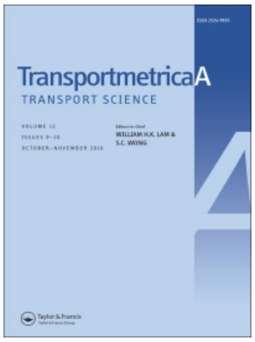

Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttra21

Mode choice in strategic freight transportation models: a constrained Box–Cox meta-heuristic for multivariate utility functions

Bart Jourquin

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2021.1937375

Taylor & Francis
Taylor & Francis Group

Mode choice in strategic freight transportation models: a constrained Box–Cox meta-heuristic for multivariate utility functions

Bart Jourquin

CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium

ABSTRACT

Modal choice models used for freight transportation studies covering inter-regional or international areas are difficult to set up because of the dearth of information about explanatory factors. While cost and transit time are known as being important explanatory variables, they are generally correlated to each other, and their coefficient computed with a Logit model can have unexpected signs.

Box-Cox transformations (BCT) of the independent variables can help to overcome this problem. If solutions to identify the BCT parameter that maximises the likelihood of a model are well known, the process is not straightforward once it must respect the constraints that the variables' coefficient estimators take the expected signs.

This paper presents a shotgun hill climbing meta-heuristic with backtracking capabilities, able to quickly identify Box-Cox λ parameters to use when multiple variables must be transformed. The algorithm appears to be efficient and effective and produces stable and statistically valid solutions.

ARTICLE HISTORY
Received 15 October 2020
Accepted 26 May 2021

KEYWORDS
Freight transport; modal choice; Multinomial Logit; Box–Cox transforms; hill climbing; heuristic

1. Introduction

Modal choice models are important tools supporting transport policy decisions. Used to compute modal elasticities for instance, they can be very useful in cost–benefit or traffic impact analysis for planned infrastructure (de Jong, Gunn, and Ben-Akiva 2004; Liedtke and Carrillo Murillo 2012; Beuthe et al. 2014; Rothengatter 2019). When applied to large inter-regional or international areas, they mostly rely on basic explanatory variables, such as transportation costs, transit times or trip lengths.

Other variables, related to the level of service, such as safety, flexibility, frequency, losses during transport or reliability may also play a role in decision-making. Their relative importance varies with the type of transport considered (Beuthe and Bouffioux 2008; Arencibia et al. 2015) but cost and transit time are often cited among the most important explanatory variables (Cullinane and Toy 2000), while trip length is deemed to be implicitly included in

CONTACT Bart Jourquin bart.jourquin@uclouvain.be CORE, Université catholique de Louvain, Louvain-la-Neuve B-1348, Belgium

these two variables. Note that, in the context of passenger transport, Gaudry and de Lapparent (2015) have attempted to include distance as an additional explanatory variable which, they suggest, takes into account a specific 'attitude to distance' distinct from the 'attitude to time or cost'. In the case of freight transportation, we may speculate that this variable could be taken as a proxy for the increased probability of an incident in the course of a longer transport which affects the security and safety of the cargo as well as the transport reliability, factors of importance for the shippers and consignees. While this is an issue of some interest worth mentioning, it will not be deeply investigated here because of lack of additional information needed for a good assessment of this hypothesis. Thus, the focus of this paper is put on the technique of estimating a freight transport Logit model with one, two or three variables (cost, transit time and distance) that are, each of them and separately, Box–Cox transformed. The combination of the three variables is tested in order to assess the robustness of the proposed heuristic.

Fortunately, transportation costs, transit times and trip lengths are also among the figures that can be gathered when a model covers extended inter-regional or international areas with large numbers of origins and destinations. However, these independent variables may only poorly explain mode choice decisions as they are often correlated, which is problematic. Indeed, the interpretation of a regression coefficient is that it represents the mean change in the dependent variable for each unit change in an independent variable when all of the other independent variables remain constant. When independent variables are correlated, changes in one variable are associated with shifts in other variables. In such cases, the coefficient estimates are unstable: their value and even their sign 'swing' between the independent variables. Moreover, they also may have a weak statistical power (Greene 2012; Adeboye, Fagoyinbo, and Olatayo 2014).

Yet, at the level of aggregation considered in this paper, if one considers the simple economic concept of own elasticity of demand, the coefficient estimators for cost and transit time should be negative as well documented in the literature review (Beuthe, Jourquin, and Urbain 2014; Jourquin and Beuthe 2019). Similarly, it can be expected that a longer transport distance would have generally a negative impact on the attractiveness (and thus the probability of choice) of a transportation mode.

Several methods can be found in the literature to overcome multicollinearity. The aim of this paper is not to present an extensive review of these methods, but to put the focus on one line of thought, based on Box–Cox transforms (Box and Cox 1964) of independent variables. Beside the fact that this technique can be useful to improve the Log-Likelihood of a model, it is sometimes considered as an elegant and efficient way to help overcome multicollinearity and to re-establish the expected signs of the estimators (Fridstrøm and Madslien 1994; Gaudry 2016). The Box–Cox transformation (BCT, equation 1) belongs to the family of power transforms, used to create a monotonic transformation of data using power functions. The chosen value for λ offers a lot of flexibility.

$$\mathrm{BC}(x,\lambda) = \begin{cases} \frac{x^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log(x), & \text{if } \lambda = 0 \end{cases} \tag{1}$$

By 'bending' the functional form, BCT provides more degrees of freedom for model estimation. For instance, transforming a linear specification into a non-linear function can be used to be in line with more sophisticated theories such as non-constant values of time.

In order to estimate the optimal value of the λ parameter(s), toolboxes using different concepts such as profile likelihood, Bayesian statistics or Newton–Raphson greedy algorithms are most often used (Sakia 1992; Bierlaire 2003; Ishak and Ahmad 2018; Soleymani 2018). Simulated annealing was also used for the maximum-likelihood estimation (Robert and Casella 2004), and neural networks have been implemented for a fast identification of Box–Cox transformation parameters (Hong 2006). Unfortunately, to the best of our knowledge, none of these methods explicitly take care of the expected sign of the estimators of the independent variables (and of their level of significance).

There is thus a need to propose a method that efficiently identifies a set of λ transform parameters that explicitly integrates the constraints related to the expected signs of the estimators and their level of significance. The latest constraint is also important as the Logit model with BCT explanatory variables can for instance be used to compute cost or transit time elasticities. In such a context, the level of significance of the estimators cannot be ignored to assess the robustness of the derived elasticities.

Obviously, a brute-force approach can be used, which tests all the possible combinations of lambdas in a given range and for a given step (granularity). However, the number of combinations to test growth exponentially with the number of variables that must be Box–Cox transformed. For instance, if three variables must be transformed, 68,921 runs are needed to identify the corresponding three lambdas when they are searched in the $[-2, +2]$ range with a step of 0.1, which is a rather classical range and granularity (Ishak and Ahmad 2018; Soleymani 2018).

The main contribution of this paper is a heuristic that can quickly identify a good combination of λ's when several independent variables must be Box–Cox transformed. The heuristic provides, when possible, estimators for the transformed variables that have the expected sign and have a least the level of significance chosen by the modeller. As the introduction of these constraints in the traditional max Log-Likelihood paradigm maybe questionable, attention is paid to the statistical validity of the identified solutions.

The problem to solve and the proposed heuristic are illustrated using a large real-World dataset.

