

Research Article

Bayesian Disaggregate and Aggregate Calibration of Path Logit Choice Models

Giulio Erberto Cantarella¹ and Antonino Vitetta² 

¹Dipartimento di Ingegneria Civile, Università degli Studi di Salerno, Fisciano, SA 84084, Italy

²Dipartimento di ingegneria dell'Informazione, delle Infrastrutture e dell'Energia Sostenibile, Università degli Studi Mediterranea di Reggio Calabria, Feo di Vito, Reggio Calabria 89122, Italy

Correspondence should be addressed to Antonino Vitetta; vitetta@unirc.it

Received 18 March 2023; Revised 21 June 2023; Accepted 28 June 2023; Published 15 July 2023

Academic Editor: Yajie Zou

Copyright © 2023 Giulio Erberto Cantarella and Antonino Vitetta. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In transport demand analysis, the calibration of a model means estimation of its (endogenous) parameters from observed data with an inference statistical estimator. Indeed, these considerations apply to any choice behaviour model, such as those derived from Random Utility Theory or any other choice modelling theory. Calibration of choice models can be carried out from disaggregate vs. aggregate data, while inference statistical estimators can be specified through Bayesian vs. Classic (or Frequentist) approaches. In this paper, the resulting Bayesian or Classic disaggregate or aggregate calibration methods are discussed, analysed in detail, and compared from the mathematical point of view. These methods are applied to calibrate Logit choice models for describing path choice behaviour at national scale on a small sample. The Logit choice model can be derived from Random Utility Theory (or be considered an instance of the Bradley–Terry model). Path choice set definition is also discussed, and specialised indicators are used for result comparison. The main contributions of this study concern the use of two different estimation approaches, Bayesian vs. Classic, adopting and introducing some indicators of goodness of estimation. The results of this work, relating to the sample of users adopted, show that the Bayesian approach provides a better estimate than the Classic approach because the calibrated parameters are more stable, the specific constants of the alternatives decrease, and the resulting models show better values of clearly right indicator.

1. Introduction

User path choice behaviour models are one of two main elements of any method for travel demand assignment to a transportation network, the other being the arc cost-flow functions modelling user driving behaviour and resulting congestion, a part of the OD demand flow matrix (details can be found in [1] and [2]).

Traditionally, within assignment methods, the path choice behaviour is modelled

- (i) By assuming the path choice set defined by
 - (a) Either exhaustive enumeration of all elementary paths

- (b) Or selective enumeration of some of them according to some given criteria (i.e., efficient paths as defined in [3])

- (ii) By defining the user choice strategy among alternatives in the path choice set.

From a more general point of view, Mansky [4] proposed to model any choice behaviour as a two-level decision:

- (i) By an explicit definition of the perceived choice set
- (ii) By a description of choice among alternatives in the choice set.

This approach includes the first one as a special case and it is described below in detail for path choice behaviour

considering that the application described in this paper focuses on the explicit approach.

(i) *Alternative Choice Set*

- (a) *Generation of Perceived Alternatives.* Given a pair of origin and destination, the routing alternatives can be generated through the exhaustive approach (considering all elementary, say loop-less, paths) or the selective approach (only some elementary paths). The exhaustive approach may be less effective because commonly users likely perceive only a few alternatives. For this reason, first some candidate paths are generated considering some reasonable criteria [5, 6].
- (b) *Definition of the Perceived Choice Set.* Then, from the candidate alternatives, the perceived choice set may be defined through the application of deterministic [5] or probabilistic perception models [7, 8].
- (ii) *Choice Strategy among Alternatives.* The choice of the alternative from the perceived choice set is commonly described through Random Utility Models (RUM; see [5, 9, 10]), Fuzzy Utility Models (FUM; [11–13]; see also [14, 15]) Quantum Utility Models (QUM; [16, 17]), psychological choice modelling, and ranking models (for instance, [18–20]) as well as applying any other choice modelling theory, for instance, comparing pairs of alternatives [21, 22] for the probability estimation.

According to most used utility-based choice theories, given the choice set defined as in point I, each user moving between o and d pair i

- (a) Knows all paths in the *choice set*
- (b) Associates to each path in the choice set a *perceived utility*
- (c) Chooses the maximum perceived utility (or minimum perceived disutility) path
- Furthermore, according to any Uncertainty Utility Theory,
- (d) The perceived utility is modelled by a continuous uncertainty number due to several sources of uncertainty regarding the users as well as the modellers.

In RUT, the perceived utility is modelled by a continuous random variable, with expected value called *systematic utility*; thus, the choice probability of an alternative is given by the probability that its perceived utility is equal to maximum among all alternatives; hence, the path choice proportions are assumed defined by the path choice probabilities. When the perceived utility covariance matrix is non-singular, a *probabilistic path choice function* is obtained, fully specified by the *perceived utility probability density function (pdf)*.

Assuming the perceived utilities independent and distributed with some common parameters according to a Gumbel or a Weibull r.v., the Logit or the Weibit model

[23] is obtained, respectively, both in closed form. Independence between alternatives does not allow to model overlapping paths, and this issue can be addressed within the utility specification (i.e., C-Logit in [24]; Path Size in [25]). Other approaches, not in closed form, model overlapping paths through covariance: Probit [26] or Gammit [27] based on MVNormal or MVGamma distributions. More details are given in de Luca [28]. As already stressed, the Logit or the Weibit choice models can also be derived from the psychological Bradley and Terry [21] model (BLT).

Whichever is the choice modelling theory, *parameters of the choice model*, including those in the utility specification, have to be calibrated against observed data.

