analogously, k'$_{h(l')\text{-}d}$ as the path connecting the head of the generic link *l*' with the destination *d*, (6.12) can be re-written as:

$$\begin{aligned} V_l &= -c_l + \theta_l \cdot \ln[\sum_{k' \supseteq l} \exp(-\sum_{l' \in k'_{o-\mathrm{t}(l)}} c_{l'} - \sum_{l' \in k'_{\mathrm{h}(l)-d}} c_{l'}) / \theta)] = \\ &= -c_l + \theta_l \cdot \ln[\sum_{k'_{o-\mathrm{t}(l)} \supseteq l} \exp(-\sum_{l' \in k'_{o-\mathrm{t}(l)}} c_{l'} / \theta) \cdot \sum_{k'_{\mathrm{h}(l)-d} \supseteq l} \exp(-\sum_{l' \in k'_{\mathrm{h}(l)-d}} c_{l'} / \theta))] = \\ &= -c_l + \theta_l \cdot \{\ln[\sum_{k'_{o-\mathrm{t}(l)} \supseteq l} \exp(-\sum_{l' \in k'_{o-\mathrm{t}(l)}} c_{l'} / \theta)] + \ln[\sum_{k'_{\mathrm{h}(l)-d} \supseteq l} \exp(-\sum_{l' \in k'_{\mathrm{h}(l)-d}} c_{l'} / \theta))]\} = \\ &= -c_l + \theta_l \cdot \{\ln[\sum_{k'_{o-\mathrm{t}(l)} \supseteq l} \exp(-C_{k'_{o-\mathrm{t}(l)}} / \theta)] + \ln[\sum_{k'_{\mathrm{h}(l)-d} \supseteq l} \exp(-C_{k'_{\mathrm{h}(l)-d}} / \theta)]\} = \end{aligned} \tag{6.13}$$

Finally, defining as $Y_{\mathrm{t}(l)_o}, Y_{\mathrm{h}(l)_d}$, respectively, the logsum of utilities of all paths connecting origin *o* with the node t(*l*) and the logsum of utilities of all paths connecting the node h(*l*) with the destination *d*, (6.13) can be more succinctly written as:

$$V_l = -c_l + \theta_l \cdot Y_{\mathrm{t}(l)_o} + \theta_l \cdot Y_{\mathrm{h}(l)_d} \tag{6.14}$$

The numerator of the link probability can be expressed as MNL choice probability among links as:

$$\begin{aligned}
&\exp(V_l / \theta) = \exp((-c_l + \theta_l \cdot Y_{\mathrm{t}(l)_o} + \theta_l \cdot Y_{\mathrm{h}(l)_d}) / \theta) = \\
&= \exp(-c_l / \theta + \delta_l \cdot Y_{\mathrm{t}(l)_o} + \delta_l \cdot Y_{\mathrm{h}(l)_d}) = \\
&= \exp(-c_l / \theta) \cdot \exp(\delta_l \cdot Y_{\mathrm{t}(l)_o}) \cdot \exp(\delta_l \cdot Y_{\mathrm{h}(l)_d}) = \\
&= \exp(-c_l / \theta) \cdot [\sum_{l' \in EBS(\mathrm{t}(l))} \exp(V_{l'/\mathrm{h}(l')} / \theta_l)]^{\delta_l} \cdot [\sum_{l' \in EBS(d)} \exp(V_{l'/d} / \theta_l)]^{\delta_l} = \\
&= \exp(-c_l / \theta) \cdot [\sum_{l' \in EBS(\mathrm{t}(l))} \exp(-c_{l'} + \theta \cdot Y_{\mathrm{t}(l')} / \theta_l)]^{\delta_l} \cdot \ldots \\
&\ldots \cdot [\sum_{l' \in EBS(d)} \exp(-c_{l'} + \theta \cdot Y_{\mathrm{t}(l')} / \theta_l)]^{\delta_l}
\end{aligned} \tag{6.15}$$

Remembering the Dial's assumption (6.7), the numerator can be finally re-written as the numerator in the (6.8) that we want to proof:

$$\begin{aligned}
&\exp(V_l / \theta) = \exp(-c_l / \theta) \cdot \left[W_{\mathrm{t}(l)_o} \cdot W_{\mathrm{h}(l)_d} \right]^{\delta_l} = \\
&= \left[W_{\mathrm{t}(l)_o} \cdot \exp(-c_l / \theta_l) \cdot W_{\mathrm{h}(l)_d} \right]^{\delta_l}
\end{aligned} \tag{6.16}$$

The proof for the denominator is immediately obtained, because of the latter is a sum referred to the links of the set Ω_i of the same quantities in (6.16).
Putting together (6.5) and (6.8), the (6.1) can be purely expressed as a function of Dial's weights and CoNL parameters as:

$$p(k) = \sum_{i \in \mathrm{I}} w^i \cdot \left(\prod_{l \in L_k, k \in \mathrm{K}_{l^i}} \frac{w_l}{W_{\mathrm{h}(l)}} \right) \cdot \left\{ \frac{\left[W_{\mathrm{t}(l)_o} \cdot \exp(-c_l / \theta_l) \cdot W_{\mathrm{h}(l)_d} \right]^{\delta_l}}{\sum_{l' \in \Omega_i} \left[W_{\mathrm{t}(l')_o} \cdot \exp(-c_{l'} / \theta_{l'}) \cdot W_{\mathrm{h}(l')_d} \right]^{\delta_{l'}}} \right\} \quad \forall k \in \mathrm{K}_{\mathrm{od}} \tag{6.17}$$

This means the computation of route choice probabilities can be made as a function of links and nodes quantities, without need of path enumeration.

6.2.2 Specification of mixing components algorithm without explicit enumeration

As widely described in Chapter 5, particularly in Section 5.3.1, the CoNL mixing components specification represents the main step to operationalize the model. For this purpose, Section 5.4.1 proposes a recursive algorithm able to perform it, in a double version. In both versions, indicating with *i* the generic iteration, the procedure explores the network, defining a set Ω_i of links representing the nests of the generic component (also defined network level) *i*. Then, in the explicit enumeration procedure, it is possible to match each path with the nests (that, in turn, are links included in the path). The first version, i.e. the double-step version, explores the network for each o-d pair, thanks to the fact that, once defined the choice set K_{od}, the procedure deduces a set $\mathrm{L}_{od}^{eff\_o}$ of efficient links (with reference to the origin *o*) within which performing the procedure for searching the network levels. This means performing n_C^2 times the procedure, but also that each one of them the procedure considers also the links within $\mathrm{L}_{od}^{eff\_o}$ (see Section 5.4.1.1). The second version, i.e. the single-step version, instead, performs n_C network explorations, but it does not restrict the attention to the links belonging to $\mathrm{L}_{od}^{eff\_o}$

(see Section 5.4.1.2). In an explicit enumeration procedure, it is possible to perform the single-step version, obtaining a set of levels Ω_i, but successively matching their links with the paths belonging to K_{od}. This leads to have several empty nests (see the examples in Figure 5.7).
The first problem to deal with an implicit enumeration procedure refers the possibility to adapt the procedures described in Section 5.4.1 without possibility to know the set K_{od}. In fact, in an implicit enumeration procedure there is no possibility nor to deduce $L_{od}^{eff\_o}$ in the same way (i.e., starting from a choice set K_{od}), neither to successively match paths and nests. However, the set exhaustive set of the efficient links is available downstream having performed the minimum cost path algorithm. Thus, the procedure must consider the set $L_{od}^{eff\_od}$ of all the efficient links with reference both to the origin *o* and to the destination *d*. This means the implicit enumeration procedure restricts the set of available links. Furthermore, it is not possible to work with a single-step procedure, because the latter is based on the successive paths-nests matching.
However, the results on toy networks applications in Section 6.2.5 show how the restriction can be not so relevant in many cases. Furthermore, the real network application of the explicit CoNL route choice model described in Section 6.4 confirms that the double-step procedure is by far the more computational efficient procedure in real-world cases.
The second problem to deal is the possibility to implement the check for MNL levels. In fact, as described in Section 5.4.1, this is performed to avoid levels that does not contribute to reproduce correlations. An MNL level occurs in two ways: one alternative per nest or full set of alternatives in a nest. It is possible to check it thanks to the Dial weights computation. In fact, a nest *l* containing one a only alternative verifies, simultaneously, the two properties:

$$W_{\mathrm{t}(l)} = \exp(-C_{\mathrm{o\text{-}t}(l)} / \theta) \tag{6.18}$$

$$W_{\mathrm{h}(l)} = \exp(-C_{\mathrm{h}(l)-\mathrm{d}} / \theta) \tag{6.19}$$

The (6.18) derives from the consideration that if one path includes node *n*, then only one path appears in the log-sums within (6.3). In the case the (6.18) and (6.19) is verified for all the links belonging to Ω_i, the level is deleted from the set of levels or, equivalently, the procedure assigns null CoNL weight to it. The check for a unique full nest, instead, is easily performed by considering that the only way to obtain it is to have a one-nest level. In fact, it is not possible, by construction, to have an empty nest when using the double-step procedure.
The two checks mentioned must be performed when building the mixing components specification.

6.2.3 SNL CoNL algorithm

The recursive equation (6.17) contains terms that are implicitly computable, thanks to the (2.82), (6.5) and (6.8). The mixing components identification procedure described in Section 5.4.1 must be preliminarily applied for identifying the network's levels.
Processing each levels and each link of the generic level, (6.8) must be computed as a function of Dial node weights. This implies, for each link, a forward and a backward exploration respectively from the head and the tail of the processed link.
In terms of flows, being $d_{(o\text{-}d)}$ the generic (*o-d*) matrix entry, a final link flow can be expressed as:

$$\begin{aligned} f_l &= \sum_{\text{o-d}} f_l^{\text{o-d}} = \sum_{o-d} d_{\text{o-d}} \cdot p(l / \text{o-d}) = \\ &= \sum_{\text{o-d}} d_{\text{o-d}} \cdot [\sum_{i \in I} w^i \cdot p(l^i / \text{o-d})] = \\ &= \sum_{\text{o-d}} \sum_{i \in I} w^i \cdot d_{\text{o-d}} \cdot p(l^i / \text{o-d}) = \sum_{\text{o-d}} \sum_{i \in I} f_l^{i,\text{o-d,pre-load}} \end{aligned} \tag{6.20}$$

Wherein p(l/o-d) represents the assignment map element for link l, that is obviously a function of all the CoNL link choice probabilities.

Each nest $l \in \Omega_i$ with its probability p(l/o-d), defines a portion of total demand $d_{\text{o-d}}$ to be assigned on the network, passing through the link l. Thus, each $f_l^{\text{o-d}}$ represents also a flow to be distributed to all the paths connecting (o-d) that cross the link l. Thus, it is intuitive to build a procedure processing each level of the network (i^{th} NL mixing components of the CoNL), each link l (nesting group of the generic NL) and assigning a flow $f_l^{i,\text{o-d,pre-load}}$ to: the link l, all links l' belonging to paths connecting origin o and the node t(l) and all the links l' belonging to the paths connecting the node h(l) with the destination d. This assignment can be easily performed exploiting the Dial's procedure, respectively, from o to t(l) and from h(l) to d, while processing each link of each network's level.

These considerations can be summarized in the algorithm shown below. Defining:

- →**Dialweights**$^{\text{o-d}}_{n1(n2)}$ a sub-routine that computes the Dial weight, on the efficient sub-network with reference to the o-d, for the node $n2$, starting from node $n1$;
- →**Dialflows**$^{\text{o-d}}_{n1(n2)}$ a sub-routine that assigns the $f_l^{i,\text{o-d,pre-load}}$ as defined in (6.20), to all the links l' of the efficient sub-network with reference to the o-d, from node $n1$ to node $n2$;

Defining a coefficient of variation cv, the algorithm proceeds as shown in Table 6.1. The algorithm allows to compute link flows consistent with the CoNL route choice model (Papola et al., 2018), with the only adding restriction of the double Dial efficiency. Several tests will be set out in the next paragraphs.

SNL CoNL algorithm

for each o-d

$$\theta = \frac{\sqrt{6}\cdot cv \cdot C_{\text{o-d,min}}}{\pi}$$

for each $i \in \mathrm{I}$

for each $l \in \Omega_i$

$$\delta_l = \begin{cases} \max\left\{\delta_{\min}; \sqrt{1-\frac{c_l}{C_{od,\min}}\cdot\frac{\sum_{i\in \mathrm{I}} c^i_{l,mean}}{\sum_{i\in \mathrm{I}^l} c^i_{l,mean}}}\right\} & \text{if} \quad 1-\frac{c_l}{C_{od,\min}}\cdot\frac{\sum_{i\in \mathrm{I}} c^i_{l,mean}}{\sum_{i\in \mathrm{I}^l} c^i_{l,mean}} > 0 \\ \delta_{\min} & \text{otherwise} \end{cases}$$

$$\theta_l = \delta_l \cdot \theta$$

→ **Dialweights**$^{\text{o-d}}{}_{\text{o}(t(l))}$ **Output**: $W_{\mathrm{t}(l)_o}$

→ **Dialweights**$^{\text{o-d}}{}_{\text{h}(l)(\text{d})}$ **Output**: $W_{\mathrm{h}(l)_d}$

$$p^i(l^i) = \left[W_{\mathrm{t}(l)_o}\cdot \exp(-c_l/\theta_l)\cdot W_{\mathrm{h}(l)_d}\right]^{\delta_l}$$

$$den_{p^i(l^i)} = den_{p^i(l^i)} + p^i(l^i)$$

next

for each $l \in \Omega_i$

$$p^i(l^i) = \frac{p^i(l^i)}{den_{p^i(l^i)}}$$

→ **Dialflows**$^{\text{o-d}}{}_{\text{o}(t(l))}$

→ **Dialflows**$^{\text{o-d}}{}_{\text{h}(l)(\text{d})}$

next

next

next

Table 6.1: SNL CoNL algorithm without explicit enumeration of routes.

6.2.4 Simplified SNL CoNL algorithm

The procedure described in the previous sections, for each o-d pair substantially depends on the Dial weights and flows computation, for each link *l* of each network level *i*. Albeit it is effective for reproducing CoNL route choice flows, it requires to perform twice the subroutines for weights and twice the sub-routine for link flows. This happens because a different variance parameter θ_l for each link determines different Dial weights value, according to (6.6) and (6.7). The different values of θ_l are due, in turn, to the different nesting parameters value δ_l, because of the computation of (5.15). Accepting the simplification of keeping fixed the nesting parameters value makes not necessary to perform the two procedures for weights and flows twice for each link of each level.
In fact, it is possible to compute once the Dial weights prior to the algorithm (the preliminary procedure already did it). Furthermore, this allows to update in memory a temporary matrix, whose generic entry is the $f_l^{i,\text{o-d,pre-load}}$, loading the pairs o-t(*l*) and h(*l*)-d. However, it has been noted that the dimension of the matrix is not n_C^2, but n_N^2.
Defining:

- OD_{temp} as the matrix whose entries are all the node to node temporary flows;
- → **Dialflows$^{\text{o-d}}_{n1}$(•)** a sub-routine that assigns the $f_{l'}^{i,\text{o-d,pre-load}}$ as defined in (6.20), to all the links *l'* of the efficient sub-network with reference to the o-d, from node *n*1 to all nodes *n*2, in a single-step exploration;

The algorithm can be simplified as shown in Table 6.2 and the procedure reduces significantly the computation time. Obviously, the resulting link flows are biased with reference to the link flows resulting from the rigorous version of the algorithm. However, several tests will be set out in the next sub-section, to compare the performance of the simplified algorithm with reference to the first one, showing that the last version is able to significantly reduce the computation time, introducing a not significant bias in the link flows computation.