After a presentation of the data and the modal choice model in sections 2 and 3, section 4 illustrates the benefits of using Box–Cox transforms.

Section 5 explains, demonstrates and discusses the results of a meta-heuristic, which clearly breaks the combinational nature of the problem. It can be defined as belonging to the family of shotgun hill climbing algorithms with backtracking capabilities (Russell and Norvig 2003; Witten, Frank, and Hall 2011; Christian and Griffiths 2016).

2. Dataset description

The dataset used in the context of this paper is gathered from the ETISPlus European FP7 research programme. It provides an information system useful for assessing European transport policies: it combines data, analytical modelling with maps and an online interface for accessing the data (Szimba et al. 2012). Public access to several deliverables and many data is provided, among which origin-destination matrixes and digitised networks. For this paper, the Origin-Destination (OD) matrixes for the year 2010, aggregated at the NUTS-2 regional level are used. These matrixes contain annual transported tons for each OD pair, and are available for road, railway and inland waterways (IWW) transport, with figures for 10 categories of commodities (NST-R chapters[1] 0–9). It is important to note that these matrixes are based on observations (statistics) but are partially constructed using models and tools for calibration and validation.

This dataset has been imported into an open-source transportation network model (Nodus, http://nodus.uclouvain.be).

As illustrated in Table 1, the datasets corresponding to the different categories of commodities are quite different in sizes (number of OD cells), geographical spread (average trip length), volumes (average t.km per OD cell) and modal split (market shares). This diversity is useful to test the robustness of the heuristic that will be presented later.

Nodus allows retrieving, for each OD pair, each mode and each group of commodities, the loading, unloading, transit and transhipment costs. The cost functions C takes into account all the carrier's costs during transport: labour, capital invested in vehicles, fuel, maintenance, insurance, handling and storage costs, services directly linked to a transport, plus all residual indirect costs like those of administrative services. Beside these costs, the software also provides the total transit times T (travel time + loading and unloading duration) and trip length L. Actually, the C, T and L values are gathered from the assignment of each modal OD matrix on its corresponding digitised network. Thus, for each OD relation and each group of commodities, C, T and L values are computed for each available mode.

The loading and unloading costs *ld_cost* and *ul_cost* are fixed costs, though they vary with the mode and the transported goods. The transhipment costs associated with transport chains are not explicitly modelled in this exercise. The loading factors of the vehicles are taken from the ECCONET research project (Beuthe et al. 2014). They are exogenous but specific for each group of commodities and type of vehicle (truck, train or one of the 6 types of barges included in the model). The travelling unit cost or moving cost, *mv_cost,* depends on the length and on the average commercial speed for the considered mode, as well as on the transported commodity g. For a given link l belonging to a network of mode m, it is computed as:

$$mv\_cost^{g}_{l,m} = \frac{Average\ speed_{m}}{Speed_{l,m}} * length_{l} * unit\ mv\_cost^{g}_{m} \tag{2}$$

As the unit *mv_cost* also contains time-related costs, the *Average speed / Speed* ratio allows for taking into account higher/lower than average costs on slower/faster segments of the

Table 1. Content of the demand matrixes.

	OD cells			Avg. trip length (km)			Avg. flow (1000 t.km)			Market share (%)		
NST-R	Road	IWW	Rail	Road	IWW	Rail	Road	IWW	Rail	Road	IWW	Rail
0	14,586	4741	14,519	805	948	877	7561	2308	1039	87.6	4.6	7.8
1	15,195	5085	15,119	795	994	868	8495	1940	258	93.7	4.6	1.8
2	3773	1725	3759	412	743	460	2616	6966	14,677	19.4	13.8	66.8
3	6138	2356	6096	514	763	568	6936	8927	2363	67.4	23.2	9.4
4	6482	2642	6441	460	681	511	17,528	3317	5171	84.6	4.6	10.9
5	11,778	4380	11,731	701	972	766	5003	1096	1775	77.9	4.5	17.6
6	12,913	4481	12,826	635	903	698	17,913	6344	5438	79.1	7.5	13.4
7	4622	2298	4604	476	781	525	2148	969	5913	32.9	4.3	62.8
8	13,608	5106	13,571	737	992	804	6310	2502	1521	80.0	9.7	10.4
9	25,399	6725	25,250	976	1163	1056	13,623	2647	1651	86.5	5.5	8.0

network. Indeed, $Average\ Speed_m$ represents the average speed for mode m on the total network and $Speed_{l,m}$ is the average speed on link l.

The total cost C_m^g of a route between an origin and a destination for a vehicle of type m transporting commodities of type g is thus equal to:

$$C_m^g = ld\_cost_m^g + ul\_cost_m^g + \sum_l^L mv\_cost_{l,m}^g, \tag{3}$$

where L is the set of successive links representing the route.

Similarly, the total transit time has fixed elements (the loading and unloading duration ($ld\_duration$ and $ul\_duration$) and a variable part (the travel duration that depends on length and speed on the successive links along the route). Thus:

$$T_m^g = ld\_duration_m^g + ul\_duration_m^g + \sum_l^L mv\_duration_{l,m}, \tag{4}$$

with

$$mv\_duration_{l,m} = length_l / speed_{l,m}. \tag{5}$$

The exact nature of this dataset (partially constructed OD matrixes containing aggregated regional yearly transported tonnages and computed values for the independent variables) must be kept in mind as it has a clear impact on the variance. The presence of the length and speed variables in the definitions of C and T also explain why both variables are correlated.

3. Modal choice model and examples of problematic solutions

As outlined in the previous section, three independent variables, cost C, transit time T and trip length L, as well as the transport demand (OD matrixes) are available. As they are different for each group g of commodities, it is appropriate to estimate a separate model for each of these groups.

The explanatory variables being specific to each mode, but not to shippers, the analysis applies the McFadden's Conditional Logit model (McFadden 1973). Depending on which independent variables to include in the utility function, the model can be written as

$$Pr_m^g = \frac{\exp(\alpha^g C_m^g + \delta_m^g)}{\sum_{j=1}^n \exp(\alpha^g C_j^g + \delta_j^g)}, \tag{6}$$

$$Pr_m^g = \frac{\exp(\alpha^g C_m^g + \beta^g T_m^g + \delta_m^g)}{\sum_{j=1}^n \exp(\alpha^g C_j^g + \beta^g T_j^g + \delta_j^g)}, \tag{7}$$

$$Pr_m^g = \frac{\exp(\alpha^g C_m^g + \beta^g T_m^g + \gamma^g L_m^g + \delta_m^g)}{\sum_{j=1}^n \exp(\alpha^g C_j^g + \beta^g T_j^g + \gamma^g L_m^g + \delta_j^g)}, \tag{8}$$

where Pr_m^g is the probability to choose mode m when transporting commodity g, and n represents the number of modes in the choice set. Parameters α^g, β^g and γ^g are not mode specific (Conditional Logit). However, since the model is solved separately for each group of commodities g, these coefficients can vary from group to group. As a cost increase, a

Table 2. Example of problematic solutions.