From the statistical point of view, estimators can be specified according to

(i) Bayesian (B) statistical inference:

Parameters are assumed described by random variables and (point/interval) estimates are defined through the posterior distribution, given by the prior distribution times the likelihood of the observations (duly normalised); the resulting estimators can easily be fitted in a dynamic context applying the well-known Bayes' theorem.

(ii) Classic or Frequentist (C) statistical inference:

Parameters are assumed having a (deterministic) value (even though non-knowledgeable), and their estimates are given by the point of maximum of the likelihood, say the joint probability of observing the sample of input data, where the maximum is taken among all possible samples (or other methods such as Least Squares).

The Bayesian statistical estimator is a random variable, and the uncertainty is defined by a probability distribution. The Classic statistical estimator is a point value based on the Frequentist approach. Bayesian methods assume prior pdf (probability density function) specification and obtain posterior pdf specification, possibly different in specification and parameters; in Classic estimation, a prior belief is not considered.

Bayesian methods have recently been adopted for travel demand estimation and traffic models in dynamic context. The dynamic travel demand is estimated in Yu et al. [29] considering normal prior distribution and posterior distribution obtained from added observed traffic counts; the errors decrease with the observations. Considering the generalized Bayesian approach and path choice, Zhu et al. [30, 31] analysed the traffic models in stochastic transportation systems with user equilibrium and non-user equilibrium conditions and the convergences with numerical studies and day-to-day dynamics for path.

In transportation demand analysis, the estimation of endogenous parameters of the demand model from observed data is commonly called calibration, developed through the steps of specification of the functional form, parameter estimation, and validation with formal and informal statistic indicators. For this reason, the term calibration is used below to define the estimate of the

endogenous travel demand model parameters, with specific reference to the application to path choice. Calibration can be carried out from:

- (i) Disaggregate data: the input data are the alternative effectively chosen by each user in a sample, assuming that their choice decisions are independent.
- (ii) Aggregate data: the input data are the frequencies or the number of time that each alternative is observed.

The parameter calibration for path choice has been carried out through Classic estimators [5, 24] or Bayesian ones [32, 33] with multinomial prior Logit or mixed Logit probability. Washington et al. [32] reported theoretical aspects about calibration of multinomial Logit models, applied to path choice, through Classic or Bayesian estimators.

In this paper, the Bayesian calibration approach is analysed in detail and compared with the Classic one by calibrating Logit path choice models at national scale. Path choice set definition is also discussed. The paper is structured as follows. Section 2 presents the methodology. Section 3 discusses the results of an application to a real case. Conclusions and further developments are given in Section 4. The main contributions of this paper concern the use of

- (i) Two different calibration approaches, Bayesian vs. Classic (or Frequentist) (Section 2).
- (ii) Two different sources of information, observed choices or frequencies (Section 2).
- (iii) Some indicators of goodness of calibration (Section 3).

2. Methodology

Path choice behaviour for each user class (u.c.) n can be described by applying any model derived from a discrete choice theory (such as RUM, FUM, and QUM quoted above). Applying Random Utility Theory (RUT), it is assumed that for a journey:

- (i) S_n is the choice set containing the perceived alternatives for the u.c. n .
- (ii) Each user in the u.c. n associates a *perceived utility* $U_{k,n}$ to each path k belonging to the choice set S_n of perceived alternatives.
- (iii) Each user in the u.c. n chooses the (an) alternative of maximum perceived utility.
- (iv) The perceived utility $U_{k,n}$ is modelled as a (continuous) random variable, with pdf $f_U(\bullet)$, due to several sources of uncertainty both for users and modeller; it has expected value $E(U_{k,n}) = v_{k,n}$, called *systematic utility* (several specifications can be considered for the pdf of $U_{k,n}$).

According the above assumptions, the choice proportion $p_{k,n}$ of path k is given by the (mass) probability that the perceived utility of path k is the maximum (that is the choice probability):

$$\begin{aligned} p_{k,n} &= p_n(k|\boldsymbol{\theta}_n, S_n) \\ &= P(U_{k,n} \geq U_{h,n}, \forall k, h \in S_n), \end{aligned} \quad (1)$$

where $\boldsymbol{\theta}_n$ is the choice function parameter vector including the perceived utility pdf parameters, whose meaning depends on the choice model specification; this vector may be different with the u.c. n . According to equation (1), the choice probability $p_n(k|\boldsymbol{\theta}_n, S_n)$ depends on the values of the systematic utility $v_h \forall h \in S_n$.

The systematic utility is commonly assumed, both in research and practice, a linear combination of a vector \mathbf{x} of attributes, such as travel time, monetary costs due to tolls, fees, . . . , through utility parameters $\boldsymbol{\theta}_v$, included in vector $\boldsymbol{\theta}$:

$$v_{k,n} = \boldsymbol{\theta}_v' \cdot \mathbf{y}_{k,n}. \quad (2)$$

All parameters in vector $\boldsymbol{\theta}$ are to be calibrated against observations, and they are usually considered deterministic parameters. Recently, they have also been considered random variables Θ with a joint pdf $f(\boldsymbol{\theta}|S_n)$, to model for instance users' heterogeneity, VoT distribution, etc.