Simplified SNL CoNL algorithm

for each o-d

$$\theta = \frac{\sqrt{6}\cdot cv\cdot C_{\text{o-d,min}}}{\pi}$$

$$\delta_l = \delta_{\min}$$

$$\theta_l = \delta_l \cdot \theta$$

for each $i\in \mathrm{I}$

for each $l\in\Omega_i$

$$p^i(l^i) = \left[W_{\mathrm{t}(l)_o}\cdot \exp(-c_l/\theta_l)\cdot W_{\mathrm{h}(l)_d}\right]^{\delta_l}$$

$$den_{p^i(l^i)} = den_{p^i(l^i)} + p^i(l^i)$$

next

for each $l\in\Omega_i$

$$p^i(l^i) = \frac{p^i(l^i)}{den_{p^i(l^i)}}$$

$$f_l^{i,\text{o-d,pre-load}} = w^i\cdot d_{\text{o-d}}\cdot p(l^i/\text{o-d})$$

$$f_l = f_l + f_l^{i,\text{o-d,pre-load}}$$

$$OD_{o-\mathrm{t}(l)} = OD_{\mathrm{h}(l)-d} = f_l^{i,\text{o-d,pre-load}}$$

next

next

for each $n_1, n_2\in \mathrm{N}$

→ **Dialflows$^{\text{o-d}}_{n1}$(•)**

next

next

Table 6.2: Simplified SNL CoNL algorithm without explicit enumeration of routes.

6.2.5 Experimental results

This sub-section shows the resulting link flows of the SNL CoNL with the two algorithms described in Section 6.2.3 and 6.2.4. Some results are shown on the following three test networks: the Braess network, the mesh network with bypass (see Section 5.5.3) and the Sioux-Falls network.

A comparison between the SNL Probit flows and the SNL CoNL is here proposed. Particularly, giving the implicit enumeration CoNL identical results of the explicit enumeration CoNL, the comparison relates the two algorithms proposed in the previous section (see Table 6.1 and Table 6.2).

The Braess network is tested with reference to the o-d pair **1-4**, with a total demand equal to 1000. For helping the comprehension of the flows differences entity, an average absolute value of errors indicator (AAVE) is computed, with reference to the SNL Probit flows. The results, reported in Table 6.3, show negligible differences between the first and the second algorithm for SNL CoNL. In fact, when the value of δ_{MIN} and δ's, respectively, in the first general algorithm and in the second simplified algorithm are set to be equal, the results are practically identical. Thus, with reference to SNL Probit, the difference in terms of flows is very similar, i.e. about 33 veic/h, representing the 3.3% of the total demand.

The second network's results are reported in Table 6.4 with reference to the o-d pair **1-12**. The simplified CoNL gives substantially different results when δ_{MIN} is set to be small. A deepening about the δ_{MIN} value and its meaning is widely discussed in the following section. Setting δ_{MIN} on a value of 0.3/0.4 makes the results very similar between the two algorithms and very similar to SNL Probit too (2.4 and 2.5 % of the total o-d demand, respectively, for the first and the second algorithm).

The Sioux-Falls network shows another interesting aspect. In fact, as shown in Section 5.5.5, the o-d pair **1-15** has 16 efficient routes. The Table 5.6 reports the route costs, that are very different from each other. For working on an exhaustive choice-set, the explicit enumeration CoNL needs a big number of Monte-Carlo iterations. The implicit procedure, instead, has the advantage of being exhaustive on the efficient sub-network. In Table 6.6, the explicit and implicit enumeration procedures are contrasted. In the explicit case, the computation time is shown as a function of the Monte Carlo iterations that are required for obtaining an exhaustive choice-set with reference to the o-d pair **1-15**. It can be seen that the implicit enumeration procedure gives the advantage of being an exhaustive procedure on the efficient sub-network, with a computation time significantly lower. The computation times reported in Table 6.6 appear as non-linearly increasing with the number of iterations. This is essentially due to the check on the unicity of the random generated path, that is performed at each iteration. In fact, the when the number of generated paths rises, then the number that the procedure must check rises.

Links		SNL Probit	SNL CoNL - implicit enumeration (δ_l computed with (6.24))			SNL CoNL - implicit enumeration (δ_l fixed)					
tail	**head**	$\alpha = cv^2 * C_{od,min}$	$\delta_{min} \leq 0.3$	$\delta_{min}=0.35$	$\delta_{min}=0.4$	$\delta \to 0$	$\delta=0.1$	$\delta=0.2$	$\delta=0.3$	$\delta=0.35$	$\delta=0.4$
1	2	689	639	640	642	625	629	634	638	640	642
1	3	311	361	360	358	375	371	366	362	360	358
2	3	311	279	280	284	271	269	267	276	278	284
2	4	377	361	360	358	375	371	366	362	360	358
3	4	623	639	640	642	625	629	634	638	640	642
AAVE SNL Probit=			**33**	**33**	**32**	**34**	**35**	**35**	**33**	**33**	**32**

Table 6.3: Braess network, o-d pair **1-4** *– Comparison of link flows.*

Links		SNL Probit	SNL CoNL - implicit enumeration (δ_l computed with (6.24))					SNL CoNL - implicit enumeration (δ_l fixed)				
tail	head	$\alpha = cv^2 * C_{od,min}$	δ_{min}=0	δ_{min}=0.1	δ_{min}=0.2	δ_{min}=0.3	δ_{min}=0.4	δ_{min}=0	δ_{min}=0.1	δ=0.2	δ=0.3	δ=0.4
1	2	478	412	413	416	424	437	384	392	410	427	444
1	4	323	376	380	383	381	377	365	369	366	364	363
1	9	199	212	208	200	196	186	252	239	224	209	193
2	3	239	188	190	197	208	225	171	176	195	214	229
2	5	239	224	223	219	215	212	212	217	214	213	215
3	5	121	59	61	69	82	100	19	27	57	83	102
3	6	124	129	129	128	127	125	152	148	138	131	127
4	5	214	232	235	238	238	235	212	218	219	220	222
4	7	109	144	144	145	143	142	152	151	147	144	140
5	6	304	257	261	268	275	279	222	231	245	258	270
5	8	264	258	258	259	259	269	222	231	245	258	270
6	9	429	386	390	396	402	404	374	379	383	389	397
7	8	109	144	144	145	143	142	152	151	147	144	140
8	9	373	402	402	403	402	410	374	382	393	402	410
9	10	655	776	771	747	709	675	973	941	825	742	690
9	11	345	224	229	253	291	325	27	59	175	258	310
10	12	655	776	771	747	709	675	973	941	825	742	690
11	12	345	224	229	253	291	325	27	59	175	258	310
AAVE SNL Probit=			**53**	**52**	**45**	**34**	**24**	**109**	**98**	**66**	**41**	**25**

Table 6.4: Mesh grid network with bypass, o-d pair **1-12** *– Comparison of link flows.*

Links		SNL Probit	SNL CoNL - implicit enumeration (δ_l computed with (6.24))					SNL CoNL - implicit enumeration (δ_l fixed)				
tail	head	$\alpha = cv^2 * C_{od,min}$	$\delta_{min}=0$	$\delta_{min}=0.1$	$\delta_{min}=0.2$	$\delta_{min}=0.3$	$\delta_{min}=0.4$	$\delta_{min}=0$	$\delta_{min}=0.1$	$\delta=0.2$	$\delta=0.3$	$\delta=0.4$
1	2	160	160	157	151	144	137	184	178	169	158	146
1	3	840	840	843	849	856	863	816	822	831	842	854
2	6	160	160	157	151	144	137	184	178	169	158	146
3	4	320	338	338	340	342	343	352	351	347	347	348
3	12	520	502	505	509	514	520	463	471	484	496	507
4	5	124	102	105	114	119	121	151	144	136	132	130
4	11	196	236	233	226	223	222	201	206	211	214	218
5	6	20	0	0	0	0	0	0	0	0	0	0
5	9	118	102	105	114	119	121	151	144	136	132	130
6	5	13	0	0	0	0	0	0	0	0	0	0
6	8	167	160	157	151	144	137	184	178	169	158	146
7	18	49	0	0	0	0	0	0	0	0	0	0
8	7	49	0	0	0	0	0	0	0	0	0	0
8	9	3	2	2	2	2	3	2	2	2	2	3
8	16	114	158	155	149	142	134	182	176	167	155	143
9	10	121	104	108	117	121	124	153	147	138	134	132
10	15	255	188	192	210	226	238	189	194	213	232	246
10	17	0	5	5	5	5	7	4	4	4	4	6
11	10	118	88	89	98	110	121	40	51	79	102	120
11	14	330	363	359	341	325	315	341	344	330	317	308
12	11	252	215	215	213	213	214	180	188	198	205	210
12	13	268	287	290	296	301	306	284	283	286	291	296
13	24	268	287	290	296	301	306	284	283	286	291	296
14	15	343	386	382	366	354	347	348	351	341	335	334
14	23	7	0	0	0	0	0	0	0	0	0	0
16	10	16	0	0	0	0	0	0	0	0	0	0
16	17	108	158	155	149	142	134	182	176	167	155	143
17	19	108	163	159	154	147	141	186	180	171	160	149
18	16	10	0	0	0	0	0	0	0	0	0	0
18	20	39	0	0	0	0	0	0	0	0	0	0
19	15	134	163	159	154	147	141	186	180	171	160	149
20	19	26	0	0	0	0	0	0	0	0	0	0

Links		SNL Probit	SNL CoNL - implicit enumeration (δ_l computed with (6.24))					SNL CoNL - implicit enumeration (δ_l fixed)				
tail	**head**	$\alpha = cv^2 * C_{od,min}$	$\delta_{min}=0$	$\delta_{min}=0.1$	$\delta_{min}=0.2$	$\delta_{min}=0.3$	$\delta_{min}=0.4$	$\delta_{min}=0$	$\delta_{min}=0.1$	$\delta=0.2$	$\delta=0.3$	$\delta=0.4$
20	21	3	0	0	0	0	0	0	0	0	0	0
20	22	10	0	0	0	0	0	0	0	0	0	0
21	20	0	0	0	0	0	0	0	0	0	0	0
21	22	176	194	192	181	174	170	243	232	205	187	175
22	15	268	263	266	271	273	274	276	275	275	273	270
23	14	20	24	24	25	28	32	7	7	10	18	26
23	22	82	69	74	89	99	104	33	44	71	87	95
24	21	173	194	192	181	174	170	243	232	205	187	175
24	23	95	92	98	114	127	137	40	51	81	104	121
		AAVE SNL Probit=	**21**	**20**	**18**	**17**	**18**	**32**	**29**	**22**	**18**	**17**

Table 6.5: Sioux-Falls network, o-d pair **1-15** *– Comparison of link flows.*

Monte-Carlo # iterations	100	500	1000	5000	10000
Number of different efficient routes generated	12	13	13	15	16
Choice-set generation/Total time [sec]	0.71	2.03	4.1	28	86
Explicit SNL CoNL/Total time [sec]	2.3	4.1	6.2	40	106
Implicit SNL CoNL/Total time [sec]			0.4		

Table 6.6: Sioux-Falls network – Computation time.

6.3 An in depth analysis of CoNL route choice and Daganzo and Sheffi correlations

This subsection investigates the capability of the CoNL route choice model to reproduce Daganzo and Sheffi (1977) target correlations. The analysis is carried out contrasting the CoNL with the main route choice models that are capable to handle complex correlation patterns. Obviously, the first model proposed in this comparison is the Multinomial Probit (MNP; Daganzo and Sheffi, 1977; Yai et al., 1997). As recalled in Section 6.1.2, it represents the natural candidate to aim this target, since its underlying hypothesis of multivariate Normal distribution of the random residuals and the fact that the covariance matrix enters directly into the integral (2.15). The second model compared is the Link Nested Logit model (Vovsha and Bekhor, 1998). The LNL is surely the first GEV model proposed with the purpose of reproducing flexible covariance matrices effects on probabilities, even if it does not allow to compute covariances in closed form. Because of that, the investigation of its real capability of reproducing Daganzo and Sheffi correlations represents, to the best of my knowledge, an open question for route choice modelling. Three main LNL formulations are investigated, namely the null nesting parameters formulation (Vovsha and Bekhor, 1998), the arithmetic mean in inclusion's parameters formulation (Prashker and Bekhor, 1998) and a geometric mean in inclusion's parameters formulation (Marzano, 2006). The latter has not been fully explored yet in its potential. About the first one, instead, the further aim of this analysis is testing the capability of the LNL with a fixed value of the nesting parameters, since a null value for them actually represents an undesirable condition in many cases (see the properties of the nest probabilities described in Section 2.2.1.1). Another interesting particularization of the CNL is the Pair Combinatorial model (Chu, 1989; Prashker and Bekhor, 1998; Gliebe et al, 1999), because it is built on the idea of specifying the model with a nest for each pair of alternatives. This concept is apparently very suitable in contexts with expected high correlation effects like route choice. The LNL and PCL correlations are computed using both the theoretical formulation of the covariances between two paths utilities, i.e. the standard double integral one, and the (Marzano et al., 2013) and (Marzano, 2014) procedure, i.e. the more efficient equivalent single-integral formulation. The integral simulation is based on a Matlab R2017b code, based on a function of Matlab library implementing the method of global adaptive quadrature (Shampine, 2008). Finally, the CoNL route choice is tested also by implementing some minor changes in the models parameters computation. Particularly, it has been reminded that the application of CoNL route choice is founded on two main steps: the network levels identification and the nesting parameters computation. The last one is essentially based on (5.11) and (5.15), wherein the average value of the link costs of each level plays a central role.

The first small change lies in numerator and denominator computation of (5.11), wherein, the $c_{l,mean}$ value is computed in a weighted way, on the basis of how many times a generic link l appears in all the levels, given an o-d pair. In other words, defining n_L^i as the number of links constituting a generic network level i, while n_i^l represents the number of levels wherein the link l appears, the (5.11) becomes:

$$c_{l,mean}{}^{i} = \frac{1}{n_{\mathrm{L}}^{i}} \sum_{l \in \Omega_i} \frac{c_l}{n_i^l} \qquad \forall i \in \mathrm{I} \tag{6.21}$$

The second small change consists of testing the capability of achieving Daganzo and Sheffi correlations by substituting the average value computed by (6.21) with the maximum and the minimum value of the costs.

The comparison is carried out on some of the toy networks described in Section 5.5. For the first two networks, the results are directly discussed and shown in terms of correlation matrices. For the other networks, given their size and complexity, some appropriate synthetic indicators are adopted.

The Daganzo and Sheffi (1977) network depicted in Figure 5.5, represents the simplest correlations scenario, given by a typical 3x3 block-diagonal (i.e. Nested structured) covariance matrix. The only value of interest is represented by the correlation between paths 12_a3 and 12_b3. As shown in Section 5.5.1, the CoNL collapses to a simple NL, notoriously capable to handle this correlations scenario. Thus, the CoNL is perfectly capable to reproduce it with an opportune setting of the parameter δ_{MIN} (see Section 5.3.2). In this case, clearly, the opportune value is 0. In this way, except for some numeric approximation, all the models are capable to reproduce the target correlation value, whatever value of h (see Section 5.5.1 for details).