NST/R	α^g		β^g	
0	−1.9607321	(***)	0.1071265	
1	−1.5512147	(***)	−1.1649947	(***)
2	−4.7085641	(***)	0.6887534	(***)
3	−0.5562238	(**)	−2.3806673	(***)
4	−3.0464042	(***)	−2.3464773	
5	−2.3983015	(***)	0.4403564	(***)
6	−0.3742473	(**)	−1.8106375	(***)
7	−2.8706081	(***)	−0.0156810	
8	−2.4279871	(***)	−0.1181572	
9	−2.1342931	(***)	0.6702142	(***)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

longer transit time or a longer itinerary would reduce the attractiveness of a transport mode we expect that these coefficients have negative values. Finally, δ_m^g are the calculated intercepts for each mode and group of commodities. As one mode must be considered as the reference mode (road transport in case presented in this paper), its δ_m^g is set to 0.

One could argue that, in some cases, positive estimators for the independent variables might be accepted. For instance, a company using moving stock may appreciate longer transit times. However, these cases are specific to transport tasks that can only be identified at a disaggregated level. This is not expected at the level of aggregation suitable for strategic models with annual transported volumes at a regional level, which is the case illustrated in this paper.

For OD matrixes containing aggregated volumes, a weighted Logit methodology is applied whereby the transported tonnages weight the mode choice observations in the Log-Likelihood functions (Rich, Holmblad, and Hansen 2009). The weighted Conditional Logit problem is solved using the 'mnLogit' R package(Hasan et al. 2019), a faster and parallelised version of the well-known mLogit R package (Croissant 2019).

As an example of the difficulties of estimation that can be encountered with such a data set, Table 2 illustrates a problematic set of solutions obtained for the bivariate case with a classical log transform applied to C and T. It appears that the β^g estimator has an unexpected positive sign for groups of commodities 0, 2, 5 and 9. Moreover, the same β^g estimator has the correct sign but is not significant at all for groups 4, 7 and 8.

Such problematic results find their origin in the nature of the dataset and the rather simple functional form of the utility functions. Unfortunately, as explained in the introduction, these are often the only explanatory variables that can be gathered for freight transport models that cover large geographic areas.

The next section explains how relevant Box–Cox transforms of these variables can help satisfying the assumption about the signs of the estimators. It also discusses the impact of the method on the statistical fit of the models.

4. Box–Cox parameters estimation: a brute force approach

The goal of the game is thus to identify the values of λ's that maximise the Log-Likelihood of the Logit model, considering two constraints:

(1) The estimators of the independent variables should have the expected sign.
(2) The estimators should have a minimal level of significance, chosen by the practitioner.

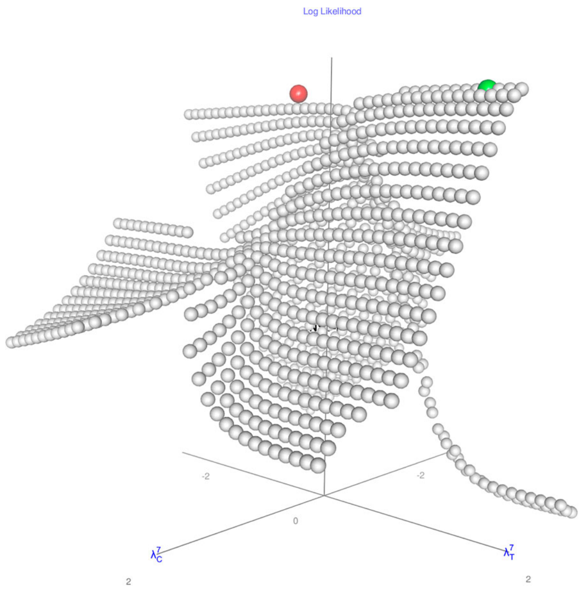

Figure 1. Log-Likelihoods of valid solutions and unconstrained max (bivariate model).

The values of C_m^g, T_m^g and L_m^g can be Box–Cox transformed before their respective estimators α^g, β^g, γ^g being computed. In order to identify the 'best' combination of λ's, one can use a simple 'brute force' approach, during which the transformations are performed using all the possible λ_C^g,λ_T^g and λ_L^g combinations taken in the range $[-2, 2]$ with a step of 0.1. For a given λ, 41 values are thus possible, which corresponds to the number of Logit models that must be computed in the case of the univariate utility function described in equation (6). For the bivariate case (equation 7), $41^2 = 1681$ Logit models must be computed, and 68,921 runs are needed for the utility function with three variables (equation 8). As the Logit models are separately solved for the 10 groups of commodities, the number of runs is obviously multiplied by 10. The number of runs becomes thus rapidly very large, and this is the reason why a heuristic will later be presented in this paper, which drastically reduces the λ's combinations to test while producing results that are equal or very close to the optimal ones identified by the brute force approach.

The impact of BCT on the Log-Likelihood of the Logit model is illustrated in Figure 1 for the bivariate case with C and T variables. All the possible combinations of λ_C^g and λ_T^g are tested and only the combinations that produce the expected (negative) sign for the estimators of the two independent variables are retained, but only if they are significant (at least at 0.05). Such combinations will henceforth further be referred to as 'valid' solutions. The Log-Likelihood of the Logit of each valid solution is represented along the vertical axis for the commodities of group NST/R 7.[2] The two horizontal axes represent the values of λ_C^2 and λ_T^2. Each grey dot represents a valid solution, among which the green dot is the one with the highest Log-Likelihood. The red dot represents the unconstrained max Log-Likelihood, i.e. the combination of λ's that maximises the Log-Likelihood but produces at least an estimator with the unexpected sign (invalid solution). The red and green dots are enlarged to make them more visible.

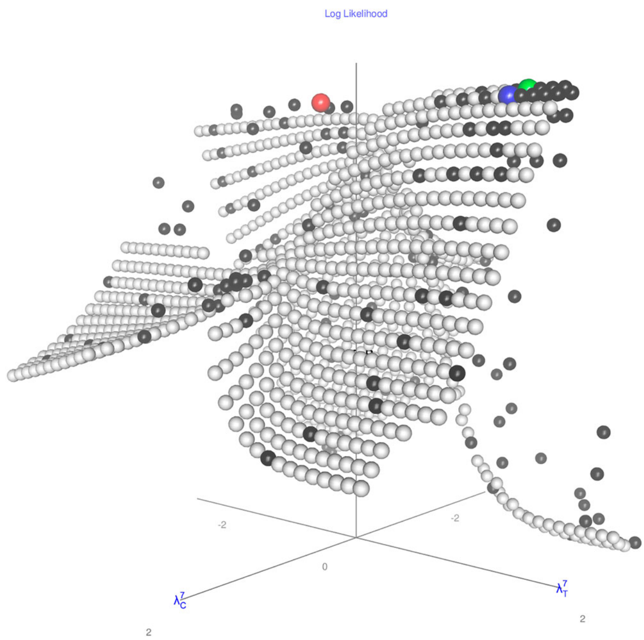

Figure 2. Set of solutions tested by the heuristic (bivariate model).

As a step of 0.1 is used between each λ, 1681 dots could be plotted but, excluding the red one, only 1075 are drawn. Indeed, only 1075 combinations of λ_C^2 and λ_T^2 result in valid solutions. All the other combinations produce either an unexpected sign or have at least one estimator that is not significant.

Figure 1 makes clear that no continuum exists between all the valid solutions. It also shows that the green and the red dots are not necessarily 'neighbours'. These two characteristics must be tackled by the heuristic presented in the next section, which task is to quickly converge towards the valid solution with the highest likelihood (the green dot). By definition, a heuristic doesn't guarantee that the optimal solution is found (as later illustrated by Figure 2). This could have an impact on the effectiveness of the algorithm and the statistical robustness of the solution. The fact that the optimal valid solution is surrounded by a series of valid solutions, whose LL's are practically identical, should, however, facilitate the work. These aspects will be thoroughly discussed in sections 5.2 and 5.3.