For each u.c. n different RUMs are obtained depending on the distribution of the perceived utility, as already said. In this paper, multinomial Logit is used, where the perceived utilities $U_{k,n}$ are independently and identically distributed (i.i.d.) as a Gumbel r.v. with mean given by the systematic utility $v_{k,n}$ and dispersion parameter θ_o ; this parameter plays the role of utility scale factor, and thus it may not be distinguished from $\boldsymbol{\theta}_v$; therefore, the only actual parameters of the model are $\boldsymbol{\theta} = \boldsymbol{\theta}_v/\theta_o$. Equation (2) becomes

$$v_{k,n} = \boldsymbol{\theta}' \cdot \mathbf{y}_{k,n}, \quad (3)$$

and if all parameters in vector $\boldsymbol{\theta}$ are assumed deterministic variables, the choice probabilities can be defined in closed form:

$$p_n(k|\boldsymbol{\theta}, S_n) = \frac{\exp(v_{k,n}/\theta_o)}{\sum_{h \in S_n} \exp(v_{h,n}/\theta_o)}, \quad (4)$$

and as already noted, the Logit choice model can also be obtained as a psychological one.

As already said, calibration of Random Utility Models for path choice (or of any other model), that is, estimation of parameters $\boldsymbol{\theta}$, can be carried out from:

- (i) Disaggregate data: the input data are the alternative effectively chosen by each user within a sample of N users, assuming that their choices are stochastically independent.
- (ii) Aggregate data: the input data are the frequencies, that is, the number of times that each alternative is chosen, with N being the sum of frequencies over all alternatives.

Similar considerations may apply for other RUMs such as Probit, Weibit, Gammit, or other psychological choice models that will be discussed in a future paper.

In the following notation, S_n is omitted to simplify notation. For application's purpose, each user class n is assumed to be made up of one user.

2.1. The Classic (or Frequentist) Approach. According to the Classic (or Frequentist) approach of estimation, each parameter is assumed having a given (even though non-knowledgeable) value, and its estimate is given by the point of the maximum of the likelihood, say the joint probability of observing the sample of input data over all possible samples under the same conditions. This approach shows several drawbacks, such as the assumption of the availability of infinite many samples under the same conditions, as well known in the literature.

The joint distribution of the sample of the observations given the values of parameters θ can be specified by the disaggregate or the aggregate approaches described above.

2.1.1. Disaggregate. Given a sample of N users, choice decisions are assumed independent, for a user n , the alternative x_n effectively chosen is observed, and values x_n are collected in vector $\mathbf{x} = [\dots, x_n, \dots]^T$.

The likelihood for the N observed chosen alternatives is the joint probability of observing the alternative effectively chosen by each user.

Assuming that the random variables X_n describing the alternative chosen by each user n are i.i.d. as the integer r.v. X with probability mass function p_x , is specified as in the previous subsection:

$$\begin{aligned} P(\mathbf{x}|\theta) &= p_{\mathbf{x}}(\mathbf{x}|\theta) \\ &= \prod_n p_{\mathbf{x}}(x_n|\theta). \end{aligned} \quad (5)$$

2.1.2. Aggregate. Given M alternatives, the number of times (frequency), z_m , that each alternative m is observed is assumed independent, and values $z_m \geq 0$ are collected in the vector $\mathbf{z} = [\dots, z_m, \dots]^T$, to avoid any confusion with vector \mathbf{x} , with $\sum_m z_m = N$.

Assuming that \mathbf{z} is a realization of a multinomial random variable \mathbf{Z} with probabilities $p_z(m)$, specified above, the likelihood of observing the vector \mathbf{z} is given by

$$\begin{aligned} P(\mathbf{z}|\theta) &= p_{\mathbf{z}}(\mathbf{z}|\theta) \\ &= \prod_m p_{\mathbf{z}}(z_m|\theta) \\ &= \left(\frac{(\sum_m z_m)!}{(\prod_m z_m!)} \right) \cdot \prod_m p_{\mathbf{z}}(m|\theta)^{z_m}. \end{aligned} \quad (6)$$

2.1.3. Estimation method. For estimating parameters θ , the likelihood L is considered as a function of θ given by $L(\theta; \mathbf{x}) = P(\mathbf{x}|\theta)$. Thus, according to the maximum likelihood method, the estimates θ_{ML} are given by the solution of the maximisation problem: $\text{Max}_{\theta} L(\theta; \mathbf{x})$.

Very often a logarithmic transformation simplifies the problem without changing the solution: $l(\theta; \mathbf{x}) = \ln(L(\theta; \mathbf{x}))$; thus, the estimates θ_{ML} are given by the solution of $\text{Max}_{\theta} l(\theta; \mathbf{x})$. In Classic statistical inference, θ_{ML} maximises the probability to observe the really observed sample with respect to all the others.

(i) Disaggregate:

According to the above assumptions, the ML estimates θ_{ML} are given by

$$\theta_{ML} = \text{Arg Max}_{\theta} \sum_n \ln(p_n(x_n|\theta)). \quad (7)$$

(ii) Aggregate:

According to the above assumptions, the ML estimates θ_{ML} are given by

$$\theta_{ML} = \text{Arg Max}_{\theta} \sum_m z_m \ln(p_m(k|\theta)). \quad (8)$$

Following the Least Squares method, the estimation θ_{LS} is given by

$$\theta_{LS} = \text{Arg Min}_{\theta} \sum_n \left(\left(\frac{x_n}{\sum_n x_n} \right) - p_n(k|\theta) \right)^2. \quad (9)$$

2.2. The Bayesian Approach. According to Bayesian estimation, parameters are assumed described by random variables and estimates are given by the median or the mean or the mode (if the posterior pdf is unimodal) of their posterior joint pdf given by their prior joint pdf times the likelihood of the observation (duly normalised). This approach shows several useful features, as well known in the literature, such as it is suitable for successive re-estimations when further information is collected, possibly in a dynamic context, and it allows consistent definition of credible interval estimators.