The Braess network depicted in Figure 5.8, instead, represents a typical Cross Nested structured correlation matrix. The case here tested is the one characterized by a ratio a/b=0.8, i.e. the same one examined in Section 5.5.2.

The LNL with the first specification and the CoNL route choice perfectly reproduce the target correlation matrix. It has been noted that LNL with null nesting parameters is impossible to be simulated, so a simulation with a bigger value of δs is required. The minimum value for δs allowing for integral computation is 0.11, so the results shown in Table 6.7 actually shows a biased value. However, the Braess case surely represents a boundary case well analysed in Papola (2004) and Abbé et al. (2007) and the Papola's conjecture comes to the aid, giving a perfect target value for the correlations (0.44). Thus, there is a great confidence that the actual result of LNL correlation value is the right one.

The network depicted in Figure 5.10 surely represents a very challenging correlation scenario, since the presence of the long bypass (link *1-9*). The choice set of routes connecting the o-d pair **1-12** consists of 18 routes with similar route costs. Another very interesting network is the Sioux-Falls network, already examined in Section 5.5.5., with reference to the o-d pair **1-15** and an efficient choice-set of 16 paths. In Table 6.8 and Table 6.9 the ID of routes are summarized, in order to briefly recall them in the target correlation matrix depicted in
Table 6.10 and
Table 6.11. For immediately evaluating the distance between the values of correlations reproduced by the models and the target correlation matrix, a mean square error indicator is used. The mean square error (MSE) indicator is computed both as a function of the correlations among the single alternatives and as a function of the correlations among differences of route utilities. The results are shown in Table 6.12 and Table 6.13, where the best values of the MSE as a function of δ_{MIN} are bolded. The references *FCM* and *RCM* represent, respectively, the full correlation matrix, related to the space of the alternatives, and the reduced

matrix, related to the space of the utility differences. Finally, the route choice probabilities MSE is proposed in the third column, with reference to the Multinomial Probit probabilities.

In the first network, the best values in terms of probabilities are given by the LNL, but with a different value for the δ's. In the second network, instead, the modified CoNL provides the best values. Particularly, increasing the value of δ's gives a positive effect on route choice probabilities, confirming the intuition for which null nesting parameters potentially have the unpleasant effect of giving too deterministic probabilities within the nests. Furthermore, the mathematical properties of the G.E.V. models explains this phenomenon, because the density function of a M.E.V. distribution does not depend exclusively on the first and second order moments. In fact, in a CNL generating function, a lot of terms appear (inclusion and nesting parameters), and often a one-to-one relationship between a given correlation matrix and a vector of structural parameters cannot be identified. The same concept seems to apply to CoNL, that is characterized by a cumulative function wherein several parameters appear (CoNL weights and nesting parameters). Thus, reproducing a target correlation matrix is not synonymous with reproducing target probabilities, differently than MNP model. This effect is very clear in the PCL and the other LNL formulations case. The PCL, in particular, gives the most biased values of correlations in both cases, but not the worst in terms of probabilities. In fact, Table 6.13 shows how the *FCM* and *RCM* potentially exhibit different trends. This actually does not represent a great improvement in the comprehension of the phenomenon, because of the already discussed relationship between covariances and probabilities in the G.E.V. models. However, it stresses the concept that reproducing a target correlation matrix helps but it can be not sufficient to guarantee a target probability vector.

Finally, the results show that CoNL route choice model using (6.21) for nesting parameters calculation seems to work better than the original one, both in terms of correlations and in terms of probabilities. Particularly, substituting $c_{l,min}$ in (6.21) gives, in these cases, the best results, even better than those published in Papola et al. (2018).

Target (Daganzo and Sheffi, 1977)

Paths	*1234*	*124*	*134*
1234	1,00	0,44	0,44
124	0,44	1,00	0,00
134	0,44	0,00	1,00

LNL (Vovsha and Bekhor, 1998; δ_{min}=0.11)

Paths	*1234*	*124*	*134*
1234	1,00	0,40	0,40
124	0,40	1,00	0,00
134	0,40	0,00	1,00

LNL (Prashker and Bekhor, 1998)

Paths	*1234*	*124*	*134*
1234	1,00	0,29	0,29
124	0,29	1,00	0,00
134	0,29	0,00	1,00

LNL (Marzano, 2006)

Paths	*1234*	*124*	*134*
1234	1,00	0,36	0,36
124	0,36	1,00	0,00
134	0,36	0,00	1,00

PCL (Gliebe et al., 1999)

Paths	*1234*	*124*	*134*
1234	1,00	0,35	0,35
124	0,36	1,00	0,24
134	0,36	0,00	1,00

CoNL route choice (δ_{MIN}<0.3)

Paths	*1234*	*124*	*134*
1234	1,00	0,44	0,44
124	0,44	1,00	0,00
134	0,44	0,00	1,00

CoNL route choice (δ_{MIN}=0.4)

Paths	*1234*	*124*	*134*
1234	1,00	0,42	0,42
124	0,42	1,00	0,00
134	0,42	0,00	1,00

Table 6.7: Braess network – Target correlation matrix and reproduced correlation values for the compared route choice models.

Routes	ID	Routes	ID
1 2 3 5 6 9 10 12	k_1	*1 2 5 8 9 11 12*	k_{10}
1 2 3 5 6 9 11 12	k_2	*1 4 5 6 9 10 12*	k_{11}
1 2 3 5 8 9 10 12	k_3	*1 4 5 6 9 11 12*	k_{12}
1 2 3 5 8 9 11	k_4	*1 4 5 8 9 10 12*	k_{13}
1 2 3 6 9 10 12	k_5	*1 4 5 8 9 11 12*	k_{14}
1 2 3 6 9 11 12	k_6	*1 4 7 8 9 10 12*	k_{15}
1 2 5 6 9 10 12	k_7	*1 4 7 8 9 11 12*	k_{16}
1 2 5 6 9 11 12	k_8	*1 9 10 12*	k_{17}
1 2 5 8 9 10 12	k_9	*1 9 11 12*	k_{18}

*Table 6.8: Mesh network with bypass, o-d pair **1-12**: ID of the routes.*

Routes	ID	Routes	ID
1 2 6 8 9 10 15	k_1	*1 3 4 11 10 17 19 15*	k_9
1 2 6 8 9 10 17 19 15	k_2	*1 3 4 11 14 15*	k_{10}
1 2 6 8 16 17 19 15	k_3	*1 3 12 11 10 15*	k_{11}
1 3 4 5 6 8 9 10 15	k_4	*1 3 12 11 10 17 19 15*	k_{12}
1 3 4 5 6 8 16 17 19	k_5	*1 3 12 11 14 15*	k_{13}
1 3 4 5 9 10 15	k_6	*1 3 12 13 24 21 22 15*	k_{14}
1 3 4 5 9 10 17 19 15	k_7	*1 3 12 13 24 23 14 15*	k_{15}
1 3 4 11 10 15	k_8	*1 3 12 13 24 23 22 15*	k_{16}

*Table 6.9: Sioux-Falls network, o-d pair **1-15**: ID of the routes.*

Paths	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}	k_{13}	k_{14}	k_{15}	k_{16}	k_{17}	k_{18}
k_1	1,00	0,88	0,60	0,49	0,77	0,65	0,72	0,60	0,31	0,20	0,51	0,40	0,10	0,00	0,10	0,00	0,10	0,00
k_2	0,88	1,00	0,49	0,62	0,65	0,78	0,60	0,73	0,20	0,34	0,40	0,54	0,00	0,15	0,00	0,15	0,00	0,15
k_3	0,60	0,49	1,00	0,88	0,46	0,35	0,31	0,20	0,72	0,60	0,10	0,00	0,51	0,40	0,31	0,20	0,10	0,00
k_4	0,49	0,62	0,88	1,00	0,35	0,49	0,20	0,34	0,60	0,73	0,00	0,15	0,40	0,54	0,20	0,34	0,00	0,15
k_5	0,77	0,65	0,46	0,35	1,00	0,87	0,63	0,51	0,32	0,21	0,42	0,31	0,11	0,00	0,11	0,00	0,11	0,00
k_6	0,65	0,78	0,35	0,49	0,87	1,00	0,51	0,65	0,21	0,35	0,31	0,45	0,00	0,15	0,00	0,15	0,00	0,15
k_7	0,72	0,60	0,31	0,20	0,63	0,51	1,00	0,87	0,58	0,46	0,53	0,41	0,11	0,00	0,11	0,00	0,11	0,00
k_8	0,60	0,73	0,20	0,34	0,51	0,65	0,87	1,00	0,46	0,60	0,41	0,55	0,00	0,15	0,00	0,15	0,00	0,15
k_9	0,31	0,20	0,72	0,60	0,32	0,21	0,58	0,46	1,00	0,87	0,11	0,00	0,53	0,41	0,32	0,21	0,11	0,00
k_{10}	0,20	0,34	0,60	0,73	0,21	0,35	0,46	0,60	0,87	1,00	0,00	0,15	0,41	0,55	0,21	0,35	0,00	0,15
k_{11}	0,51	0,40	0,10	0,00	0,42	0,31	0,53	0,41	0,11	0,00	1,00	0,87	0,58	0,46	0,37	0,26	0,11	0,00
k_{12}	0,40	0,54	0,00	0,15	0,31	0,45	0,41	0,55	0,00	0,15	0,87	1,00	0,46	0,60	0,26	0,40	0,00	0,15
k_{13}	0,10	0,00	0,51	0,40	0,11	0,00	0,11	0,00	0,53	0,41	0,58	0,46	1,00	0,87	0,58	0,46	0,11	0,00
k_{14}	0,00	0,15	0,40	0,54	0,00	0,15	0,00	0,15	0,41	0,55	0,46	0,60	0,87	1,00	0,46	0,60	0,00	0,15
k_{15}	0,10	0,00	0,31	0,20	0,11	0,00	0,11	0,00	0,32	0,21	0,37	0,26	0,58	0,46	1,00	0,87	0,11	0,00
k_{16}	0,00	0,15	0,20	0,34	0,00	0,15	0,00	0,15	0,21	0,35	0,26	0,40	0,46	0,60	0,87	1,00	0,00	0,15
k_{17}	0,10	0,00	0,10	0,00	0,11	0,00	0,11	0,00	0,11	0,00	0,11	0,00	0,11	0,00	0,11	0,00	1,00	0,87
k_{18}	0,00	0,15	0,00	0,15	0,00	0,15	0,00	0,15	0,00	0,15	0,00	0,15	0,00	0,15	0,00	0,15	0,87	1,00

Table 6.10: Mesh network with bypass: target Daganzo and Sheffi correlation matrix for routes connecting the o-d pair **1-12**.

Paths	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}	k_{13}	k_{14}	k_{15}	k_{16}
k_1	1,00	0,74	0,46	0,63	0,07	0,32	0,10	0,21	0,00	0,00	0,21	0,00	0,00	0,00	0,00	0,00
k_2	0,74	1,00	0,58	0,41	0,21	0,10	0,46	0,00	0,37	0,00	0,00	0,37	0,00	0,00	0,00	0,00
k_3	0,46	0,58	1,00	0,07	0,53	0,00	0,18	0,00	0,18	0,00	0,00	0,18	0,00	0,00	0,00	0,00
k_4	0,63	0,41	0,07	1,00	0,51	0,66	0,39	0,47	0,24	0,28	0,34	0,12	0,14	0,14	0,13	0,14
k_5	0,07	0,21	0,53	0,51	1,00	0,39	0,51	0,30	0,43	0,32	0,15	0,30	0,16	0,16	0,15	0,15
k_6	0,32	0,10	0,00	0,66	0,39	1,00	0,66	0,57	0,29	0,34	0,41	0,14	0,17	0,17	0,16	0,17
k_7	0,10	0,46	0,18	0,39	0,51	0,66	1,00	0,29	0,67	0,30	0,14	0,54	0,15	0,15	0,14	0,15
k_8	0,21	0,00	0,00	0,47	0,30	0,57	0,29	1,00	0,67	0,58	0,60	0,32	0,17	0,17	0,16	0,16
k_9	0,00	0,37	0,18	0,24	0,43	0,29	0,67	0,67	1,00	0,52	0,32	0,69	0,15	0,15	0,14	0,14
k_{10}	0,00	0,00	0,00	0,28	0,32	0,34	0,30	0,58	0,52	1,00	0,17	0,15	0,57	0,17	0,37	0,17
k_{11}	0,21	0,00	0,00	0,34	0,15	0,41	0,14	0,60	0,32	0,17	1,00	0,67	0,58	0,33	0,31	0,33
k_{12}	0,00	0,37	0,18	0,12	0,30	0,14	0,54	0,32	0,69	0,15	0,67	1,00	0,52	0,29	0,28	0,29
k_{13}	0,00	0,00	0,00	0,14	0,16	0,17	0,15	0,17	0,15	0,57	0,58	0,52	1,00	0,35	0,53	0,34
k_{14}	0,00	0,00	0,00	0,14	0,16	0,17	0,15	0,17	0,15	0,17	0,33	0,29	0,35	1,00	0,61	0,77
k_{15}	0,00	0,00	0,00	0,13	0,15	0,16	0,14	0,16	0,14	0,37	0,31	0,28	0,53	0,61	1,00	0,68
k_{16}	0,00	0,00	0,00	0,14	0,15	0,17	0,15	0,16	0,14	0,17	0,33	0,29	0,34	0,77	0,68	1,00

Table 6.11: Sioux-Falls network: target Daganzo and Sheffi correlation matrix for routes connecting the o-d pair **1-15**.

	CoNL (Papola et al., 2018)			CoNL modified with $c_{l,mean}$			CoNL modified with $c_{l,min}$			CoNL modified with $c_{l,max}$		
δ_{MIN}	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities
0	**3,71**	**2,22**	3,09	**1,52**	**0,48**	10,46	**1,50**	**0,47**	9,68	**2,67**	**1,10**	9,52
0.1	3,87	2,31	2,78	1,64	0,52	8,18	1,64	0,53	7,43	2,88	1,17	7,15
0.2	4,40	2,58	1,90	2,12	0,72	2,63	2,02	0,69	2,21	3,59	1,45	2,09
0.3	5,65	3,12	**0,58**	3,31	1,31	0,58	2,93	1,17	**0,47**	5,09	2,13	**0,58**
0.4	8,72	4,62	0,65	6,20	2,78	**0,55**	4,84	2,26	0,50	8,06	3,61	0,71

	LNL (Vovsha and Bekhor, 1998 with δ's=δ_{MIN})			LNL (Prashker and Bekhor, 1998)			LNL (Marzano, 2006)			PCL (Gliebe et al., 1999)		
δ_{MIN}	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities
0	**0,96**	**0,50**	28,18	**61,71**	**27,29**	7,75	**16,41**	**9,69**				
0.1	0,99	0,57	22,91	61,71	27,29	7,75	16,41	9,69				
0.2	1,20	0,92	7,70	61,74	27,34	7,68	16,41	9,69	**3,79**	124,98	76,23	**5,21**
0.3	2,03	1,82	1,42	61,79	27,51	7,60	16,41	9,69				
0.4	4,29	3,71	**0,35**	61,91	27,93	**7,45**	16,41	9,69				

Table 6.12: Mesh network with bypass, o-d pair 1-12 –MSE indicator for reproduced correlation values.