It is worth noting (Table 3) that, for the univariate and bivariate cases, all the λ's obtained for the optimal valid solutions are strictly within the recommended [−2, 2] range (Ishak and Ahmad 2018; Soleymani 2018), showing that there is no need to explore more 'extreme' values, that would exaggeratedly distort the data at the risk of suspecting the modeller of making arrangements so that his model sticks, at all costs, to his data.[3] It is also the case for the trivariate case with an exception for λ_T^2 and λ_T^7 which are equal to 2 (and may be larger than 2). Again, the trivariate case is only considered in this paper in order to test the robustness of the heuristic.

The comparison of the max Log-Likelihoods of the models without any BCT of the independent variables with those of the models with a constrained BCT of the variables is also

Table 3. Optimal λ's for 1, 2 and 3 independent variables.

	1	2		3		
NST/R	λ_C	λ_C	λ_T	λ_C	λ_T	λ_L
0	0.4	0.4	−1.0	0.5	−1.0	−0.1
1	0.2	0.1	−0.1	0.1	−0.2	1.7
2	0.2	0.4	1.1	0.7	2.0	1.1
3	0.4	0.9	0.7	0.8	0.0	1,7
4	0.2	0.0	0.1	−1.6	−0.4	0.5
5	0.2	0.5	1.1	0.5	1.1	0.6
6	0.5	0.7	0.3	0.8	0.3	0.5
7	0.1	0.1	1.7	0.2	2.0	−0.1
8	0,1	0.3	1.6	0.3	1.5	1.8
9	−0.6	0.2	1.2	0.2	1.2	1.7

Table 4. Log-Likelihood of the models with and without constrained BCT.

	Bivariate model			Trivariate model		
NST/R	No BCT	Constrained BCT	Difference	No BCT	Constrained BCT	Difference
0	(−17387.62)	−17157.17	1.33%	(−17257.66)	−17139.74	0.68%
1	−10482.26	−10103.74	3.61%	(−10443.95)	−10099.01	3.30%
2	−7722.79	−7577.96	1.88%	−7721.49	−7629.89	1.19%
3	−9980.87	−9965.90	0.15%	(−9904.08)	−10015.38	−1.12%
4	−7541.37	−7310.68	3.06%	(−7498.05)	−7362.88	1.80%
5	−20440.55	−20357.96	0.40%	−20381.83	−20273.65	0.53%
6	−20608.50	−20456.75	0.74%	−20608.01	−20454.92	0.74%
7	−10248.36	−10049.80	1.94%	(−10172.04)	−10032.11	1.38%
8	−22102.26	−21626.12	2.15%	−22096.47	−21603.30	2.23%
9	−32383.97	−32182.45	0.62%	(−32383.00)	−32174.97	0.64%

interesting to assess the statistical validity of the method. Indeed, one must avoid promoting an approach that imposes constraints to force a model to behave as expected without sufficient statistical control.

As shown in Table 4, the Log-Likelihood of the constrained models are, excepted for NST/R 3 in the trivariate model, always higher than those of the models without any BCT. Yet, in the constrained Box–Cox models, all the estimators have the expected sign and at least a level of significance of 0.05. This is, however, not the case for some of the non-BCT models: the figures between parenthesis correspond to models producing estimators with an unexpected sign, and the underlined ones identify models with at least one non-significant estimator. Note that the non BCT trivariare model for NST/R 3 corresponds anyway to a non-valid solution.

5. A simple but efficient meta-heuristic

A systematic test of all the values of λ for a Box–Cox transform of a single independent variable in the range [−2, 2] with a step of 0.1, i.e. 41 different values, is undoubtedly feasible. However, the exhaustive combinations of λ's in multivariate cases rapidly becomes a very long computing task because of the combinational nature of the problem. With the objective to drastically reduce the number of λ's combinations to test, this section proposes a simple, efficient and generic heuristic, useable for a combination of N λ's. It is tested for cases with $N = 2$ and 3.

5.1. Description

The basic idea of this heuristic is to explore the neighbourhood of a given combination of λ's in N dimensions, starting from an initial valid combination. Two characteristics previously outlined must be taken into account:

- The optimal valid combination cannot be considered as being located in the close neighbourhood of the best unconstrained combinations.
- The continuum of valid solutions is not guaranteed as they can be located in different 'clusters'.

A specific 'hill climbing' algorithm (Russell and Norvig 2003) was developed. It is a mathematical optimisation technique which belongs to the family of local searches. It is an iterative algorithm that starts from an arbitrary solution, and further tries to find better solutions by making incremental changes to the solution until no further improvement can be obtained. Since only convergence to a local maximum can be guaranteed, repeated alternative starting values are tested to have a better chance to locate the global (valid) maximum, and hence the maximum Log-Likelihood. This is usually referred to as a 'shotgun' or 'random-restart' hill climbing meta-heuristic (Christian and Griffiths 2016).

Generally, once all the solutions around an initial solution explored, the strategy used in the hill-climbing algorithms is to pursue the exploration towards the solution that presents the highest improvement (steepest climb). Such an approach implicitly considers that there exists a continuum between all the valid solutions. However, it already has been pointed out that this is not always the case for the problem discussed in this paper. Therefore, the algorithm tests all the solutions in each dimension and further explores all the solutions with the expected signs having a higher Log-Likelihood than the current solution. As multiple search paths are explored, the algorithm must 'remember' the solutions to (re)start from. Consequently, the algorithm belongs to the family of hill climbing with backtracking capabilities heuristics (Witten, Frank, and Hall 2011).

An important drawback of this strategy is that a same solution has a high probability to be encountered along several search paths. In order to avoid time consuming recomputing of already computed solutions, a hash table – a data structure that implements an associative array mapping keys (combinations of λ's) and values (solved model for these λ's) – is used to store all the already computed results. It is checked each time the result of a Logit model is needed during the search. The computing of a λ's specific Logit model is thus only performed when it is not yet present in the hash table, while the retrieval of an already computed solution is almost immediate.

Beside the definition and the initialisation of some global variables (Pseudo-code 1), the heuristic has two major phases: the identification of an initial combination of λ's to start from and the systematic exploration of its neighbourhood until no better solution is found.

To start with (Pseudo-code 2), a set of *nbDraws* combinations of λ's that produce valid solutions are randomly drawn, among which the solution with the highest Log-Likelihood is retained as starting solution for further exploration. Our experience shows that 5 random draws of an initial solution is enough when $N = 2$ and that 10 draws appear to be sufficient to obtain efficient results when $N = 3$ (see the section 5.2 below).

```
N                     number of lambas
range                 range to search λ's in (from -range to +range)
step                  interval between 2 successive values of λ
nbDraws               number of valid initial λ combinations to randomly draw
solutionsToExplore    list of solutions to explore around (empty)
bestSolution          best solution found so far (empty)
```

Pseudo-code 1. Definition of the global variables.

```
n = 0
while (n < nbDraws) {
  lambdas = random draw of a combination of N lambdas in range;
  solution = retrieveOrComputeLogit(lambdas);
  if (isValid(solution)) {
    n = n + 1;
    if (solution$logLik > bestSolution$logLik) {
      bestSolution = solution;
    }
  }
}
initialise solutionsToExplore with bestSolution;
```

Pseudo-code 2. Identify a good initial combination of λ'.

```
exploreAround(solution, step) {
  for each lambdas around(step) of the solution {
    newSolution = retrieveOrComputeLogit(lambdas);

    if (isValid(newSolution) and newSolution$logLik > bestSolution$logLik) {
      bestSolution = newSolution;
    }

    if (hasExpectedSigns(newSolution) and newSolution$logLik > solution$logLik) {
      add newSolution to solutionsToExplore;            }
    }
}
```

Pseudo-code 3. Details of the 'exploreAround' function.

```
repeat {
  solution = first element of solutionsToExplore;
  exploreAround(solution, step);
  remove solution from solutionsToExplore;
} until solutionsToExplore is empty
```

Pseudo-code 4. Explore the neighbourhood until no better solution is found.