The Bayesian approach considers that the *posterior* $P(\theta|\mathbf{x})$ conditional joint pdf of the parameters θ , conditioned to the observation vector \mathbf{x} , is equal to the product between the *prior* ($P(\theta)$) and the *likelihood* ($P(\mathbf{x}|\theta)$) probabilities, duly normalised by factor $(1/P(\mathbf{x}))$:

$$P(\theta|\mathbf{x}) = \left(\frac{1}{P(\mathbf{x})} \right) \cdot P(\mathbf{x}|\theta) \cdot P(\theta). \quad (10)$$

The initial knowledge about the parameter θ is described by the prior ($P(\theta)$), and then multiplying it by the likelihood $P(\mathbf{x}|\theta)$ from a set of observations \mathbf{x} (and normalising it) gives an update of the knowledge as described by the posterior function $P(\theta|\mathbf{x})$. It should be noted that the normalising factor $(1/P(\mathbf{x}))$ does not depend on parameters θ .

2.2.1. Prior Specification. For the vector θ of parameters, a prior conditioned joint pdf $f(\theta)$ —of the parameters θ —can be assumed known. It derives from the past information available about θ . It could derive from previous

estimation procedures or previous experiences about it (i.e., sign, probable interval, and probabilistic distribution).

Assuming θ as continuous random variables, the prior probability $P(\theta)$ is given by the prior conditioned joint pdf of θ , $f(\theta)$:

$$P(\theta) = f(\theta). \quad (11)$$

Assuming that all the parameters θ_j are i.i.d. with pdf $f(\theta_j)$ the a prior probability has the specification:

$$P(\theta) = \prod_j f(\theta_j). \quad (12)$$

2.2.2. Likelihood Specification. Our knowledge of parameters θ as modelled by the prior pdf can be improved from observations. The likelihood L —of the observations over the parameters—can be specified in two ways:

- (a) Observing the alternative chosen by a sample of users, \mathbf{x} .
- (b) Observing the frequencies relative to the alternative chosen by a sample of users, \mathbf{z} .

As above, it is repeated for reader's convenience.

(i) Disaggregate:

$$\begin{aligned} P(\mathbf{x}|\theta) &= p_X(\mathbf{x}|\theta) \\ &= \prod_n p_X(x_n|\theta). \end{aligned} \quad (13)$$

(ii) Aggregate:

$$\begin{aligned} P(\mathbf{z}|\theta) &= p_z(\mathbf{z}|\theta) \\ &= \prod_m p_z(\mathbf{z}_m|\theta) \\ &= \left(\frac{(\sum_m z_m)!}{(\prod_m z_m!)} \right) \cdot \prod_m p_z(m|\theta)^{z_m}. \end{aligned} \quad (14)$$

Note that the probability mass function $p_n(k|\theta)$ can be specified in several ways, and some of them are reported in Section 2. For the probability evaluation, the meaning of the parameter θ depends on the type of model.

The likelihood can be considered as a function of θ :

$$\begin{aligned} L(\theta|\mathbf{x}) &= P(\mathbf{x}|\theta), \\ &\text{or} \\ L(\theta|\mathbf{z}) &= P(\mathbf{z}|\theta). \end{aligned} \quad (15)$$

2.2.3. Normalising Factor. As said above, the normalising factor ($1/P(\mathbf{x})$) depends on observations but does not depend on parameters θ .

2.2.4. Posterior Definition. As reported above, the posterior $P(\theta|\mathbf{x})$ conditional joint pdf of the parameters θ is the probability given to parameters θ after observations have been collected.

(i) Disaggregate:

$$P(\theta|\mathbf{x}) = \left(\frac{1}{P(\mathbf{x})} \right) \cdot P(\mathbf{x}|\theta) \cdot P(\theta). \quad (16)$$

(ii) Aggregate:

$$P(\theta|\mathbf{z}) = \left(\frac{1}{P(\mathbf{z})} \right) \cdot P(\mathbf{z}|\theta) \cdot P(\theta). \quad (17)$$

2.2.5. Point and Interval Estimators. In both cases, a point estimate of parameters θ may be given by the mean, the median, the mode, or any other location indicator of the posterior distribution, depending on the error function. Credible intervals are easily defined from the posterior joint pdf, as well.

(i) Disaggregate:

$$P(\theta|\mathbf{x}) = \left(\frac{1}{P(\mathbf{x})} \right) \cdot L(\theta|\mathbf{x}) \cdot P(\theta). \quad (18)$$

(ii) Aggregate:

$$P(\theta|\mathbf{z}) = \left(\frac{1}{P(\mathbf{z})} \right) \cdot L(\theta|\mathbf{z}) \cdot P(\theta). \quad (19)$$

(1) *Normalising Factor.* It can be easily noted that the normalising factor does not affect the point of maximum, and thus it can be neglected.