	CoNL (Papola et al., 2018)			CoNL modified with $c_{l,mean}$			CoNL modified with $c_{l,min}$			CoNL modified with $c_{l,max}$		
δ_{MIN}	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities
0	**6,01**	**3,56**	3,65	**2,96**	**1,87**	4,99	**4,96**	**2,52**	2,41	**6,26**	**4,06**	4,05
0.1	6,25	3,71	2,96	3,07	1,94	4,38	5,08	2,59	1,93	6,53	4,22	3,37
0.2	7,03	4,21	1,34	3,43	2,17	2,64	5,50	2,86	0,81	7,41	4,74	1,66
0.3	8,53	5,11	**0,71**	4,29	2,66	1,06	6,27	3,39	0,3	9,09	5,64	0,75
0.4	11,07	6,49	0,72	6,16	3,62	**0,4**	7,52	4,30	**0,26**	11,85	7,06	**0,73**

	LNL (Vovsha and Bekhor, 1998 with δ's=δ_{MIN})			LNL (Prashker and Bekhor, 1998)			LNL (Marzano, 2006)			PCL (Gliebe et al., 1999)		
δ_{MIN}	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities	FCM	RCM	Choice probabilities
0	3,00	**0,29**	11,67									
0.1	2,94	0,34	9,96									
0.2	2,74	0,58	4,78	**51,80**	**27,83**	**4,66**	**14,98**	**9,63**	**1,55**	**78,47**	**49,81**	**5,49**
0.3	**2,71**	1,15	1,73									
0.4	3,39	2,35	**0,59**									

Table 6.13: Sioux-Falls network, o-d pair 1-15 - MSE indicator for reproduced correlation values.

6.4 Experimental results on real data with unrestricted acyclic choice set

In Papola et al. (2018) the capability of CoNL route choice model to fit real observations is not analysed. For bridging this gap, this sub-section presents the results of the application of CoNL route choice on a set of real observed route choices. A dataset of 219 trajectories of drivers moving inside the Regione Campania network has been collected, with the aid of a smartphone application for Andoid systems called *Algoroute*[18]. The Regione Campania graph consists of 539.863 links and 244.019 nodes, subdivided in 1.112 zones (see Figure 6.1) with the aid of the software TransCad. After the map-matching and the data-cleaning operations, 197 observed routes have been considered in the analysis, each one in its time slot.

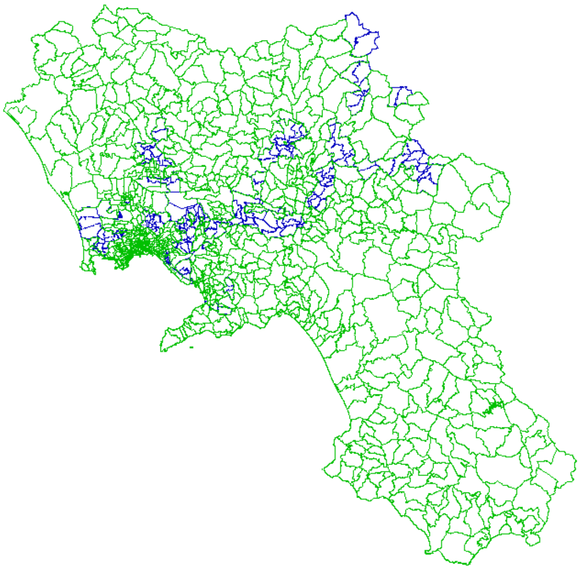


Figure 6.1: Regione Campania zoning (TransCad screenshot).

The CoNL route choice has been contrasted with Multinomial Logit, C-Logit and P.S.-Logit models, in terms of probabilities to reproduce the observed choices, through the log-likelihood computation with standard values for the parameters.

Obviously, working with an explicit enumeration model with a restricted choice set can give numerical problems in the likelihood computation. In fact, a route that is not efficient with reference to its origin o, with the current procedure, gives null probability to be chosen and, consequently, a null value for the likelihood and a negative infinitive value for the log-likelihood. Thus, using the same procedure for identifying the network levels, i.e. the efficient sub-network based procedure described in Section 5.4.1, the explicit allocation of routes to the links/nests belonging to routes has been modified. Practically, the routes are added a posteriori to the CoNL specification for each level. In the case no link/nest l of a generic level i belongs to a given non-efficient route k, the route k constitutes a single alternative-nest. For better clarifying the idea, consider the grid network and the choice set depicted in Figure 5.1. The considered levels of network are represented in Figure 5.4, but the routes at the lowest level of the trees representation are those of Figure 5.1. The resulting specification is depicted in Figure

[18] The Android application *Algoroute* has been developed by Inputspace S.r.l.., a spinoff of the Department of Civil, Architectural and Environmental Engineering of University of Naples Federico II.

6.2, wherein the whole exhaustive acyclic choice-set connecting the o-d pair **1-4** is allocated, at the lowest level of the representation, within the nests.

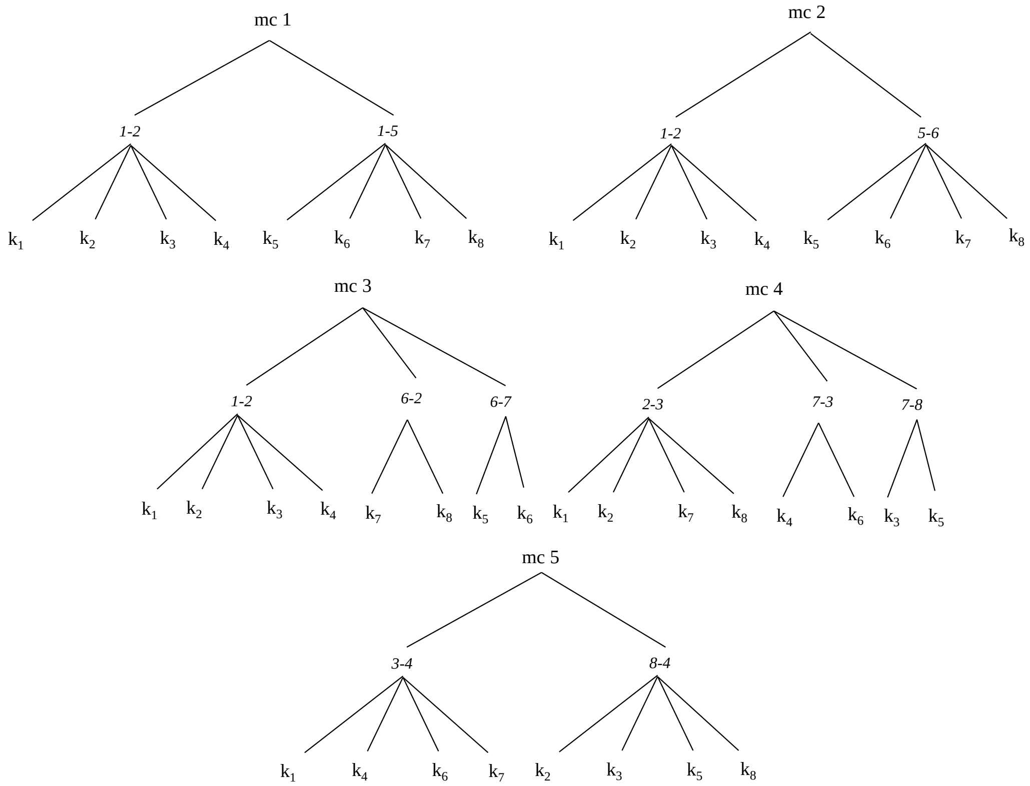


Figure 6.2: Grid network (Figure 5.1): CoNL specification for o-d pair **1-4** *with the exhaustive acyclic choice-set.*

For the choice-set generation and the choice probabilities computation of each model, a stochastic assignment software has been developed using Matlab R2018.a. Figure 6.3 represents a screenshot of the software menu.

Totally, 3.599 alternative routes are generated with the Monte Carlo technique, on 142 different o-d pairs.

For all the models the systematic utilities are specified purely as a function of the route costs, computing these as:

$$V_k = \beta_c \cdot C_k = \beta_c \cdot \sum_l a_{lk} \cdot c_l \qquad \forall k \tag{6.22}$$

For fixing a scale and for interpreting the negative effect of the route cost on the systematic utility, a cost parameter β_c=-1 is assigned to each routes utilities. In the C-Logit model formulation, the parameter β_0 in (2.86) is set to be 1, as well as the β_{PS} and the γ for the Path Size Logit in (2.91). The CoNL is implemented, instead, in its original version with a practically null δ_{MIN}.

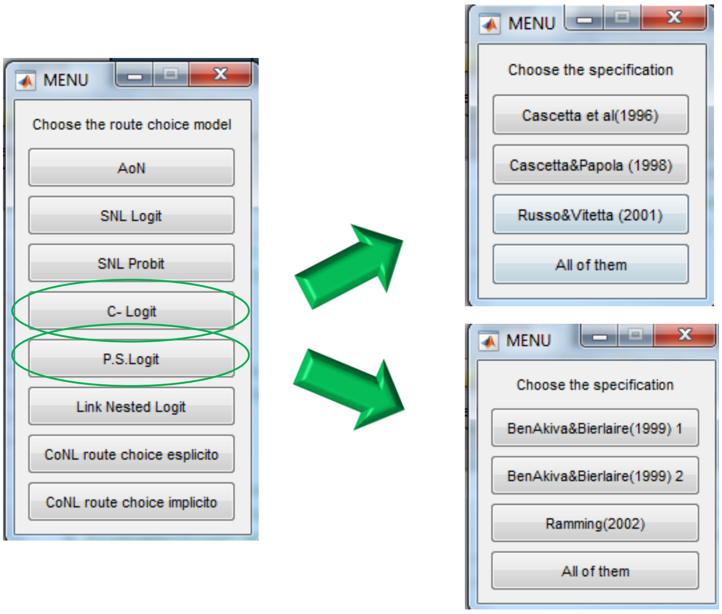


Figure 6.3: Screenshot of the SNL software developed in Matlab R2018a.

As shown in Table 6.14, the CoNL route choice provides the best performance in reproducing observed choices.

	MNL	C-Logit	PS-Logit	CoNL
LL **(cv=0.1)**	-770,0	-708,0	-983,1	**-700,7**
LL **(cv=0.2)**	-586,6	-525,6	-793,0	**-515,3**
LL **(cv=0.3)**	-527,5	-467,2	-731,2	**-455,9**

Table 6.14: Log-likelihood computation for the compared models.

6.5 Conclusions and future steps

The chapter proposes some advance on CoNL route choice model, with reference to the model version described in Chapter 5. An implicit enumeration algorithm for computing link flows consistent with the CoNL route choice model on the efficient sub-network has been illustrated. In the order, Section 6.2 provides a theoretical demonstration, some considerations on the adaptation of the mixing components procedure for the implicit algorithm, an operative general procedure, a simplified version of the latter and some experimental results on toy networks. Section 6.3, instead, carries out an in depth analysis on the real capability to reproduce target correlation matrices given by the Daganzo and Sheffi (1977) paradigm, contrasting the results with other models, both in terms of reproduced covariances (in the space of utilities and in the space of utility differences) and in terms of probabilities. Particularly, in the same sub-Section, some minor changes to the CoNL nesting parameters formulation are tested and general conclusions about the relationship between correlations and choice probabilities for the G.E.V. models are derived. Finally, Section 6.4 performs the validation of the original CoNL route choice model (Papola et al., 2018), on a dataset of real observed route choices. In addition, the set of observed not efficient routes is considered into the specification and a way to do it is proposed.

First, the first implicit enumeration procedure is effective to reproduce CoNL route choice flows on the efficient sub-network, but with an high computational cost due to the numerous networks explorations. Thus, the simplified procedure, i.e. the one with fixed values for the nesting parameters, is by far less expensive, reducing considerably (at least one order of magnitude) the computation times, at a cost of a very small bias in the computed CoNL flows.
Second, the CoNL route choice and the Link Nested Logit with fixed nesting parameters values provides the best performances in reproducing a target Daganzo and Sheffi correlation matrix and a target MNP probabilities vector. The two phenomena do not go in the same direction, so a G.E.V. model that is capable to catch complex correlation patterns effects is very suitable, but the correlations are not the only parties in presence. In fact, in the G.E.V. models, there is not a one-to-one relationship between choice probabilities and covariances, while there is an influence of the moments of order higher than the second. However, weighting the link costs in the nesting parameters CoNL computation, and replacing the average cost value in (5.11) with the minimum or maximum values, can give a good improvement in terms of choice probabilities values, with reference to the ones already published in Papola et al. (2018).
Third, the CoNL route choice without the restriction due to the efficient routes, shows its superiority in reproducing observed route choices with reference to the MNL, C-Logit and P.S.-Logit.
Surely, a first future step in the current research topic is represented by a real data validation of the SNL CoNL with implicit enumeration, searching for a procedure that relaxes the restriction due to the efficient sub-network. The correlations study, insetad, can be surely improved by performing more tests. The real data validation for CoNL route choice with the explicit path enumeration can be expanded, considering a bigger set of trajectories and more utility attributes (socio-demographic, number of turn-lefts, number of traffic lights encountered and so on) can be put in the model specification, to estimate the model parameters. The set of compared models can be enlarged and, finally, the mathematical S.U.E. properties using CoNL route choice model can be analysed, to ensure the existence and unicity of the solution of the equilibrium fixed point problem.

Chapter 7: Conclusions of the thesis

This thesis investigates the potential of the combination of random utility models (CoRUM; Papola, 2016) for travel demand analysis and any application of discrete choice modelling. In the current work, several theoretical advance and some specific transport-field applications are carried out. The CoRUM framework is very general and allows handling several discrete choice modelling crucial issues. The latter are of particular interest when modelling the travel behaviour. The thesis follows two main research paths: the analysis of the theoretical potential of the CoRUM (Chapter 3 and Chapter 4) and the applications of the CoRUM to the route choice modelling (Chapter 5 and Chapter 6).

Chapter 1 mentions all the relevant discrete choice modelling issues. This thesis addresses, in particular, two of them: inter-correlations the problems related to the error structure and the inter-intra-respondent taste heterogeneity. The two problems represent the vehicle with which Chapter 2 has been presented. In the original work of Papola (2016), the CoRUM has been tested in its flexibility to capture flexible substitution patterns among choice alternatives, in discrete choice contexts characterized by different degrees of correlations. Particularly, thanks to its closed form of covariances, the combination of Nested Logit's (CoNL) specification is appeared to be a very intuitive, simple and powerful tool to deal with the correlations problem. In fact, to the best of my knowledge, the CoNL model is the only one who ensures a flexible correlation pattern with a closed form statement for the covariances and correlations. Given these preconditions, the CoRUM framework, and particularly the CoNL specification, has opened interesting scenarios for possible applications, as well as having different possibilities to broaden its horizons.