During the exploration process, corresponding to the second phase of the heuristic, every solution with the expected signs (regardless of their level of significance) and having a higher Log-Likelihood than the one of the current start solution is added to the list of solutions to explore from later. All the encountered valid solutions are compared to the best one found so far and replace it if better (Pseudo-code 3).[4] Once all the solutions around

```
for each currentStep in (0.4, 0.2, 0.1) {
  repeat {
    solution = first element of solutionsToExplore;
    exploreAround(solution, currentStep);
    remove solution from solutionsToExplore;
  } until solutionsToExplore is empty

  initialise solutionsToExplore with bestSolution;
}
```

Pseudo-code 5. Explore the neighbourhood until no better solution is found.

the current starting point explored, it is removed from the list. Consequently, this list grows and shrinks dynamically, and the exploration ends when the list is empty, meaning that no better valid solution can be found (Pseudo-code 4).

The particularity of this hill climbing algorithm is that the exploration isn't limited towards the steepest direction, but that all the directions in which the slope increases are tested, as long as the signs of the estimators of the computed solutions are expected.

This strategy can further be optimised (Pseudo-code 5) using successive values for *step*, starting from a coarse value and ending with the final granularity of 0.1. Values of 0.4, 0.2 and 0.1 were used for the case presented in this paper. This strategy helps to rapidly converge towards the neighbourhood of an interesting (or even best) solution using large steps, then exploring the surrounding with gradually smaller steps.

The two phases are repeated several times (shotgun meta-heuristic) in order to try to locate the global maximum. Again, the use of a hash table helps to limit the number of Logits to compute as, from run to run, already computed solutions can be directly fetched.

Figure 2 gives an idea on how the heuristic finds its path to a solution. This example, for NST/R 7, has been chosen because the solution found by the heuristic (blue dot) doesn't correspond to the exact solution under constraint (green dot). All the black dots represent the solutions tested by the heuristic. Some seem to be in the middle of nowhere; they correspond to random draws of the first phase, that are not valid solutions. The density of black dots becomes higher in the neighbourhood of the final solution. In this example, only 146 Logit computations (black dots + the blue dot) were needed (instead of 1 641) to identify a solution, which Log-Likelihood is very close to the best solution (green dot).

5.2. Performance of the heuristic

This heuristic is implemented in an R script (Appendix 2) and its results are compared to the exact solutions computed for the dataset described earlier using the brute force approach. As the heuristic starts from a randomly drawn solution, it was run 50 times in order to measure its average and worst-case performances.

The next two tables give some performance indicators for the bivariate and trivariate cases.[5] For each group of commodities, one finds:

- The average, smallest (best-case) and largest (worst-case) number of Logit computations needed by the heuristic.

Table 5. Performances indicators for the bivariate model (with 5 shots).

NST-R	Logit computations (brute force = 1,681) Average	Best	Worst	Hits (50)	Log-Likelihood (Δ% with brute force optimal) Average Δ	Best-case Δ	Worst-case Δ
0	80	60	109	50	0%	0%	0%
1	66	44	122	38	0%	0%	−0.1%
2	84	65	127	49	0%	0%	0%
3	72	52	97	50	0%	0%	0%
4	63	45	85	11	−0.1%	0%	−0.2%
5	69	49	96	49	0%	0%	0%
6	67	51	97	50	0%	0%	0%
7	80	59	115	35	−0.1%	0%	−0.2%
8	78	51	132	47	−0.1%	0%	−0.9%
9	77	57	111	50	0%	0%	0%

Table 6. Performances indicators for the trivariate model (with 10 shots).

NST-R	Logit computations (brute force = 68,921) Average	Best	Worst	Hits (50)	Log-Likelihood (Δ% with brute force optimal) Average Δ	Best-case Δ	Worst-case Δ
0	704	609	821	50	0%	0%	0%
1	482	394	575	50	0%	0%	−1%
2	876	725	1082	1	−0.5%	0%	−1.2%
3	977	845	1126	50	0%	0%	0%
4	1391	1208	1592	3	−0.3%	0%	−0.8%
5	408	298	480	50	0%	0%	0%
6	538	464	611	50	0%	0%	0%
7	556	431	693	30	0%	0%	0%
8	448	281	548	48	0%	0%	−0.3%
9	669	587	795	32	0%	0%	0%

- The number of ‘hits’ (when the solution of the heuristic corresponds to the one identified by the brute-force algorithm).
- The average, best-case and worst-case differences (in %) between the Log-Likelihood of the solutions found by the heuristic and the those of the brute force approach.

When the heuristic is applied to the bivariate case (Table 5), it converges after an average of 74 Logit computations, i.e. about 4% of what is needed to obtain the exact solution with a brute force approach. Even if the final solution can differ from the exact one (cfr. the number of hits), its Log-Likelihood is always very close (a difference of less than 1%) to the best one, even for the worst cases.

Table 6 shows that, when applied to the trivariate case, the heuristic finds a solution after an average of 705 Logit computations, which represents only 1% of the runs needed to find the exact solution with the brute force algorithm. The heuristic clearly breaks the combinational logic of the problem. Nevertheless, the Log-Likelihoods of the solutions are very close to the best ones, even for the worst cases.

Knowing that, on a decent recent computer, the computing time of one Logit model for this dataset takes about 1 second, 20 hours are needed to solve the brute force trivariate case for one group of commodities. The heuristic converges after an average of 11 minutes and 45 seconds. It is thus efficient.

Table 7. Best-case parameters for the bivariate model.

				Intercepts				Estimators			
NST/R	Max LL	λ_C	λ_T	IWW		Rail		C		T	
0	−17157.2	0.4	−1.0	−3.138	***	−1.930	***	−0.453	***	−0.713	***
1	−10103.7	0.1	−0.1	−1.563	***	−3.013	***	−1.233	***	−1.150	***
2	−7578.0	0.4	1.1	−2.710	***	−0.944	***	−1.065	***	−0.003	***
3	−9965.9	0.9	0.7	−0.728	***	−1.128	***	−0.236	***	−0.103	***
4	−7310.7	0.0	0.1	1.373	**	2.334	***	−3.099	***	−2.017	***
5	−20358.0	0.5	1.1	−2.764	***	−0.021		−0.376	***	−0.007	***
6	−20456.7	0.7	0.3	2.302	***	2.017	***	−0.163	***	−0.730	***
7	−10049.8	0.1	1.7	−4.601	***	−1.300	***	−2.054	***	0.000	*
8	−21626.1	0.3	1.6	−3.365	***	−1.318	***	−1.047	***	−0.001	***
9	−32182.5	0.2	1.2	−1.418	***	−0.245	***	−0.687	***	−0.015	***

Table 8. Worst-case parameters for the bivariate model.