(2) *Prior pdf.* Several approaches are available to define the prior distribution such as conjugate priors and improper priors. As the size of the sample goes to infinity, the results of the Bernstein–von Mises theorem can be used:

- ① The posterior tends to the likelihood, and the prior role becomes less and less relevant:

$$\begin{aligned} \lim_{N \rightarrow \infty} P(\theta|\mathbf{x}) &= L(\theta|\mathbf{x}), \\ &\text{or} \end{aligned} \quad (20a)$$

$$\lim_{N \rightarrow \infty} P(\theta|\mathbf{z}) = L(\theta|\mathbf{z}).$$

- ② The likelihood tends to be normally distributed:

$$\begin{aligned} \lim_{N \rightarrow \infty} L(\boldsymbol{\theta}|\mathbf{x}) &\sim \text{Normal}(\boldsymbol{\theta}|\mathbf{x}), \\ &\text{or} \\ \lim_{N \rightarrow \infty} L(\boldsymbol{\theta}|\mathbf{z}) &\sim \text{Normal}(\boldsymbol{\theta}|\mathbf{z}). \end{aligned} \quad (20b)$$

Thus, for large samples ① the mode $\boldsymbol{\theta}_{\text{BM}}$ of the posterior pdf goes to the point of the maximum of the Likelihood, ② the mean and the median goes to mode, therefore:

$$\begin{aligned} \boldsymbol{\theta}_{\text{BM}} &= \text{Arg Max}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\mathbf{x}), \\ &\text{or} \\ \boldsymbol{\theta}_{\text{BM}} &= \text{Arg Max}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\mathbf{z}). \end{aligned} \quad (21)$$

Moreover, this theorem links Bayesian inference with Classic or Frequentist inference.

(3) *Remark.* In most cases, it is convenient to maximise a log-transformation of the objective function, which does not affect (as any increasing transformation) the point of maximum. Assuming the (naïve) pair-wise conditional independence between the N observations and that the m parameters are i.i.d. with pdf $f(\bullet)$ and applying equations (13) and (14) for defining the likelihood in equation (21), the estimators are given by

(i) Disaggregate:

$$\boldsymbol{\theta}_{\text{BM}} = \text{Arg Max}_{\boldsymbol{\theta}} \sum_j \ln(f(\theta_j)) + \sum_n \ln(p_x(x_n|\boldsymbol{\theta})). \quad (22)$$

(ii) Aggregate:

$$\begin{aligned} \boldsymbol{\theta}_{\text{BM}} &= \text{Arg Max}_{\boldsymbol{\theta}} \sum_j \ln(f(\theta_j)) + \ln\left(\prod_m p_z(m|\boldsymbol{\theta})^{z_m}\right) \\ &= \text{Arg Max}_{\boldsymbol{\theta}} \sum_j \ln(f(\theta_j)) + \sum_m z_m \cdot \ln p_z(m|\boldsymbol{\theta}), \end{aligned} \quad (23)$$

where $\ln((\sum_m z_m)!)/(\prod_m z_m!)$ has been omitted since it does not affect the point of maximum.

2.2.6. *Application in a Dynamic Context.* The Bayesian estimators can easily be fitted in a dynamic context applying the well-known Bayes' theorem. Indeed, once the posterior distribution has been defined, it can be used as the prior distribution when other data, possibly of different types, will be available.

3. Results and Discussion

In this section, results of an application are discussed. The whole (small) sample contains 216 randomly sampled independent interviews (one interview for each user). Data were collected through an on-road revealed preference (RP) survey, carried out in Italy at the national scale. Reference was to short/long extra-urban car trips. Different OD pairs, between main cities at least partially connected through

motorways, were considered such that the observed distances are in the range of 20–1100 km and the travel times are in the range of 0.2–10.5 h.

The travel survey's main purpose was to test some calibration methods for route choice models at national scale; therefore, the whole OD matrix was not estimated since it is not relevant for the analysis of the user route choice behaviour; of course, it would be much relevant for any kind of assignment that is not a topic of this paper.

The level of service attributes for the Italian national road transport system and the comparison between the generated and the chosen alternatives are carried out with a road network composed of all motorways and main roads. It has around 5000 nodes and 16000 links.

The main purpose of the experimentation is the comparison of different methodologies for the model parameter estimation; for this aim, a small sample has been used for computation simplicity. For practical applications, larger samples must be used.

Alternative choice set is generated with a multicriteria and deterministic approach considering the criteria that guarantee a high degree of coverage of the interviews with a significant choice of path for each criterion.

The user choice strategy among alternatives is described by a Logit choice model, as a psychological choice model or obtained as a RUM assuming perceived utility, independent Gumbel distributed with common scale parameter θ_0 . The calibration of the distribution parameter, including those of the systematic utility specification, is based on Bayesian and Classic approaches.

3.1. Main Assumptions and Preliminary Data Analysis.

The quantitative characteristics (attributes) considered for each path, which give the best statistical results in the calibration, are

- (i) $y_{Tk,n}$, travel time (in hour) with parameter θ_T .
- (ii) $y_{Pk,n}$, percentage of distance (in decimal) on motorway links with parameter θ_P .
- (iii) $y_{Lk,n}$, label variables (0 or 1) considered in the path generated with the criterion of minimum time with parameter θ_L (it is 1 if the path of minimum time is equal to the path with another adopted criterion).

The expected value of the perceived utility has the specification:

$$\frac{v_{k,n}}{\theta_0} = \theta_T \cdot y_{Tk,n} + \theta_P \cdot y_{Pk,n} + \theta_L \cdot y_{Lk,n}, \quad \forall k \in S_n. \quad (24)$$

It is assumed that each entry of the prior vector $\boldsymbol{\theta} = [\theta_T; \theta_P; \theta_L]'$ is independently distributed with a normal probability function $f_N(\boldsymbol{\theta})$, with

- (i) Expected value $\boldsymbol{\mu}_{\boldsymbol{\theta}} = [\mu_{\theta_T}; \mu_{\theta_P}; \mu_{\theta_L}]' = [-2 \text{ h}^{-1}; 2; 2]'$.
- (ii) Standard deviation $\boldsymbol{\sigma}_{\boldsymbol{\theta}} = [\sigma_{\theta_T}; \sigma_{\theta_P}; \sigma_{\theta_L}]' = [1 \text{ h}^{-1}; 1; 1]'$ (and all covariances, $\sigma_{\theta_T P}, \sigma_{\theta_T L}, \sigma_{\theta_P L}, \sigma_{\theta_L T}, \sigma_{\theta_L P}$, equal to 0 since independent).