First, the possibility of theoretically generalizing the model, adding more capabilities, such as allowing accommodating taste variation and heteroskedasticity, seemed a natural step to be completed. Second, an in depth analysis on how this flexibility could improve forecasting power of the model appeared to be opportune. In fact, making good forecasts is essential for transport policy. The market shares of users who choose a certain transport alternative (e.g., whether to move or not, when, how, on what route and so forth) are essential quantities to estimate when analysing the travel demand. Therefore, the sensitivity of the market shares to changes in the values of the attributes involved plays a fundamental role. Finally, the particular CoNL specification, with its own peculiarity of being a closed form models both in terms of probabilities and covariances/correlations, was indicated by the author itself as a natural candidate for route choice modelling, wherein prior expectations in terms of correlations can be quantified (see paradigm of Daganzo and Sheffi, 1977). Since Papola (2016) provided only a few general guidelines, the operationalization of the model for route choice had not yet been explored.

The first step consists of proposing a generalization of the CoRUM framework, allowing taking into account the taste variation and the heteroskedasticity. A Combination of Mixed RUMs formulation is proposed Chapter 3. The proposed formulation allows avoiding the identification issues and the computational burdens of the Mixed Logit with joint error component / random coefficient formulation, that currently represents the theoretical more general formulation available (McFadden and Train, 2000). Such Combination of Mixed RUMs is estimated on a stated survey of 1688 observations of 211 respondents (8 choice tasks per person). The Combination of Mixed RUMs, especially when combining Mixed Nested Logit, outperforms all the other tested mixed models (Mixed Logit, Mixed NL, Mixed CNL) in terms of goodness of fit. In particular, the Cross Nested Logit with random parameters seems very hard to estimate and Nested Logit with random parameters allows only partially to reproduce inter-alternative correlations, apart from the rate due to the random parameters. The Mixed Logit with joint random coefficient and error component, instead, although its theoretical generality is very hard to specify to ensure identification of the parameters (Walker et al., 2007). In fact, such formulation requires an high awareness of its theoretical background and involves very complex simulations (exploding with the dimension of the choice set) and mathematical preliminary evaluations (see rank condition, order condition and equality condition described in Section 2.2.1.3 for ensuring the identification of the parameters). Thus, it seems that this general and powerful model have advantages that are more theoretical than practical. In the real-world applications, this means that for making it operational, several strong constraints have to be introduced (for instance, parametrized covariance matrices or non-full covariance matrices). Therefore, the Combination of Mixed Nested Logit is a compromise between the not generality of Mixed Logit with pure random coefficient specification and the computational hard treatability of joint Error Component / Random coefficient Mixed Logit formulations. However, such combination of mixed Nested Logit is widely easier to estimate and to manage. A future step may certainly be the extension of the study for more flexible sieve estimators (see description of semi-nonparametric approaches for taste heterogeneity described in Section 2.2.2.4), or allowing more general shapes of distribution for random parameters. Another one is to test the capability of the CoRUM formulation under the nonparametric approach, by combining a continuous mixture for the random residuals with a discrete flexible mixture on the marginal utilities. The advantage would be to work with a closed form expression for the choice probabilities. Furthermore, the gains of such formulations in a posterior analysis prediction seems a proper step. In fact, allowing for disentangling between different effects (such as the correlation effects and the taste heterogeneity effects) may have a strong impact on the goodness of predictions about the behaviour of individuals whose previous choices have been observed. In the current micro-data era, this seems to be a very crucial question to investigate, opening very wide horizons on individual level parameters or individual models estimation possibilities.

The second thesis step consists of a synthetic dataset based analysis on forecasting capability of the CoRUM model, particularly the CoNL specification, when contrasted with other closed-form formulations. The advantage of building a controlled environment for analysis is the knowledge of all the real values (true marginal utilities and true underlying correlations). Moreover, with such experiment, the choices on variate scenarios can be observed, and so, the models can be tested in their ability to capture them. The CoNL has been compared with the main closed form random utility models. In particular, the models compared are the Multinomial Logit, the Nested Logit and other models characterized by the same flexibility in reproducing correlation patterns, namely the Cross Nested Logit and the FinMix models. The experimental results, summarized in the Section 4.4, highlight that, when making forecasting, the importance of the sample size the analyst are working with is crucial. In fact, the models with good flexibility in reproducing correlation contexts (CoNL, CNL and FinMix) seem to be capable to reproduce correct forecasts in very challenging correlations scenarios, but they need the right number of observations In fact, also in a very simple choice scenario as the one proposed in Section 4.3 (three/four alternatives, no taste variation, a few parameters playing a role) with cross sectional data, the appropriate model (i.e. the model capable to handle that correlation pattern) needs rarely less than a thousand observations to make unbiased forecasts. The analysis confirms, instead, that using the wrong model has a strong influence also when predicting choices in variate scenarios, i.e. the scenarios wherein the values of the attributes change. Definitely, the choice of the model strictly depends upon the availability of data, not just on the trail and errors processes and the analyst's expectations (necessary pre-requisites of choice analysis). The goodness of fit indicators values *per se*, in fact, can mislead the analyst when he must predict future scenarios but he has small samples of data. Surely, the case studies could be enriched (different choice contexts, more random effects in the synthetic population) and the set of models involved can be enlarged.

The final step consists of some methodological advance on route choice modelling and is described in Chapter 5 and Chapter 6. Route choice represents the core of all assignment procedures and the problem is characterized by several sources of complexities. The dimension of the problem (dimension of network, number of involved od pairs and routes) and the routes overlapping effects on choice probabilities (routes perception as independent alternative) are the more relevant. This means an analyst, for addressing these issues, needs an instrument (route choice model) that is both computationally efficient but that also has a theoretical robust foundation. For applying CoNL model to route choice, several problems had to be solved. First, the creation of a network-based procedure for building the model specification. Second, a way to compute the prohibitive number of structural parameters of the model. Third, test the choice probabilities resulting from a specification procedure that is not manually built ad hoc by the analyst. Fourth, understand how closed form covariances based nesting parameters computation can improve the choice probabilities. Fifth, the possibility of implementing the model for computing the traffic link flows without the burdensome explicit enumeration of paths. The current thesis proposes some solutions to all the problems mentioned. The first three problems are addressed in Chapter 5. An algorithm has been proposed to specify the CoNL model, given any network, and an exact formulation for nesting parameters, derived from the Daganzo and Sheffi (1977) assumption, have been proposed. All tested networks (small, medium and large) confirm the theoretical robustness of the CoNL route choice model, with reference to all the main existing route choice models in literature. The last two problems are addressed in Chapter 6. The analysis on CoNL correlations effects on choice probabilities show the crucial importance of setting a minimum value for nesting parameters. This appears to be necessary for avoiding a "too deterministic" choice within a group of correlated (i.e. overlapped) alternatives. Finally, an implicit enumeration algorithm for CoNL route choice has been proposed and tested on some small and medium networks. The algorithm works on efficient sub-network (see Dial, 1971) and gives the same results of CoNL route choice model with the explicit enumeration, without the computational burdens due to the enumeration of paths. A simplified version of the implicit enumeration algorithm is also proposed, for consistently reducing computation time, with a not substantial bias of the CoNL route choice probabilities. Finally, the CoNL route choice model has been tested in reproducing observed route choices, on a dataset of about 200 trajectories of drivers moving into the network of Regione Campania. The model showed its superiority over other models in the literature that are commonly implemented in the commercial software. However, the research is still open. The goodness of fit of the novel CoNL route choice model must be tested in estimation on an opportune dataset of observations (some thousands). The implicit enumeration algorithm could be tested on a real network with the availability of traffic counts. Finally, a S.U.E. mathematical programming formulation for CoNL route choice has to be provided, in order to ensure the convergence and the unicity of the fixed-point problem equilibrium solution.

Contributi a questo lavoro

Quando si è studenti universitari si comprano libri solo se costretti dai docenti. Quando si è professionisti si acquistano volentieri libri che possano risolvere dubbi pratici in poco tempo. Quando si fa ricerca, invece, si pagherebbe oro pur di avere libri che sviscerano i propri argomenti di interesse con chiarezza, rigore e dettaglio. Avendo avuto la fortuna di vivere la parabola laurea-professione-ricerca, ci sono una serie di autori a cui mi piacerebbe, sinceramente, stringere la mano. Le citazioni che gli ho dedicato in bibliografia rendano merito al loro lavoro, cui sono personalmente grato per tutto ciò che ho appreso, spesso grazie a letture fatte in orari improponibili.

Dall'altro lato, invece, dovrei menzionare tutti gli studenti a cui ho provato a trasferire qualche mia conoscenza. La responsabilità di dover chiarire loro dubbi di qualsiasi tipo, in orario di ricevimento o di lezione, mi ha spinto a farmi domande di qualità sempre maggiore e, dunque, a comprendere meglio molte delle cose che ho riportato nel Capitolo 2. Lo sforzo di includerli tutti sarebbe improponibile e, pertanto, mi limito a rivolgere a loro un pensiero di gratitudine.

Premesso ciò, ci sono delle persone che, invece, ho il dovere di menzionare direttamente, perché, in un modo o nell'altro, questa tesi include anche i loro sforzi.

Comincerei da quei (pochi) tesisti, il cui lavoro si è incrociato con il mio. Anzitutto vorrei citare il lavoro di Vittorio Fontana e Carlo Malafronte. Il loro sforzo sulla raccolta di dati da traiettorie GPS e il modello di offerta da loro implementato hanno prodotto la base su cui è stato possibile costruire il confronto su rete reale di Section 6.4, consentendo una validazione su dati reali del modello CoNL route choice. Vorrei menzionare anche Ciro Buonocore, che da poco si è aggregato al nostro ristretto gruppo di ricerca. Anche se non è parte attiva in questo lavoro, il fatto che si occupi di tutto ciò di cui non riesco ad occuparmi io direttamente mi ha consentito di lavorare più serenamente alla mia ricerca in questi ultimi mesi di dottorato. Inoltre, mi offre la possibilità, quotidianamente, di provare ad aiutarlo nel suo percorso di formazione come PhD.

Fra le esperienze più significative di questi tre anni ci sono stati sicuramente due seminari che ho avuto il piacere di seguire a Londra, organizzati dall' Institute of Choice Modelling di Leeds. Le lezioni che ho avuto modo di seguire in quella sede, in particolare quelle del prof. Stephane Hess, sono state il maggiore stimolo che ha portato alla luce il capitolo 3 di questa tesi.

Ritengo doveroso citare anche due docenti che, in maniera del tutto gratuita e disinteressata, hanno espresso un parere dettagliato su questo lavoro come valutatori: il prof. Michele Ottomanelli del Politecnico di Bari e il prof. Stefano De Luca dell'Università degli studi di Fisciano. Ai consigli del secondo devo, in particolare, la struttura finale dei capitoli con la quale la tesi si presenta.

Un ringraziamento particolare va al prof. Vittorio Marzano, che è stato quanto di più prossimo ad un co-tutor io abbia avuto in questo periodo. C'è molto del suo lavoro nei capitoli 4 e 5 di questa tesi. Una sua proposta è stata da stimolo per gli approfondimenti proposti in Section 6.3 sulle correlazioni riprodotte dai modelli di scelta del percorso. Lo ringrazio inoltre per il

contributo sempre brillante, colto ed incisivo che riesce a dare alle nostre discussioni, nonché per tutti gli illuminanti consigli che ha saputo offrirmi in questi tre anni.
Il principale ringraziamento per questo lavoro lo devo al mio tutor: il prof. Andrea Papola. Lui mi ha convinto a fare il dottorato di ricerca quando era ormai fuori dal mio insieme di scelta per il futuro. Lui ha avuto l'idea, in quel Maggio 2015, di introdurre un ingegnere che si occupava di strutture, senza la minima conoscenza del mondo dei trasporti e della modellazione di scelte discrete, nel suo gruppo di ricerca. Lui ha creduto che io potessi fare ricerca in questo campo quando io non lo credevo minimamente possibile. Dal suo articolo sul CoRUM sono nati tutti i contenuti di questa tesi, che altro non sono che la generalizzazione e qualche particolarizzazione della sua intuizione originaria. Sopra ogni altra cosa, lo ringrazio per la serenità sempre trasmessami nel lavoro, aggiungendo che augurerei a chiunque di avere un tutor con i suoi intuito, intelligenza, umanità ed attenzione per la persona. Infine, lo ringrazio per tutte le nostre discussioni, spesso davanti a un caffè, sui modelli di scelta discreta. Con la certezza che, da queste discussioni, siano nati tanti chiarimenti (spero reciproci) e con la speranza che, da esse, sia nato anche qualche avanzamento di qualità sul choice modelling.

Ringraziamenti personali

Il lavoro di ricerca è qualcosa di totalizzante. Spesso, il tempo che si dedica al lavoro da un piacere che fa dimenticare quanto se ne stia sottraendo ad altro. In questo altro, però, c'è la base su cui si fonda la serenità per cui il lavoro stesso può essere affrontato. Pertanto, con lo sforzo di non seguire un ordine specifico, c'è qualcuno a cui è necessario che io renda merito.
A vario titolo ho condiviso esperienze di convegni e corsi di formazione con tanti colleghi, che hanno reso ogni trasferta un'esperienza unica e stimolante.
Tutti i colleghi di dipartimento con cui ho avuto a che fare lungo questo percorso hanno spesso fatto sì che quest'ultimo divenisse, per me, un po' una seconda casa, in cui rifugiarsi quando le insicurezze del mondo esterno attanagliano spietatamente.
Andrea Papola non è solo il mio tutor. E' un collega, un amico, una specie di fratello maggiore. Anche volendo tralasciare il lavoro, è la persona da cui ho imparato di più in questi ultimi anni. Le nostre chiacchierate sono spesso state anche la valvola di sfogo per affrontare tanti problemi che col lavoro c'entrano poco. Soprattutto, gli devo il merito di aver sempre compreso i miei momenti di difficoltà e di aver avuto la pazienza di aspettarne di migliori.
Mia madre e mio padre meritano tanto. A loro sono totalmente grato per una serie non elencabile di cose. Volendone citare una sola, devo a loro il fatto di essere potuto rimanere a Napoli per tutto questi anni dopo la laurea. A loro devo anche tutta la mia riconoscenza per aver rispettato sempre le mie scelte. Il loro supporto è stato imprescindibile per me perché potessi portare a termine questo dottorato, soprattutto nell'ultimo anno. Devo inoltre dire grazie a mia madre per la vicinanza che mi ha offerto, avendo sempre avuto la maturità di perdonare tanti miei sfoghi e l'intelligenza di offrirmi il suo supporto in maniera mai invadente. Spero che questo piccolo traguardo, e il fatto di essere diventata nonna, bastino, almeno in parte, a ricambiare tutti i suoi sforzi di genitore.
Casa è dove qualcuno ti aspetta. E' per questo che a Carmy devo un grazie incommensurabile. Il fatto di aver sopportato sempre i miei orari e le mie assenze dovute al lavoro, senza avermi fatto minimamente pesare nulla, rappresenta una rarità, che ormai per me è normalità, ma che non darò mai per scontato. La ringrazio per tutti sacrifici che ogni giorno fa con amore, e per il fatto che, a qualsiasi ora torni, lei ha sempre da parte per me un abbraccio ed un sorriso.
Infine, dal 21 Marzo 2018, c'è qualcuno che illumina le mie giornate e che ha introdotto nella mia vita una sorta di effetto scala. Da quando c'è lei tutto intorno ha subito un effetto ridimensionante. Le cose intorno a me sono le stesse, ma la loro importanza si è drasticamente ridimensionata. Parlo di mia figlia Isabel. Con la speranza che magari un giorno possa prendere questo lavoro, leggerlo, comprenderlo e, soprattutto, dirmi che ormai non serve più nulla.