				Intercepts				Estimators			
NST/R	Max LL	λ_C	λ_T	IWW		Rail		C		T	
0	−17157.2	0.4	−1	−3.138	***	−1.930	***	−0.453	***	−0.713	***
1	−10111.9	0.2	0	−1.885	***	−3.343	***	−1.010	***	−0.797	***
2	−7578.0	0.4	1.2	−2.734	***	−0.968	***	−1.066	***	−0.002	***
3	−9965.9	0.9	0.7	−0.728	***	−1.128	***	−0.236	***	−0.103	***
4	−7328.5	−1.1	−0.2	3.529	***	4.547	***	−23.080	***	−5.010	***
5	−20365.9	0.4	0.8	−2.641	***	0.071		−0.492	***	−0.025	***
6	−20456.7	0.7	0.3	2.302	***	2.017	***	−0.163	***	−0.730	***
7	−10067.9	−0.1	−0.3	−4.303	***	−0.785	***	−3.355	***	−0.402	***
8	−21811.4	−0.1	−0.4	−3.263	***	−1.179	***	−2.818	***	−0.600	***
9	−32182.5	0.2	1.2	−1.418	***	−0.245	***	−0.687	***	−0.015	***

If solutions with less or nonsignificant parameter values can be valid, the question of the trade-off between significance and the overall likelihood value obtained by the different combinations of the lambda parameters arises. In order to objectify this point Table 7 gives, for the bivariate model, the best-case values (which also correspond to the brute force solution) of all the estimated parameters and their level of significance. If one except the significance level for the intercept for NST/R 5 commodities transported by rail, all the coefficients are at least significant at 0.01.

The corresponding figures for the worst cases can be found in Table 8. Apart from the transit time coefficient for NST/R 7, all the coefficients have the same level of significance in both tables, showing that almost no trade-off appears between level of significance and Log-Likelihood.

5.3. Model results

Once the Logit models using the Box–Cox transformed variables applied to the dataset, the computed modal split can be compared to the observations. Usually, some goodness of fit statistics is used to assess the validity of a model. An interesting review and discussion about this can be found in Parady, Ory, and Walker (2021). The authors conclude that the reliance on goodness-of-fit measures rather than validation performance is unwise, especially because transport models are mostly based on observational studies. Therefore a 'three-level' approach (Zhang 2013; Jourquin 2016) is applied here. The first level consists

Table 9. r computed at the OD level (all groups).

Mode	*r* (best case)	*r* (worst case)
Road	0.95	0.95
IWW	0.73	0.73
Rail	0.87	0.86

Table 10. r computed at the network level (all groups).

Mode	*r* (best case)	*r* (worst case)
Road	0.97	0.97
IWW	0.89	0.89
Rail	0.88	0.88

of a comparison between the observed and estimated global market shares. At the second level, the correlation coefficient *r* between the observed and estimated tonnages for each OD pair, each group of commodities and each mode is computed. Finally, an analysis of the *r*'s between the observed and computed volumes along the segments of the networks is presented. Note, however, that this three-level approach was proposed for validation purpose, i.e. when a model is not run on the same dataset the estimators were computed with. Therefore, the first level (comparison of the modal shares at the aggregated level) is useless and not presented here. It is indeed well known that a Logit model with a complete set of alternative-specific data always reproduce the sample market shares.

The figures presented in Tables 9 and 10 are obtained using the bivariate utility function (equation 7) with the Box–Cox transforms identified by the heuristic. More precisely, figures are given for the best (when the heuristic converges towards the optimal values identified by the brute force approach) and worst cases (the solution with the lowest Log-Likelihood identified by the heuristic among 50 runs).

To compute these figures, a specific modal choice module is developed for Nodus, which uses the best- and worst-case λ_C^g and λ_T^g and their corresponding α_C^g, α_T^g, δ_m^g estimators obtained from the Logit models. The results are gathered from the output of multimodal assignments of the OD matrixes[6] on the European networks, also performed by Nodus. The model provides the volume transported by mode *m* between each origin *O* and destination *D* and for each group of commodities *g*. This permits the computation of the correlation coefficients *r* between the calculated volumes and those found in the ETIS modal matrixes. The *r*'s appear in Table 9, and show that the results obtained using the best-case λ's are almost identical to those obtained when the 'worst' valid λ's are used.

The figures presented in Table 9 neglect the role that the network topology can play. As no observed count data is available along the segments of the networks, the results of separate assignments for each mode (the observed OD matrix of a mode assigned to its own network) are used as a proxy of the actual transport volumes on each link. These reference flows are then compared to those obtained by the multimodal assignment. Table 10 gives the resulting correlation coefficients between the two sets of volumes assigned to the same network segments. At this level also, the heuristic appears to produce very stable results, as there is no difference between the best and worst cases.

Table 11. Elasticities computed using the estimated parameters.

	Cost elasticities					
	Computed using best-case λ's			Computed using worst-case λ's		
	Road	IWW	Rail	Road	IWW	Rail
Road	*−0.10 to −2.65*	*0.03 to 0.21*	*0.03 to 1.71*	*−0.10 to −2.65*	*0.03 to 0.21*	*0.03 to 1.71*
	−0.32	0.07	0.20	−0.27	0.08	0.17
IWW	*0.57 to 2.09*	*−0.42 to −2.31*	*0.03 to 1.47*	*0.58 to 1.54*	*−0.42 to −3.61*	*0.03 to 1.50*
	1.15	−0.93	0.24	1.02	−1.08	0.22
Rail	*0.65 to 3.25*	*0.03 to 0.19*	*−0.65 to −2.65*	*0.49 to 3.25*	*0.03 to 0.22*	*−0.73 to −2.65*
	1.23	0.10	−1.29	1.02	0.11	−1.15

As the best case λ's obtained by the heuristic always correspond to the exact solution identified by the brute force algorithm (Table 5) and that the correlation coefficients presented in Tables 9 and 10 are (almost) identical for the best and worst cases, one can conclude that the heuristic is effective and produces stable results.

The estimators obtained with the identified λ's can further be used to compute elasticities. The values of the own and cross elasticities can indeed be derived directly from the estimated conditional logit with Box–Cox transformed independent variables (Jourquin and Beuthe 2019). Table 11 gives the computed own and cross elasticities (extreme values are printed in italic and average values are underneath) as they can be compared with the many estimates that are found in the literature. It goes beyond the scope of this paper to present an in-depth discussion on the obtained elasticities, but there is no doubt that they are comparable to what is published elsewhere (see for instance the review proposed by Beuthe, Jourquin, and Urbain 2014). The interesting point here is that the elasticities obtained using the best-case λ's are only slightly different from those computed with the worst-case λ's, which is another indicator of the robustness of the heuristic. Elasticities with respect to time and distance also could be computed, but they are of less immediate interest for assessing the merits of the proposed heuristics.

6. Conclusion

Modal choice models used in the context of strategic freight transportation studies covering large inter-regional or international areas are generally difficult to set up because of lack of explanatory data. Transportation costs, transit times and trip lengths are often among the only figures that can be gathered. These variables are rather tightly correlated so that their estimated coefficients may turn out nonsignificant or even have unexpected signs. This may even be the case when the model includes only the cost and distance variables, even though these variables are known as being very important in mode choice decision and should have a negative effect on the modes' attractiveness. Indeed, when aggregated data is used, such as annual transported volumes between European regions, the signs of the estimators for these explanatory variables are expected to be negative.