3.1.1. Alternative Choice Set. In Italy, it is extremely a rare case that an extra-urban OD pair between main cities is connected by more than three paths. We did not find any two (out of more than 100) province capitals (population range is roughly from 25 000 to 2 500 000 citizens) for which a commonly used navigator suggests more than three (actually different) paths, mostly due to the Italian geography and the available road facilities.

Starting from the sample of 216 users, at most three paths are generated for each user according to three criteria: minimum travel time path; maximum motorway path; and minimum monetary cost path. These criteria have been selected as the most effective with respect to the revealed chosen path, from a pool of several other criteria.

The three criteria adopted ensure that for each user at least one generated path has been chosen by the user in 200 cases over 216 cases. Other criteria have been tested, and the second best path is generated for each criterion, but the coverage does not increase with statistical significance. The choice set covered by the first path generated according to each criterion counts 200 interviews out of 216 (93.0%).

There are 14 users with only one path generated in the perceived choice set, and they are not further considered in the parameter calibration. The choice set with at least two available alternatives generated through the three criteria counts 186 interviews out of 216 (86.1%).

3.1.2. Choice among Alternatives. The observed paths cover the different ranges of travel distance at the Italian national scale.

The main statistic characteristics of all the generated paths relative to the 186 users are

- (i) Travel time between 0.2 h and 10.3 h.
- (ii) Distance between 23 km and 1069 km.
- (iii) Monetary cost between 2 € and 190 €, due to toll and fuel.
- (iv) Percentage on motorway between 0% and 99%.

The statistics reported below do not refer to the paths generated for a single user: the variability of journey times, distances, and the percentage of use of the motorway refer to the variability among all the users belonging to the sample.

In Figure 1, a comparison between the travel time and the percentage of motorway in the path with minimum travel time is reported for all the 186 users. Long-distance paths have the tendency to have a high percentage on motorway.

3.2. Calibration or Parameter Estimation. To test the Bayesian (B) or the Classic (C) approach to calibration, two different types of information are adopted: disaggregate (D) and aggregate (A). The Logit choice model is adopted. To test the effect on the calibrated parameters, two sample are considering: 124 users (2/3 of 186) randomly extracted; 186 users (all). It simulates a dynamic context.

Considering the small number of interviews and the number of calibrated parameters, in the Bayesian approach, a numerical approximation of the prior distributions is used for obtaining the posterior distributions over a uniform grid ($50 \times 50 \times 50$) and the modes are obtained through a numerical optimization. As the numbers of the calibrated parameters and the interviews increase, the Metropolis–Hastings algorithm, say Markov chain Monte Carlo (MCMC), can be used for the Bayesian calibration.

3.2.1. Bayesian Approach. The main results of the calibration considering the Bayesian approach (B), the disaggregate data (D), the Logit specification, and the two user samples defined above are reported in Figure 2, which highlights the effects of adding 62 interviews to the initial 124. The prior of θ is a normal distribution (in the figure, the prior may look non-centered due to the range of the values on the horizontal axis). In all cases, the posterior pdfs are less dispersed than the prior pdfs. Furthermore, increasing the number of observations in the sample leads to a reduction of variance, as expected.

It is also worth noting that the probability of a parameter θ having the wrong sign decreases from the prior to the posterior as well as from the smaller to the large sample size, due to the reduction of variance.

3.2.2. Bayesian vs. Classic: Parameter Point Estimation. The main results of the calibration considering disaggregate and aggregate data are reported, respectively, in Tables 1 and 2. The estimates are given by the mode of the posterior distribution in the Bayesian approach or by the point of maximum likelihood in the Classic approach. For simplicity's sake, detailed statistics are not reported for aggregate data.

At a first glance, all calibrations look fine since each estimated parameter has the expected sign, and indeed the parameter of

- (i) The travel time is negative.
- (ii) The percentage of travel on the motorway links is positive (indicates the attraction of a high quality path).
- (iii) The label is positive (indicates the generation of the same path with several criteria).

Considering the disaggregate approach, the t-statistics in the Bayesian approach are always higher than those in the Classic approach. Furthermore, all parameters are significant at 95% significance in the Bayesian approach, while in the Classic approach, two out of three parameters are not significant with a very low t-Student statistic value for the parameter referred to as the label variable.

As expected, the logarithm of the optimal likelihood value with aggregate data is worse than the one with disaggregate data; moreover, the logarithm of the optimal likelihood value with Bayesian approach is (slightly) worse than the one with Classic approach, as expected.