References

Abbé, E., Bierlaire, M., Toledo, T., 2007. Normalization and correlation of Cross-Nested Logit models. Transportation Research Part B 41 (7), 795-808.

Allenby, G.M., Rossi, P.E., 1998. Marketing models of consumer heterogeneity. J. Econometrics 89 (1), 57–78.

Akamatsu, T., 1996. Cyclic flows, Markov process and stochastic traffic assignment. Transportation Research Part B, 30(5), 369–386.

Akamatsu, T., 1997. Decomposition of Path Choice Entropy in General Transport Networks. Transportation Science 31(4), 349–362.

Akin, J., Guilkey, D., Sickles, R., 1979. A Random Coefficient Probit Model with an Application to a Study of Migration," Journal of Econometrics, 11, pp. 233-246

Avineri, E., Prashker, J.N., 2004. Violations of expected utility theory in route-choice stated preferences. Transportation Research Record: Journal of the Transportation Research Board 1894, 222–229.

Azevedo, J.A., Santos Costa, M.E.O., Silvestre Madera, J.J.E.R., Vieira Martins., E.Q., 1993. An Algorithm for the ranking of shortest paths. European Journal of Operational Research, Vol. 69, 97–106.

Baillon, J. B., Cominetti, R., 2008. Markovian traffic equilibrium. Mathematical Programming 111(1-2), 33–56.

Bajari, P., Fox, J.T., Ryan, S.P., 2007. Linear regression estimation of discrete choice models with nonparametric distributions of random coefficients. Am. Econ. Rev. 459–463.

Bansal, P., Daziano, R., Achtnicht, M., 2018. Extending the logit mixed logit model for a combination of random and fixed parameters. Journal of Choice Modelling 27 (2018) 88-96.

Bastani, P., Weeks, M., 2013. Nonparametric estimation of mixing distributions in discrete choice models: Heterogeneity and WTP. European Economic Association & Econometric Society Parallel Meetings, Gothenburg.

Bastin, F., Cirillo, C., Toint, P.L., 2010. Estimating nonparametric random utility models with an application to the value of time in heterogeneous populations. Transp. Sci. 44 (4), 537–549.

Batley, R, Daly, A., 2006. On the equivalence between elimination-byaspects and generalized extreme value models of choice behavior. J. Math. Psych. 50(5):456–467.

Batley, R., Hess, S., 2016. Testing for regularity and stochastic transitivity using the structural parameters ofnested logit. Transportation Research Part B. Methodological 93 (2016), 355-376.

Beilner, H., Jacobs, F., 1971. Probabilistic aspects of traffic assignment. Proc. 5th Intern. Symp. On the Theory of Traffic Flow and Transportation. Berkeley, 1971, pp.183-194.

Bell, M. G. H., 1995. Alternatives to Dial's logit assignment algorithm. Transportation Research Part B 29(4), 287–295.

Bekhor, S. and Prashker, J.N., 1999. Formulations of extended logit stochastic user equilibrium assignments. Proceedings of the 14th International Symposium on Transportation and Traffic Theory, Jerusalem, Israel, 351-372.

Bekhor, S., Ben-Akiva, M.E., Ramming, S., 2001. Route Choice: Choice Set Generation and Probabilistic Choice Models. In Proc., 4th Triennial Symposium on Transportation Analysis Conference, University of Azores, Sao Miguel, Portugal, 459–464.

Bekhor S. and Prashker J., 2001. A stochastic user equilibrium formulation for the generalized nested logit model. Transportation Research Record 1752, 84-90.

Bekhor, S., Ben-Akiva, M.E., Ramming, M.S., 2002. Adaptation of Logit Kernel to route choice situation. Transportation Research Record 1805, 78-85.

Bekhor, S., Ben-Akiva, M.E., Ramming, M.S., 2006. Evaluation of choice-set generation algorithms for route choice models. Annals of Operations Research, 144, 235-247.

Bekhor, S., Toledo, T., Reznikova, L., 2008. A path-based algorithm for the Cross Nested Logit stochastic user equilibrium traffic assignment, Computer-Aided Civil and Infrastructure Engineering, 24(1), 15–25.

Ben-Akiva, M.E., 1972. The Structure of Travel Demand Models. Ph.D. dissertation, Massachusetts Institute of Technology.

Ben-Akiva, M.E., 1974. Structure of passenger travel demand models. Transportation Research Record 526, 26±42.

Ben-Akiva, M.E., Richards, M.G., 1975. A Disaggregate Multi-modal Model for Work Trips in the Netherlands, paper prepared for presentation at the 54th annual meeting of the Transportation Research Board, Washington, D.C.

Ben-Akiva, M.E., Lerman, S., 1977. Disaggregate Travel and Mobility Choice Models and Measures of Accessibility. Third International Conference on Behavioural Travel Modelling, Tanenda, Australia, 1977.

Ben-Akiva, M.E., Francois, B., 1983. Mu-homogenous generalized extreme value model, Working Paper, Department of Civil Engineering, MIT.

Ben-Akiva, M.E., Bergman, M.J., Daly, A.J., Ramaswamy, R., 1984. Modeling inter-urban route choice behaviour. In Proc., 9th International Symposium on Transportation and Traffic Theory, VNU Science Press, Utrecht, Netherlands, 299–330.

Ben-Akiva, M.E., Swait, J., 1984. Chocie models with simple probabilistic choice set generatio processes. Working paper, Department of Civile Engineering, MIT, Cambridge MA.

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete choice analysis. Cambridge, MA: MIT Press.

Ben-Akiva, M.E., Bolduc, D., Bradley, M., 1993. Estimation of travel choice models with randomly distributed values of time. Transportation Research Record 1413.

Ben-Akiva, M.E., Boccara, B., 1995. Discrete choice models with latet choice sets. International Journal of Research in Marketing 12 (1995), 9-24.

Ben-Akiva, M.E., Bolduc, D., 1996. Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper, Department of Civil Engineering, MIT.

Ben-Akiva, M.E., Ramming S., 1998. Lecture notes: discrete choice models of traveler behavior in networks. Prepared for Advanced Methods for Planning and Management of Transportation Networks. Capri, Italy.

Ben-Akiva, M.E., Bierlaire, M., 1999*b*. Discrete choice methods and their applications to short-term travel decisions. Handbook of Transportation Science, Kluwer, 26, 5–34.

Ben-Akiva, M.E., McFadden, D., Train, K., Garling, D., Gopinath, D., Walker, J.L., Bolduc, D., Borsch-Supan, A., Delquié, P., Larichev, O., Morikawa, T., Polydoropoulou, A., Rao, V., 1999. Extended framework for modelling choice behaviour.Marketing Letters 10:3 (1999), 187-203.

Ben-Akiva, M.E., McFadden, D., Garling, D., Walker, J.L., Bhat, C., Bierlaire, M., Bolduc, D., Borsch-Supan, A., Brownstone, D., Bunch, D.S., Daly, A., de Palma, A., Gopinath, D., Karlstrom, A., Munizaga, M.A., 2002. Hybrid choice models: progress and challenges. Marketing Letters 10:3 (2002), 163-175.

Bhat, C., 1995. A heteroscedastic extreme value model of intercity mode choice, Transportation Research B 29, 471–483.

Bhat, C.R., 1997. An endogenous segmentation mode choice model with an application to intercity travel. Transp. Sci. 31, 34–48.

Bhat, C., 1998a. Accommodating variations in responsiveness to level-of-service variables in travel mode choice models. Transportation Research A 32, 455–507.

Bhat, C., 2000. Incorporating observed and unobserved heterogeneity in urban work mode choice modelling. Transportation Science 34, 228–238.

Bhat, C.R., Guo, J.Y., 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. Transportation Research Part B 38 (2), 147–168.

Bhat, C.R., 2005. A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-use Decisions. Transportation Research Part B, 39(8), 679–707.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transp. Res. Part B: 45 (7), 923–939.

Bhat, C.R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. Transp. Res. Part B 46 (7), 817–833.

Bhat, C.R., Lavieri, P.S., 2018. A new mixed MNP model accomodating a variety of dependent non-normal coefficient distributions. Theory Dec. (2018) 84: 239-275.

Bierlaire M., 2001. A general formulation of the cross-nested logit model. 1st Swiss Transport Research Conference. Monte Verità, Ascona, March 1-3, 2001.

Bierlaire M., 2003. BIOGEME: a free package for the estiamation of discrete choice models, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Block, H.D., Marschak, J., 1959. Random ordering and stochastic theories of response. Cowles Foundation Discussion Paper, No. 66.

Bodapati, A.V., Gupta, S., 2004. The recoverability of segmentation structure from store-level aggregate data. Journal of Marketing Research 41(3).

Borsch-Supan, A., 1990. On the compatibility of nested logit models with utility maximization, Journal of Econometrics 43, 373-388.

Borsch-Supan, A., Hajivassiliou, V., 1993. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. Journal of Econometrics. 58, 3477368.

Bovy, P.H.L., Fiorenzo-Catalano, S., 2007. Stochastic route choice set generation: behavioral and probabilistic foundations. Transportmetrica, 3(3), 173-189.

Bovy, P. H. L., Bekhor, S., Prato, C. G., 2009. The factor of revisited path size: alternative derivation, Transportation Research Record, 2076, 132–40.

Boyd, J., Mellman, J., 1980. The effect of fuel economy standards on the U.S. automotive market: a hedonic demand analysis. Transportation Research 14A(5±6): 367±378.

Bresnahan, T.E., Stern, S., Trajtenberg, M., 1997. Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the late 1980s, Rand Journal of Economics, Vol. 28, Special Issue, pp. S17-S44.

Brownstone, D., Train, K., 1999. Forecasting new product penetration with flexible substitution patterns, Journal of Econometrics 89, 109–129.

Bujosa, A., Riera, A., Hicks, R.L., 2010. Combining discrete and continuous representations of preference heterogeneity: a latent class approach. Environ. Res. Econ. 47 (4), 477–493.

Bunch D.S., 1991. Estimability in the Multinomial Probit model. Transportation Research Part B: Methodological, Volume 25, Issue 1, 1-12.

Burda, M., Harding, M., Hausman, J., 2008. A Bayesian mixed logit-probit model for multinomial choice. J. Econ. 147 (2), 232–246.

Burrell, J. E., 1968. Multiple route assignment and its application to capacity restraint. Proceedings of rhe 4th Infernufionul Symposium on Transportation and Traffic Theory, University of Karlsruhe.

Busemeyer, J.R., Townsend, J.T. 1992. Fundamental derivations from decision field theory. Math. Soc. Sci. 23 (3), 255–282.

Busemeyer, J.R., Townsend, J.T. ,1993. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment.. Psychol. Rev. 100 (3), 432.

Butler, J., Moffitt, R., 1982. A computationally efficient quadrature procedure for the one factor multinomial probit model. Econometrica 50, 761–764.

Cardell, N., Dunbar, F., 1980. Measuring the societal impacts of automobile downsizing. Transportation Research 14A(5±6): 423±434.

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified Logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Jean-Baptiste Lesort ed., Pergamon, 696-711.

Cascetta, E., Papola, A., 1998. Modified logit route choice models for networks with overlapping paths: new formulations and empirical results. Atti del VI Convegno Euro Working Group, Göteborg, Sweden.

Cascetta, E., Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand, Transportation Research C, 9(6), 249–63.

Cascetta, E., Russo, F., Viola, F.A., Vitetta, A., 2002. A model of route perception in urban road networks. Transportation Research Part B, 36(7), 577-592.

Cascetta, E., 2006. Modelli per i sistemi di trasporto. De Agostini Scuola Spa, novara (Italy).

Cascetta, E., 2009. Transportation system analysis: Models and Applications. Springer, New York.

Castillo, E., Menéndez, J. M., Jiménez, P. and Rivas, A., 2008. Closed form expression for choice probabilities in the Weibull case, Transportation Research Part B, 42(4), 373-380.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. Rev. Econ. Stud. 47, 225–238 .

Chen, A., Pravinvongvuth, S., Xu, X., Ryu, S., Chootinan, P., 2012. Examining the scaling effect and overlapping problem in logit-based stochastic user equilibrium models. Transportation Research Part A, 46(8), 1343-1358.

Cheng, H., Yan, X., 2015. Random parameter nested logit model for combined departure time and route choice. International Journal of Transportation Science and Technology, vol. 4, no. 1, 2015, pp. 93-106.

Chipman, J., 1960. The foundations of utility. Econometrica 28, 193–224.

Chorus, C.G., Arentze, T.A. and Timmermans, H.J.P. (2008). A Random Regret Minimization model of travel choice. Transportation Research Part B, 42(1), pp. 1-18.

Chorus, C.G., 2010. A new model of Random Regret Minimization. European Journal of Transport and Infrastructure Research 10 (2), 181–196.

Chorus, C.G., Rose, J.M., Hensher, D.A., 2013. Regret minimization or utility maximization: It depends on the attribute. Environment and Planning Part B 40, 154–169.

Chu, C., 1989. A paired combinatorial logit model for travel demand analysis. Proceedings of 5th World Conference on Transportation Research, Ventura, CA.

Clark., C., 1961, The greatest of a finite set of random variables, Operations Research 9, 145- 162.

Connors R.D., Hess S., Daly A., 2012. Analytic approximations for computing Probit choice probabilities. Transportmetrica A: Transport Science, Volume 10, Issue 2, 119-139.

Daganzo, C.F., Bouthelier, F., Sheffi, Y., 1977a. An efficient approach to estimate and predict with multinomial probit models, 56th Meeting Transp. Res. Board (unpublished), cited in Daganzo (1979).

Daganzo, C.F., Sheffi, Y., 1977. On stochastic models of traffic assignment. Transportation Science, vol. 11, no. 3.

Daganzo, C.F., 1978a. A closed form random utiliti model with error terms of different variances, unpublished research notes.

Daganzo, C.F., 1979. Multinomial Probit: the theory and its application to demand forecasting. Economic Theory, Econometrics, and Mathematical Economics. University of Pennsylvania, Philadelphia.

Daganzo, C.F., Kusnic, M., 1993. Two properties of the nested logit model. Transportation Science 27(4), 395-400.

Dagsvik,, J., 1994. Discrete and continuous choice, max-stable process, and independence from irrelevant alternatives, Econometrica 62: 1179-1205.

Dagsvik,, J., 1995. How large is the class of random utility models?. Journal of mathematical Psychology, Volume 39, Issue 1, pp. 90-98.

Dagsvik, J., 2004. Random utility models for discrete choice behaviour: an introduction. Statistics Norway, No-0033, Oslo, Norway.

Daly, A.J., Zachary, S., 1978. Improved multiple choice models. In D.A. Hensher and M.Q. Dalvi (eds.), Determinants of Travel Choice, Saxon House, Westmead.

Daly, A., 2001b. Recursive nested EV model. Working Paper ITS 559, Institute for Transport Studies, University of Leeds, UK.

Daly, A., Bierlaire, M., 2006. A general and operational representation of GEV models. Transportation Research Part B 40 (4), 285-305.