Box–Cox transforms are, among other techniques, a useful tool to improve the robustness of Logit model when they are applied to the independent variables of the utility function. They allow subtle adjustments of the functional relationship between the dependent and independent variables, which is an elegant and efficient way to improve the Log-Likelihood of the model, as well as re-establish the expected signs and their significance. When several variables have to be transformed, it is, however, difficult to identify the optimal value of the transformation parameter λ to apply to each variable because of the combinational nature of the problem. This is particularly the problem met in the present research when a third variable, distance, is introduced for adding up more information in the analysis.

Several techniques can be found in the literature to (more or less quickly) identify the optimal Box–Cox transformation parameters, but none is able to take into account the expected signs of the estimators and their level of significance.

A specific shotgun hill climbing heuristic with backtracking capabilities that considerably reduces the computing effort needed to identify a (nearby) optimal combination of λ's is described. It is applied to a series of aggregated origin-destination matrixes made available by the ETISPlus European Research Project, which cover the European territory. Additive utility functions are used in a multinomial Logit model in order to compute the market share of each mode for the OD cells. Appropriate Box–Cox transforms are applied to the explanatory variables in order to maximise the Log-Likelihood of the model, with the double constraint of expected signs and level of significance of the estimated parameters.

The risk, however, exists that forcing the model to produce expected and significant estimators drives the model away from solid statistical foundations. Using those estimators in the framework of transport policy studies could therefore lead to unrealistic forecasts. This is a critical point that is discussed all along this paper. In summary, all the results presented in this paper are carefully examined from this angle, and the conclusions are reassuring.

The performance of the heuristic is discussed and the identified Box–Cox transforms (and their correspondent estimated parameters) are compared against their optimal values. It comes out that, for the tested datasets, the heuristic is 25 times faster than the brute force approach when it is applied to bivariate utility functions and 100 times faster with three variables. Nevertheless, the algorithm often converges towards the optimal solution and, even in the worst cases, the results of the modal-choice model are nearby equal to those obtained using the best parameters. The stability of the heuristic is also confirmed by the fact that own and cross elasticities computed using the best (optimal) and worst-case estimators proposed by the heuristic are very similar.

Altogether, the algorithm appears to be efficient and effective and produces stable and statistically solid solutions.

A special attention is devoted to the design of the algorithm, for which an R code is provided, so that it can be implemented and used by practitioners using open-source software. Therefore, the complete dataset and all the R scripts developed for this paper are available at https://github.com/jourquin/Box–Cox-Lambdas-Heuristic.

In this paper, the heuristic is applied to a conditional Logit, i.e. for which only one λ must be estimated for each independent variable. It would be further interesting to test it with an additional BC transformation of the dependent variable or to use it on a dataset with shipper dependent instead of mode specific explanatory variables. In the latest a case, a specific λ must be computed for each mode and each variable: if three modes are available, 6 λ's would be needed for a bivariate case, and 9 λ's for a trivariate case.

Notes

1. See Appendix 1 for a description of the content of each category.
2. The 'spatial distribution' of the valid solutions (spread and position of the dots) is very different for each group of commodities. Unfortunately, there is no place to publish the 10 figures in this paper, but they are available on request.
3. Note that for values of lambda(s) $\geq |2|$ numerical errors in the computation of the (approximated) Hessian needed during the estimation of the logit sometimes occur.
4. This can be implemented as a recursive function (called from itself several times, once for each 'dimension' of the combination of λ's). See Appendix 2.
5. While the heuristic can be applied to the univariate case, its usefulness is limited.
6. One matrix per NST/R, each one containing the aggregated demand for the 3 modes.

Acknowledgment

The author would like to thank Michel Beuthe for the many constructive discussions during the development phase of this research. He also thanks the three anonymous reviewers who carefully read the submitted manuscript and suggested several improvements.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Bart Jourquin http://orcid.org/0000-0003-3446-1027

References

Adeboye, N. O., I. S. Fagoyinbo, and T. O. Olatayo. 2014. "Estimation of the Effect of Multicollinearity on the Standard Error for Regression Coefficients." *IOSR Journal of Mathematics* 10 (4): 16–20. doi:10.9790/5728-10411620.

Arencibia, Ana Isabel, María Feo-Valero, Leandro García-Menéndez, and Concepción Román. 2015. "Modelling Mode Choice for Freight Transport Using Advanced Choice Experiments." *Transportation Research Part A: Policy and Practice* 75 (May): 252–267. doi:10.1016/j.tra.2015.03.027.

Beuthe, M., and Ch. Bouffioux. 2008. "Analysing Qualitative Attributes of Freight Transport from Stated Orders of Preference Experiment." *Journal of Transport Economics and Policy* 42 (1): 105–128.

Beuthe, M., B. Jourquin, and N. Urbain. 2014. "Estimating Freight Transport Price Elasticity in Multi-Mode Studies: A Review and Additional Results from a Multimodal Network Model." *Transport Reviews*. doi:10.1080/01441647.2014.946459.

Beuthe, M., B. Jourquin, N. Urbain, I. Lingemann, and B. Ubbels. 2014. "Climate Change Impacts on Transport on the Rhine and Danube: A Multimodal Approach." *Transportation Research Part D: Transport and Environment* 27: 6–11. doi:10.1016/j.trd.2013.11.002.

Bierlaire, Michel. 2003. "BIOGEME: A Free Package for the Estimation of Discrete Choice Models." In *Swiss Transport Research Conference*.

Box, G. E. P., and D. R. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society: Series B (Methodological)* 26 (2): 211–243. doi:10.1111/j.2517-6161.1964.tb00553.x.

Christian, Brian, and Tom Griffiths. 2016. *Algorithms to Live by: The Computer Science of Human Decisions*. First U.S. Edition. New York: Henry Holt and Company.

Croissant, Yves. 2019. *Estimation of Multinomial Logit Models in R: The Mlogit Packages* (version 1.0-2). R. https://cran.r-project.org/web/packages/mlogit/index.html.

Cullinane, Kevin, and Neal Toy. 2000. "Identifying Influential Attributes in Freight Route/Mode Choice Decisions: A Content Analysis." *Transportation Research Part E: Logistics and Transportation Review* 36 (1): 41–53. doi:10.1016/S1366-5545(99)00016-2.

de Jong, Gerard, Hugh Gunn, and Moshe Ben-Akiva. 2004. "A Meta-Model for Passenger and Freight Transport in Europe." *Transport Policy* 11 (4): 329–344. doi:10.1016/j.tranpol.2004.03.001.

Fridstrøm, L., and A. Madslien. 1994. "Own Account or Hire Freight: A Stated Preference Analysis." In *7th International Conference on Travel Behaviour*. Valle Nevado, Chile. https://www.researchgate.net/publication/268221233_Own_account_or_hire_freight_a_stated_preference_analysis.

Gaudry, Marc. 2016. "Méthodes Box-Cox, Algorithmes de TRIO et Demande de Transport : Trois Consignes Occamiennes Pour Faire Rendre Sens Aux Coefficients de Régression de Modèles Simples et Logistiques Discrets Ou Agrégés." doi:10.13140/RG.2.1.1102.5044.