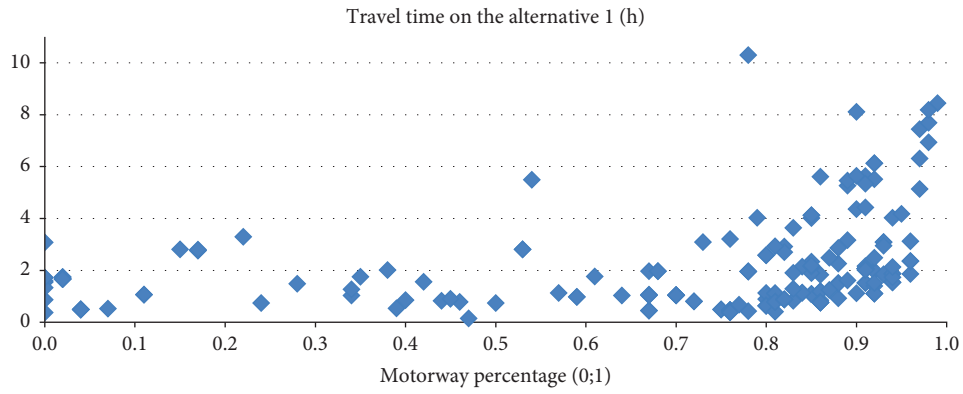


FIGURE 1: Travel time and percentage of motorway in the minimum travel time paths.

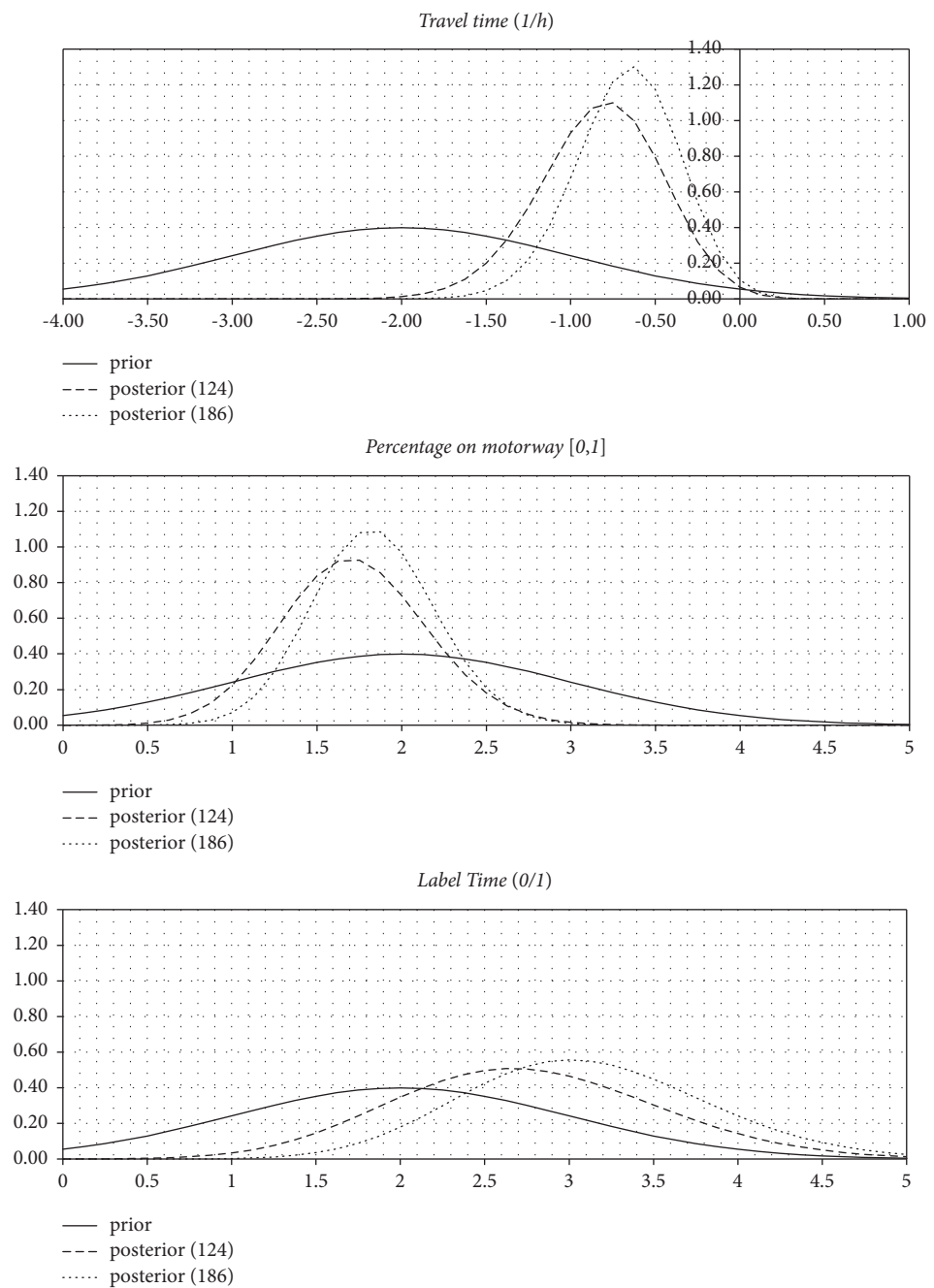


FIGURE 2: Posteriors for the calibrated parameters with disaggregate Bayesian approach.

TABLE 1: Calibration results for the path choice model, disaggregate data.

ID, approach*	1, C	2, C	3, B	4, B
Interviews	124	186	124	186
$\theta_T \in]0, +\infty[$ (1/h): travel time (<i>t</i> -Student)	-0.55 (-1.53)	-0.45 (-1.46)	-0.77 (-2.15)	-0.64 (-2.09)
$\theta_P \in [0, 1]$: percentage on motorway (<i>t</i> -Student)	1.57 (3.38)	1.78 (4.54)	1.67 (3.89)	1.78 (4.93)
$\theta_L \in \{0, 1\}$: label time (<i>t</i> -Student)	10.98 (0.15)	14.19 (0.06)	2.61 (3.26)	2.96 (4.06)
$\ln(\text{likelihood})^\wedge$	-71.71	-106.83	-72.55	-108.01
Pseudo- ρ^2	0.388	0.394	0.381	0.388

*C, Classic; B, Bayesian; \wedge statistics with parameters equal to zero; $\ln(\text{likelihood})$: -117.17 with 124 interviews; 176.36 with 186 interviews.