Dayton, M.C., Macready, G.B., 1988. "Concomitant-Variable Latent Class Models", *Journal of the American Statistical Society* 83, 173-179.

De La Barra, T., Perez, B., Anez, J., 1993. Multidimensional Path Search and Assignment. Presented at 21st PTRC Summer Annual Meeting, Manchester, United Kingdom.

De Luca, S., Di Pace, R., 2015. Evaluation of risk perception approach in route choice experiments: an application of the cumulative Prospect Theory. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, Volume 2015, 309-315.

de Palma, A., Ben-Akiva, M.E., Brownstone, D., Holt, C., Magnac, T., McFadden, D., Moffatt, P., Picard, N., Train, K., Wakker, P., Walker, J.L., 2008. Risk, uncertainty and discrete choice models. Marketing Letters 19, 269–285.

Deak, I.. 1980a. Three digit accurate multiple normal probabilities, Numerische Mathematik 35. 3699380.

Deik, I., l980b. Fast procedures for generating stationary normal vectors, Journal of Statistical Computation and Simulation 10, 2255242.

Debreu, G., 1960. Review of R.D. Luce individual choice behavior, American Economic Review 50, 186–188.

Dekker, T., 2016. Asymmetric triangular mixing densities for mixed logit models. Journal of Choice modelling 21 (2016) 48-55.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society* B 39, 1-38.

Dial, R.B. 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. Transportation Research, 5, pp. 83-111.

Diederich, A., 1997. Dynamic stochastic models for decision making under time constraints. J. Math. Psychol. 41 (3), 260–274.

Domencich T.A., McFadden D., 1975. *Urban Travel Demand – A Behavioral Analysis*, North Holland Publishing Company, Amsterdam. Oxford American Elsevier Publishing Company, Inc. – New York.

Dong, X., Koppelman, F.S., 2014. Comparison of continuous and discrete representations of unobserved heterogeneity in logit models. J. Mark. Anal. 2 (1), 43–58.

Dunne J.P., 1984. *Elasticity measures and disaggregate choice models,* Journal of Transport Economics and Policy.

Erto, P., 2004. Probabilità e statistica per le scienze e l'ingegneria (2nd edition). McGraw-Hill.

Fiebig, D.G., Keane, M., Louviere, J.J., Wasi, N., 2010. The generalized multinomial logit: accounting for scale and coefficient heterogeneity. Mark. Sci. 29 (3), 393–421.

Finney, D., 1971. Probit analysis. Cambridge University Press, Cambridge, England.

Fishman, G.S., 1973. Concepts and methods in discrete event digital simulation. Wiley, New York.

Fisk, C., 1980. Some developments in equilibrium traffic assignment. Transportation Research Part B, 14(3), 243-255.

Flötteröd, G., Bierlaire, M., 2013. Metropolis-Hastings sampling of paths. Transp. Res. Part B, 48, 53–66.

Fosgerau, M., Bierlaire, M., 2007. A practical test for the choice of mixing distribution in discrete choice models. Transp. Res. Part B 41 (7), 784–794.

Fosgerau, M., Hess, S., 2009. A comparison of methods for representing random taste heterogeneity in discrete choice models. Eur. Transp. 42, 1–25.

Fosgerau, M., Bierlaire, M., 2009. Discrete choice models with multiplicative error terms. Transportation Research Part B, 43(5), 494-505.

Fosgerau, M., Frejinger, E., Karlström, A., 2013. A link based network route choice model with unrestricted choice set. Transportation Research Part B: Methodological, 56, 70-80.

Fosgerau, M., Mabit, S.L., 2013. Easy and flexible mixture distributions. Econ. Lett. 120 (2), 206–210.

Fosgerau, M., McFadden, D., Bierlaire, M., 2013. Choice probability generating functions. The Journal of Choice Modelling 8 (2013), 1-18.

Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. Transportation Research B, vol. 40, pp. 363–378.

Frejinger, E., 2008. Route choice analysis: data, models, algorithms and applications. Ecòle Polytechnique Fédérale de Lausanne, PhD thesis.

Frejinger, E., Bierlaire, M., Ben-Akiva, M.E, 2009. Sampling of alternatives for route choice modeling. Transportation Research Part B 43(2009), 984-994.

Gao, S., Frejinger, E., Ben-Akiva, M.E., 2008. Adaptive route choice models in stochastic time dependent networks, Paper 08-1349 accepted for presentation at the 87th Annual Meeting of the Transportation Research Board, Washington, DC.

Gao, S., Frejinger, E., Ben-Akiva, M.E., 2010. Adaptive route choice in risky traffic networks: A prospect theory approach. Transportation Research, Part C: Emerging Technologies, 18(5), 727-740.

Gaudry, M., Degenais, M., 1978. The Dogit Model, Transportation Research, Part B, Vol. 13, No. 2, pp. 105-111.

Geweke, J., 1991. Efficient simulation from the Multivariate Normal and Student-t distributions subject to linear constraints and the evaluation of constraints probabilities. Meeting Computing Science and Statistics: The Twenty-Third Symposium on the Interface, Seattle.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in J. Bernardo, J. Berger, A. Dawid, and F. Smith, eds., Bayesian Statistics, Oxford University Press, New York, pp. 169–193.

Geweke, J., 1996. Monte Carlo simulation and numerical integration, in D. Kendrick and J. Rust, eds., Handbook of Computational Economics, Elsevier Science, Amsterdam, pp. 731–800.

Gliebe, J.P., Koppelman, F.S., Ziliaskopoulos, A., 1999. Route choice using a paired combinatorial logit model. Paper presented at the 78th meeting of the Transportation Research Board, Washington D.C.

Gonzalez-Valdéz, F., Ortuzar, J.d.D., 2018. The stochastic satisficing model: a bounded rationality discrete choice model. Journal of Choice Modelling 27 (2018) 74-87.

Gopinath, A.D., 1995. Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand Ph.D. Thesis. Department of Civil and Environ- mental Engineering, Massachusetts Institute of Technology. Greene, W., 2001. Fixed and random effects in nonlinear models.

Greene, W.H., 2001. Fixed and random effects in nonlinear models. Working paper EC-01-01, Stern School of Business, Department of Economics.

Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transp. Res. Part B 37 (8), 681–698.

Greene, W.H., Hensher, D.A., 2010. Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. Transportation 37 (3), 413–428.

Greene, W.H., Hensher, D.A., 2012. Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. Applied Economics, 45:14, 1897-1902.

Greene, W.H., Hensher, D.A., 2013. Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. Appl. Econ. 45 (14), 1897–1902.

Guilkey, D., Murphy, J., 1993. Estimation and testing in the random effects probit model. Journal of Econometrics 59, 301–317.

Gupta, S., Chintagunta, P.K., 1994. "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models", Journal of Marketing Research 31, 128-136.

Haghani, M., Sarvi, M., Shahosseini, Z., 2015. Accomodating taste heterogeneity and desired substitution pattern in exit choices of pedestrian crowd evacuees using a mixed nested logit model. The Journal of Choice Modelling 16 (2015) 58-68.

Hajivassiliou, V., McFadden, D., 1990. The method of simulated scores for the estimatmn of LDV models. Cowles Foundation discussion paper no. 967 (Yale University, New Haven, CT).

Hajivassiliou, V., McFadden, D., Ruud, P., 1996. Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. Journal of Econometrics 72, 85–134.

Hanckock, T.O., Hess, S., Choudhury, C.F., 2018. Decision field theory: improvements to current methodology and comparisons with standard choice modelling techniques. Transportation Research Part B, 107 (2018), 18-40.

Hara, Y., Akamatsu, T., 2014. Stochastic user equilibrium traffic assignment with a network gev based route choice model. Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management) 70, 611.

Hausman, J.A., Wise, D.A., 1978. A conditional Probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences, Econometrica 46, 403-426.

Hensher, D., Greene, W., 2003. The mixed logit model: The state of practice and warnings for the unwary. Transportation 30, 133–176.

Hess, S., Polak, J. W., 2004. Mixed Logit estimation of parking type choice. Paper presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC.

Hess, S., Bierlaire, M., Polak, J.W., 2005*a*. Mixed GEV modelling of mode choice. 5th Swiss Transport Research Conference, Monte Verità / Ascona, March 9-11, 2005.

Hess, S., Ben-Akiva, M.E., Gopinath, D., Walker, J.L., 2009. Taste heterogeneity, correlation and elasticities in latent class choice models. Compendium of Papers, 88th Annual Meeting of the Transportation Research Board. Transportation Research Board.

Hess, S., 2010. Conditional parameters estimates from mixed logit models: distributional assumptions and a free software tool. Journal of Choice Modelling 3(2) 134-152.

Hess, S., Rose, J.M., 2012. Can scale and coefficient heterogeneity be separated in random coefficient models? Transportation 39, 4.

Hess, S., Stathopoulos, A., Daly, A.J., 2012. Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. Transportation 39 (3), 565–591

Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. Journal of Choice Modelling 23 (2017) 1-8.

Hildreth, C., Houck, J., 1968. "Some Estimators for a Linear Model with Random Coefficients," Journal of the American Statistical Association, 63, pp. 584-595.

Hoogendoorn-Lanser, S., van Nes, R., Bovy, P., 2005. Path-size and overlap in multimodal networks, in *Proceedings of* 16th ISTTT Conference, College Park, MD.

Horowitz, J., Sparmann, J., Daganzo, C.F., 1982. An investigation of the accuracy of the Clark approximation for the multinomial probit model. Transportation Science 16, 382–401.

Horowitz, J., 1991. Reconsidering the multinomial probit model, Transportation Research B 25, 433–438.

Kamakura W.A., Russell, G.J., 1989. "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure", *Journal of Marketing Research* 26, 379-390.

Kamakura, W.A., Byung-Do, K., Jonathan, L., 1996. Modeling preference and structural heterogeneity in consumer choice. Mark. Sci. 15 (2), 152–172.

Kahneman, D., Tversy, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica 47: 263–291.

Katsikopoulos, K.V., Duse-Anthony, Y., Fisher, D.L., Duffy, S.A., 2000. The framing of drivers route choices when travel time information is provided under varying degrees of cognitive load. Human Factor 42 (3), 470–481.

Kazagli, E., Bierlaire, M., Flotterod, G., 2016. Revisiting the route chocie problem: A modelling framework based on mental representation. Journal of Choice Modelling 19 (2016) 1-23.

Keane, M., 1994. A computationally cficient practical simulation estimator for panel data, with applications to estimating temporal dependence in employment and wages. Econometrica 62. 95-l 16.

Keane, M., 2006. The generalized logit model: preliminary ideas on a research program. presentation at Motorola-CenSoC Hong Kong meeting, October 22, 2006.

Keane, M., Wasi, N., 2013. Comparing alternative models of heterogeneity in consumer choice behavior. J. Appl. Econ. 28 (6), 1018–1045.

Kitthamkesorn, S., Chen, A., 2013. A path-size weibit stochastic user equilibrium model. Transportation Research 57B, 378–397.

Kitthamkesorn, S., Chen, A., 2014. Unconstrained weibit stochastic user equilibrium model with extensions. Transportation Research 59B, 1–21.

Kitthamkesorn, S., Chen, A., Xu, X., 2015. Elastic demand with weibit stochastic user equilibrium flows and application in a motorised and non-motorised network, Transportmetrica A: Transport Science, 11:2, 158-185, DOI: 10.1080/23249935.2014.944241

Kohli, R., Jedidi, K., 2015. Error theory for elimination by aspects. Operations Research, Vol. 63, No. 3, 512-526.

Koppelman F.S., Sethi, V., 2000. Closed form discrete choice models. Handbook of Transport Modelling, Pergamon Press, January 2000.

Koppelman F.S., Wen, C.H., 2000. The paired combinatorial logit model: properties, estimation and application. Transportation Research Part B 34 (2), 75–89.

Lerman, S.R., Manski, C., 1977. An estimator for the multinomial probit model, 56th Meeting Transp. Res. Board (unpublished), cited in Daganzo (1979).

Leurent, F. M., 1997. Curbing the computational difficulty of the logit equilibrium assignment model. Transportation Research Part B: Methodological 31(4), 315–326.

Lisco, T., 1967. The value of commuter's travel time: A study in urban transportation, Ph.D. dissertation, Dept. of Economics, University of Chicago, Chicago, Illinois.

Loomes, G., Sugden, R., 1982. Regret-Theory: An alternative theory of rational choice under uncertainty. The Economic Journal, Vol. 92, No. 368, pp. 805-824.

Louviere, J.J., Meyer, R., Bunch, D.S., Carson, R., Dellaert, B., Hanemann, W.M., Hensher, D., Irwin, J., 1999. Combining sources of preference data for modelling complex decision processes. Marketing Letters 10:3 (1999), 205-217.

Louviere, J.J., Hensher, D.A., Swait, J.F., 2000. Stated Choice Methods and Analysis. Cambridge University Press, Cambridge.

Louviere, J.J., Train, K., Ben-Akiva, M.E., Bhat, C., Brownstone, D., Cameron, T.A., Carson, R.T., DeShazo, J.R., Fiebig, D., Greene, W., Hensher, D., Waldman, D., 2005. Recent Progress on endogeneity in choice modelling. Marketing Letters, December 2005, Volume 16, Issue 3-4, pp. 255-265.

Luce, R. D., 1959. Individual Choice Behavior. New York, NY: John Wiley & Sons, Inc.

Luce, D., Suppes, P., 1965. Preferences, utility and subjective probability, in R. Luce, R. Bush, and E. Galanter, eds., Handbook of Mathematical Psychology, John Wiley and Sons, New York, pp. 249–410.

Magidson, J., Vermunt, J., 2005. Technical Guide to Latent Gold Sofrware 4.5. Statistical Innovation.

Maher, M.J., Hughes, P.C., 1997. A probit-based stochastic user equilibrium assignment model. Transportation Research Part B, 31 (4), 341-355.

Mai, T., Fosgerau M., Frejinger E., 2015. A nested recursive logit model for route choice analysis. Transportation Research Part B: Methodological, 75, 100-112.

Mai, T., 2016. A method of integrating correlation structures for ea generalized recursive route choice model. Research Part B: Methodological, 93, 146-161.

Mai, T., Bastin, F., Frejinger, E., 2017. On the similarities between random regret minimization and mother logit: The case of recursive route choice models. Journal of Choice Modelling 23 (2017) 21-33.

Manski, C., Lerman S., 1977. The Estimation of Choice Probabilities from Choice-Based Samples. Econometrica 45, 1977-1988 (1977).

Marschak, J., 1960. Binary choice constraints on random utility indications, in K. Arrow, ed., Stanford Symposium on Mathematical Methods in the Social Sciences, Stanford University Press, Stanford, CA, pp. 312–329.

Marzano, V., Papola, A., 2004. A path-multilevel logit model for route choice which allows implicit path enumeration. Proceedings of 2004 ETC Conference, Strasbourg.

Marzano, V., 2006. Sviluppi teorici sui modelli di simulazione della scelta del percorso, Ph.D. tesi. University of Naples Federico II.

Marzano, V., Papola, A., 2008. On the covariance structure of the Cross-Nested Logit model. Transportation Research Part B 42 (2), 83-98.