Gaudry, Marc, and Matthieu de Lapparent. 2015. "Attitudes to Distance, Time and Cost in Logit Transport Choice Models." *Report TRANSP-OR 150113, Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne*. https://transp-or.epfl.ch/documents/technicalReports/gaudry_lapparent_150113.pdf.

Greene, William H. 2012. *Econometric Analysis*. 7th ed., International edition. Pearson: Pearson Series in Economics.

Hasan, Asad, Sentrana Inc, Wang Zhiyu, Alireza Mahani, and Sentrana Inc. 2019. *Fast Estimation of Multinomial Logit Models: R Package Mnlogit* (version 1.2.6). R. https://cran.r-project.org/web/packages/mnlogit/index.html.

Hong, X. 2006. "A Fast Identification Algorithm for Box–Cox Transformation Based Radial Basis Function Neural Network." *IEEE Transactions on Neural Networks* 17 (4): 1064–1069. doi:10.1109/TNN.2006.875986.

Ishak, Nur Aufa Mazni, and Sanizah Ahmad. 2018. "Estimating Optimal Parameter of Box-Cox Transformation in Multiple Regression with Non-Normal Data." In *Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)*, edited by Nor Azizah Yacob, Nur Asmaliza Mohd Noor, Nor Yuziah Mohd Yunus, Rahmah Lob Yussof, and Shaikh Abdul Karim Yamani Zakaria, 1039–1046. Singapore: Springer Singapore. doi:10.1007/978-981-13-0074-5_102.

Jourquin, Bart. 2016. "Calibration and Validation of Strategic Freight Transportation Planning Models with Limited Information." *Journal of Transportation Technologies* 06 (05): 239–256. doi:10.4236/jtts.2016.65023.

Jourquin, Bart, and Michel Beuthe. 2019. "Cost, Transit Time and Speed Elasticity Calculations for the European Continental Freight Transport." *Transport Policy* 83 (November): 1–12. doi:10.1016/j.tranpol.2019.08.009.

Liedtke, Gernot, and David Guillermo Carrillo Murillo. 2012. "Assessment of Policy Strategies to Develop Intermodal Services: The Case of Inland Terminals in Germany." *Transport Policy* 24 (November): 168–178. doi:10.1016/j.tranpol.2012.06.002.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behaviour." In *Frontiers in Econometrics*, edited by P. Zarembka, 105–142. New York, NY: Academic Press New York.

Parady, Giancarlos, David Ory, and Joan Walker. 2021. "The Overreliance on Statistical Goodness-of-Fit and under-Reliance on Model Validation in Discrete Choice Models: A Review of Validation Practices in the Transportation Academic Literature." *Journal of Choice Modelling* 38 (March): 100257. doi:10.1016/j.jocm.2020.100257.

Rich, J., P. M. Holmblad, and C. O. Hansen. 2009. "A Weighted Logit Freight Mode-Choice Model." *Transportation Research Part E: Logistics and Transportation Review* 45 (6): 1006–1019. doi:10.1016/j.tre.2009.02.001.

Robert, Christian P., and George Casella. 2004. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. New York, NY: Springer New York. doi:10.1007/978-1-4757-4145-2.

Rothengatter, Werner. 2019. "Megaprojects in Transportation Networks." *Transport Policy* 75 (March): A1–15. doi:10.1016/j.tranpol.2018.08.002.

Russell, Stuart J., and Peter Norvig. 2003. *Artificial Intelligence: A Modern Approach*. 2nd ed. Prentice Hall Series in Artificial Intelligence. Upper Saddle River, NJ: Prentice Hall/Pearson Education.

Sakia, R. M. 1992. "The Box-Cox Transformation Technique: A Review." *The Statistician* 41 (2): 169. doi:10.2307/2348250.

Soleymani, Samira. 2018. "Exact Box-Cox Analysis." *Electronic Thesis and Dissertation Repository*, April. https://ir.lib.uwo.ca/etd/5308.

Szimba, Eckhard, Markus Kraft, Jan Ihrig, Antje Schimke, Oliver Schnell, Yuko Kawabata, Sean Newton, et al. 2012. "ETISplus Database Content and Methodology." doi:10.13140/RG.2.2.16768.25605.

Witten, I. H., Eibe Frank, and Mark A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd ed. Morgan Kaufmann Series in Data Management Systems. Burlington, MA: Morgan Kaufmann.

Zhang, Mo. 2013. *A Freight Transport Model for Integrated Network, Service and Policy Design*. Delft: TRAIL Research School.

Appendices

Appendix 1. NST/R main chapters

#	Description
0	Agricultural products and live animals
1	Foodstuffs and animal fodder
2	Solid mineral fuels
3	Petroleum products
4	Ores and metal waste
5	Metal products
6	Crude and manufactured minerals, building materials
7	Fertilizers
8	Chemicals
9	Machinery, transport equipment, manufactured articles and miscellaneous articles

Appendix 2. R code of the recursive 'exploreAround' function

Global variables are underlined:

- range: absolute value of the limits to search the λ's into. Example: 2 for $[-2, +2]$
- <u>bestSolution</u>: best solution (combination of λ's) found so far
- solutionsToExplore: dynamic list of solutions that need to be explored around

Other functions are in italic:

- *getLambdas(…)*: returns a vector with the λ's of a given solution
- retrieveOrComputeLogit*(…)*: retrieves the results of the Logit model for the given combination of λ's. Compute the Logit if not yet done
- *isValid(…)*: returns true if all the estimators of the independent variables of the given solution have the expected sign and if the estimators have the minimal level of significance desired by the modeller
- hasExpectedSigns*(…)*: returns true if all the estimators of the independent variables of the given solution have the expected sign

When called, only the first two parameters must be passed to the function, the last one being used during recursiveness.

```
exploreAround <- function(solution, stepSize, dimLevel = 1) {

  # Test a step backward in the current dimension, but remain in range
  p = getLambdas(solution)
  if (p[dimLevel] - stepSize >= -range) {
    p[dimLevel] = round(p[dimLevel] - stepSize, 1)
  } else {
    p[dimLevel] = -range
  }
  newSolution = retrieveOrComputeLogit(p)

  # Is this the best solution found so far?
  if (isValid(newSolution) & newSolution$LL > bestSolution$LL) {
    bestSolution <<- newSolution
  }

  # Add this solution to the list to explore later if it has better LL and
    expected signs
  if (hasExpectedSigns(newSolution) & newSolution$LL > solution$LL) {
    solutionsToExplore[[length(solutionsToExplore)+1]] <<- newSolution
  }

  # Test a step forward in the current dimension, but remain in range
  p = getLambdas(solution)
  if (p[dimLevel] + stepSize <= range) {
    p[dimLevel] = round(p[dimLevel] + stepSize, 1)
  } else {
    p[dimLevel] = range
  }
  newSolution = retrieveOrComputeLogit(p)

  # Is this the best solution found so far?
  if (isValid(newSolution) & newSolution$LL > bestSolution$LL) {
    bestSolution <<- newSolution
  }

  # Add this solution to the list to explore later if it has better LL and
    expected signs
  if (hasExpectedSigns(newSolution) & newSolution$LL > solution$LL) {
    solutionsToExplore[[length(solutionsToExplore)+1]] <<- newSolution
  }

  # Recursive entrance to next dimension
  if (dimLevel < length(p)) {
    exploreAround(solution, stepSize, dimLevel + 1)
  }
}
```