Marzano, V., Papola, A., Simonelli, F., Vitillo, R., 2013. A practically tractable expression of the covariances of the Cross-Nested Logit model. Transportation Research Part B 57, 1-11.

Marzano, V., 2014. A simple procedure for the calculation of the covariances of any Generalized Extreme Value model. Transportation Research Part B 70, 151-162.

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

McFadden, D., 1975. On Independence, Structure, and Simultaneity in Transportation Demand Analysis. Working Paper No. 7511, Urban Travel Demand Forecasting Project, Institute of Transportation and Traffic Engineering, University of California, Berkeley.

McFadden, D., Reid, F., 1975. Aggregate travel demand forecasting from disaggregated behavioral models. Transportation Research Record: Travel Behavior and Values 534: 24±37.

McFadden D., 1978. Modelling the choice of residential location. Karlquist, A. et al. (eds.), Spatial interaction theory and residential location, North-Holland, Amsterdam, pp. 75-96.

McFadden, D., 1979, Quantitative methods for analyzing travel behavior of individuals: Some recent developments, in: D. Hensher and P. Stopher, eds., Behavioral travel modeling (Croom Helm, London), 279-318.

McFadden D, 1981. Econometric models of probabilistic choice. Manski C, McFadden D, eds. Structural Analysis of Discrete Data with Econometric Applications (MIT Press, Cambridge, MA), 198–272.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57, 995-1026.

McFadden, D., Train, K., 2000. Mixed MNL models of discrete response, Journal of Applied Econometrics **15**, 447–470.

McFadden, D., 2001. Economic choices. The American Economic Review, 91(3), 351e378.

McLachlan, G.J., Basford, K.E., 1988. Mixture Models: Inference and Applications to Clustering, Marcel Dekker, Inc., New York.

Mehndiratta, S., 1996. Time-of-day effects in inter-city business travel. PhD Thesis, University of California, Berkeley.

Nakayama, S., Chikaraishi, M., 2015. Unified closed-form expression of logit and weibit and its extension to a transportation network equilibrium analysis. Transportation Research Part B, 81 (2015), 672-685.

Newman, J.P., 2008. Normalization of network generalized extreme value models. Transportation Research Part B 42 (10), 958-969.

Ortuzar, J.d.D., 1983. Nested logit models for mixed-mode travel in urban corridors, Transportation Research A 17, 283–299.

Ortuzar, J.d.D., 2001. On the development of the nested logit model. Transportation Research Part B 35 (2001), 213-216.

Oviedo, J.L., Yoo, H.I., 2017. A latent class nested logit mdoel for ranked-ordered data with application to Cork Oak reforestation. Environ. Resource Econ. (2017) 68:1021-1055.

Oyama, Y., Hato, E., 2017. A discounted recursive logit model for dynamic gridlock network analysis. Transportation Research Part C: Emerging Technologies 85, 509–527.

Papola, A., 2000. Some developments of the cross-nested logit model. Proceedings of the 9th IATBR Conference.

Papola, A., 2004. Some developments of the cross-nested logit model. Transportation Research Part B 38 (12), 833-851.

Papola, A., Marzano, V., 2004. Applicabilità del modello hybrid Logit-Probit alla simulazione della scelta del percorso, in Proceedings of the Annual SIDT Meeting, Reggio Calabria 2004, Franco Angeli (ed.) (in Italian).

Papola, A., Marzano, V., 2013. A network generalized extreme value model for route choice allowing implicit route enumeration. Comput.-Aided Civ. Infras- truct. Eng. 28 (8), 560–580.

Papola, A., 2016. A new random utility model with flexible correlation pattern and closed-form covariance expression: The CoRUM. Transportation Research Part B: Methodological, 94, 80-96.

Papola A., Tinessa F., Marzano V., 2018. Application of the Combination of Random Utility Models (CoRUM) to route choice. Transportation Research Part B (111), 304-326.

Powell, W., Sheffi, Y., 1982. The Convergence of Equilibrium Algorithms with Predetermined Step Sizes. Transportation Science, 16(1):55-65.

Prashker, J.N., Bekhor, S., 1998. Investigation of stochastic network loading procedures. Transportation Research Record 1645, 94-102.

Prashker, J.N., Bekhor, S., 2004. Route choice models used in the stochastic user equilibrium problem: a review. Transport reviews, 24(6), 437-463.

Prato, C.G., Bekhor, S., 2006. Applying branch-and-bound technique to route choice set generation. Transportation Research Record: Journal of the Transportation Research Board, 1985, Transportation Research Board of the National Academies, Washington, D.C.,19–28.

Prato, C.G., Bekhor, S., 2007. Modeling route choice behavior: How relevant is the composition of choice set? Transportation Research Record, Issue 2003, 2007, 64-71.

Prato, C.G., 2009. Route choice modelling: past, present and future research directions. Journal of Choice Modelling 2(1), 65-100.

Prato, C.G., 2014. Expanding the applicability of random regret minimization for route choice analysis. Transportation, 41:351-375.

Ramming, M.S., 2001. Network Knowledge and Route Choice. PhD thesis, Massachusetts Institute of Technology.

Rao, C., 1973. Linear Statistical Inference and Its Applications, John Wiley and Sons, New York.

Redner, R.A., Walker, H.F., 1984. "Mixture densities, Maximum Likelihood, and the EM Algorithm", *SIAM Review* 2, 195-239.

Revelt, D., Train, K., 1998. Mixed logit with repeated choices. Reviewof Economics and Statistics 80, 647–657.

Revelt, D., 1999. Three discrete choice random coefficients papers and one police crime study. PhD Thesis, University of California, Berkeley.

Revelt, D., Train, K., 2000. Customer-specific taste parameters and mixed logit', Working Paper No. E00-274, Department of Economics, University of California, Berkeley.

Roe, R.M., Busemeyer, J.R., Townsend, J.T., 2001. Multialternative decision field theory: a dynamic connectionst model of decision making.. Psychol. Rev. 108 (2), 370.

Russo, F., Vitetta, A., 2003. An assignment model with modified Logit, which obviates enumeration and overlapping problems. Transportation, 30, 177-201.

Rust, J., 1987. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. Econometrica: Journal of the Econometric Society pp. 999–1033.

Sagebiel, J., 2011. Comparing the Latent Class Model with the Random Parameters Logit: A Choice Experiment Analysis of Highly Heterogeneous Electricity Consumers in Hyderabad, India. Paper presented at the International Choice Modelling Conference, July 04-06, 2011, Leeds, UK.

Sener, I.N., Pendyala, R.M., Bhat, C.R., 2011. Accomodating spatial correlation across choice alternatives in discrete choice models: an application to modelling residential location choice behaviour. Journal of Transport Geography 19 (2011), 294-303.

Sasaki, T., 1965. Theory of traffic assignment through absorbing markov process. Transactions of the Japan Society of Civil Engineers 1965(121), 28–32.

Shampine, L.F., 2008. "Vectorized Adaptive Quadrature in MATLAB" Journal of Computational and Applied Mathematics, 211, 2008, pp.131–140.

Sheffi, Y. and Powell, W.B., 1982. An algorithm for the equilibrium assignment problem with random link times. Networks, 12(2), 191–207.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice-Hall.

Shen, J., 2009. Latent class model or mixed logit model? A comparison by transport mode choice data, Applied Economics, 2915-2924.

Siikamaki, J., 2001. Discrete choice experiments valuing biodiversity conservation in Finland. PhD Thesis, University of California, Davis.

Siikamaki, J., Layton, D., 2001. Pooled models for contingent valuation and contingent ranking data: Valuing benefits from biodiversity conservation. Working Paper, Department of Agricultural and Resource Economics, University of California, Davis.

Simon, H.A., 1955. A behavioral model of rational choice. Q. J. Econ. 69, 99–118.

Simon, H.A., 1956. Rational choice and the structure of the environment. Psychol. Rev. 63, 129–138.

Simon, H.A., 1978. Rationality as Process and as Product of Thought. American Economic Review, May 1978 (Papers and Proceedings), 68(2), pp. 1–16.

Small, K.A., 1987. A discrete choice model for ordered alternatives. Econometrica 55 (2), pp. 409-424.

Stern, S.. 1992. A method for smoothing simulated moments of discrete probabilities in multinomial probit models. Econometrica 60, 943-952.

Swait, J., Ben-Akiva, M.E., 1987*a*. Incorporating random constraints in discrete choice models of choice set generation. Transportation Research B, 21, 91-102.

Swait, J., Ben-Akiva, M.E., 1987*b*. Empirical Test of a Constrained Choice Discrete Model: Mode Choice in Sao Paulo, Brazil Transportation Research, Vol. 21B, No. 2, 103-116.

Swait, J., 1994. A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. J. Retail Consum. Serv. 1 (2), 77–89.

Swait, J., Bernardino, A., 2000. Distinguishing taste variation from error structure in discrete choice data. Transp. Res. Part B 34 (1), 1–15.

Swait, J., 2001. Choice Set Generation within the Generalized Extreme Value Family of Discrete Choice Models. Transp. Res. Part B, 35 (2001), 643-666.

Swait, J., 2003. Flexible Covariance Structures for Categorical Dependent Variables through Finite Mixtures of Generalized Extreme Value Models. Journal of Business & Economic Statistics, 21 (1), 80-87.

Swamy, P., 1971. Statistical Inference in Random Coefficient Regression Models, Springer-Verlag, New York,..

Talvitie, A., 1972. Comparison of probabilistic modal-choice models. Highway Research Board Record 392: 111±120.

Taussig, F., 1912. Principles of economics. New York: Macmillan.

Teye, C., Davidson, P., Culley, R., 2014. Simultaneously accounting for inter-alternative correlation and taste heterogeneity among long distance travellers using mixed nested logit (MXNL) model so as to improve road toll traffic and revenue forecast. 41st European Transport Conference 2013, ETC 2013. Transportation Research Procedia 1 (2014) 24-35.

Thurstone, L. L., 1927. A Law of Comparative Judgment. Psychological Review, 34, pp. 273–86.

Tinessa, F., Papola, A., Marzano, V., 2017. The importance of choosing appropriate random utility models in complex choice contexts, M.T.-I.T.S. Conference Proceedings, 26-28 June 2017, Naples.

Train, K., 1995. Simulation methods for probit and related models based on convenient error partitioning. Working Paper No. 95-237, Department of Economics, University of California, Berkeley.

Train, K., 2001. A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit. Working Paper, Department of Economics, University of California, Berkeley.

Train, K., Sonnier, G., 2005. Mixed logit with bounded distributions of correlated partworths, in R. Scarpa and A. Alberini, eds., Applications of Simulation Methods in Environmental and Resource Economics, Springer, Dordrecht, pp. 117–134.

Train, K., 2008a. EMalgorithms for nonparametric estimation of mixing distributions. Journal of Choice Modelling 1, 40–69.

Train, K., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.

Train, K., 2016. Mixed logit with a flexible mixing distribution. Journal of Choice Modelling 19, 40–53.

Tsallis, C., 1994. What are the numbers that experiments provide? Quimica Nova 17, 468–471.

Tversky A, 1972*a*. Elimination by aspects: A theory of choice. Psych. Rev. 79(4):281–299.

Tversky A, 1972*b*. Choice by elimination. J. Math. Psych. 9(4):341–367

Umarov, S., Tsallis, C., Steinberg, S., 2008. On a q-central limit theorem consistent with nonextensive statistical mechanics. Milan Journal of Mathematics 76, 307–328.

Van der Zijpp, N.J., Catalano, F.S., 2005. Path enumeration by finding the constrained k-shortest paths. Transportation Research Part B: Methodological, 39 (6), 545-563.

Van Vliet, D., 1981. Selected node-pair analysis in Dial's assignment algorithm. Transportation Research Part B: Methodological 15(1), 65–68.

Vij, A., Carrel, A., Walker, J.L., 2011. Latent modal preferences: behavioral mixture models with longitudinal data. In: Proceedings of the Second International Choice Modeling Conference. Leeds, UK.

Vij, A., Walker, J.L., 2016. How, when and why integrated choice and latent variable models are latently useful. Transportation Research Part B, 90 (2016), 192-217.

Vij, A., Krueger, R., 2017. Random taste heterogeneity in discrete choice models: Flexible nonparametric finite mixture distributions. Transp. Res. Part B 106 (2017), 76-101.

Vovsha, P., 1997. The Cross-Nested Logit Model: Application to Mode Choice in the Tel-Aviv Metropolitan Area. Proceedings of the 76th TRB Meeting.

Vovsha, P., Bekhor, S., 1998. Link-Nested Logit model of route choice. Overcoming route overlapping problem. Transportation Research Record 1645, 133-142.

Walker, J.L., 2002. The mixed logit (or logit kernel) model: dispelling misconceptions of identification. Transportation Research Record 1805: 86–98.

Walker, J.L., Ben-Akiva, M.E., Bolduc, D., 2007. Identification of parameters in normal error component logit-mixture (NECLM) models. Journal of Applied Econometrics 22, 1095–1125.

Walker, J.L., Li, J., 2007. Latent lifestyle preferences and household location decisions. J. Geogr. Syst. 9 (1), 77-101.

Wen, C.H., Koppelman, F.S., 2001. The generalized nested logit model. Transportation Research Part B 35 (7), 627-641.

Wen, C.-H., Chou, P.L., Hong, L.Y., Chen, T.N., 2011. A latent class nested logit model of intercity bus carrier choice. Journal of the Eastern Asia Society for Transportation Studies, 9, 382–394.

Wen, C.H., Huang, W.W., Fu, C., C., P.Y., 2013. A latent class generalized nested logit model and its application to modelling carrier choice with market segmentation. Transportmetrica A: Transport Science. Volume 9, 2013, Issue 8.

Westin, R., 1974. Predictions from binary choice models. Journal of Econometrics 2: 1±16.

Westin, R., Gillen, D., 1978. Parking location and transit demand: a case study of endogenous attributes in disaggregate mode choice models. Journal of Econometrics 8: 75±101.

Williams, H.C.W.L., 1977. On the formation of travel demand models and economic evaluation measures of user benefit. Environment and Planning A 9, 285-344.

Williams, H.C.W.L., Ortúzar, J.d.D., 1982. Behavioural theories of dispersion and the mis-specification of travel demand models. Transportation Research 16B, 167-219.

Xu, X., Chen, A., 2013. C-logit stochastic user equilibrium model with elastic demand. Transportation Planning and Technology, 36:5, 463-478, DOI: 10.1080/03081060.2013.818275

Yai, T., Iwakura, S., and Morichi, S., 1997. Multinomial probit with structured covariance for route choice behavior. Transportation Research Part B: Methodological, 31(3):195–207.

Ye, X., Garikapati, V.M., You, D., Pendyala, R., 2017. A practical method to test the validity of the standard Gumbel distribution in logit-based multinomial choice models of travel behaviour. Transportation Research Part B: Methodological, 106, 173-192.

Yuan, Y., You, W., Boyle, K.J., 2015. A guide to heterogeneity features captured by parametric and nonparametric mixing distributions for the mixed logit model. 2015 AAEA & WAEA Joint Annual Meeting *July 26-28* (No. 205733), Agricultural and Applied Economics Association & Western Agricultural Economics Association.

Zhou, Z., Chen, A. and Behkor, S., 2012. C-logit stochastic user equilibrium model: formulations and solution algorithm. Transportmetrica, 8(1), 17-41.