TABLE 2: Calibration results for the path choice model, aggregate data.

ID, approach*	5, C	6, C	7, B	8, B
Interviews	124	186	124	186
$\theta_T \in]0, +\infty[$ (1/h): travel time	-1.72	-1.76	-1.13	-1.18
$\theta_P \in [0, 1]$: percentage on motorway	0.25	0.24	0.92	0.91
$\theta_L \in \{0, 1\}$: label time	1.65	1.72	2.06	2.14

*C, Classic; B, Bayesian; $\ln(\text{likelihood})$: with parameters equal to zero = -204.34; with best parameters = -148.85; pseudo- $\rho^2 = 0.272$.

TABLE 3: Statistics of the posterior parameters with disaggregate and Bayesian approach.

ID	3	3	3	3	4	4	4	4
Interviews	124	124	124	124	186	186	186	186
Statistic	Mode	Mean	Variance	C_v	Mode	Mean	Variance	C_v
Travel time (1/h)	-0.77	-0.83	0.13	-2.26	-0.64	-0.67	0.09	-2.17
Percentage on motorway [0, 1]	1.67	1.72	0.18	4.01	1.78	1.84	0.13	5.04
Label time (0/1)	2.61	2.74	0.59	3.56	2.96	3.08	0.48	4.42

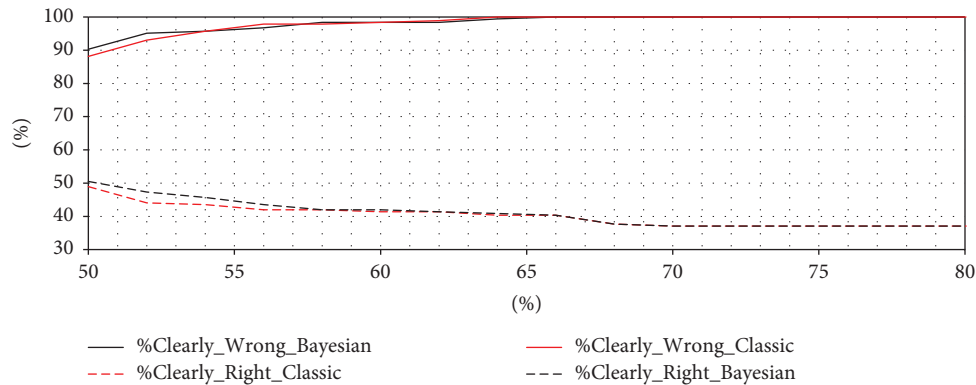


FIGURE 3: Clearly right and clearly wrong (186 users) in disaggregate approach: comparison with parameters obtained with Classic and Bayesian approaches.

For a comparison of maximum likelihood values, the traditional value of a pseudo- ρ^2 is also computed as 1 minus the ratio between the logarithm of the optimal likelihood value and of the likelihood value with all parameters equal to zero. This indicator allows comparing the results of the different methodologies by reporting the values of the indicators on the scale [0; 1]. The two estimation methods give similar values for the pseudo- ρ^2 statistic.

In all cases, the Bayesian approach provides a more stable value of the parameters. Moreover, with respect to disaggregate data, the Bayesian approach provides a much lower value of the label parameter; this result is highly

relevant as this variable describes information that is not described by the other variables (see analysis reported below).

Of course, a further advantage of the Bayesian approach over the Classic approach is the definition of the posterior probability distribution of the parameters, which support a more theoretically sound interval estimation described in the next subsection.

3.2.3. Bayesian Parameter Interval Estimation. Table 3 compares the posterior statistics, mode, mean, variance, and coefficient of variation (c_v) of the parameters calibrated by

TABLE 4: Main characteristics and utilities for the five representative users and for all samples.

<i>(a) Attributes</i>									
Criteria	MinTime			MaxMotorway			MinCost		
	$T, \gamma_{T1,n}$	$M, \gamma_{P1,n}$	$L, \gamma_{L1,n}$	$T, \gamma_{T2,n}$	$M, \gamma_{P2,n}$	$T, \gamma_{T3,n}$	$M, \gamma_{P3,n}$	$T, \gamma_{T3,n}$	$M, \gamma_{P3,n}$
User 103	1.07	0.11	0	1.11	0.58	1.12	0.00	0.00	0.00
User 102	1.57	0.42	0	1.72	0.71	2.03	0.00	0.00	0.00
User 59	1.04	0.70	0	1.14	0.78	1.45	0.00	0.00	0.00
User 88	2.70	0.82	0	2.82	0.89	3.73	0.00	0.00	0.00
User 87	5.14	0.97	1	—	—	7.84	0.00	0.00	0.00
Average (186 users)	2.12	0.69	0.37	2.29	0.83	2.99	0.00	0.00	0.00

<i>(b) Percentage contribution of the attribute component in the utilities (disaggregate, 186 users)</i>									
ID, approach	2, C			4, B			Cost		
	Motorway		Time	Motorway		Time	Motorway		Cost
Utility per attribute	$\theta_T \gamma_{T1,n}$	$\theta_P \gamma_{P1,n}$	$\theta_L \gamma_{L1,n}$	$\theta_T \gamma_{T2,n}$	$\theta_P \gamma_{P2,n}$	$\theta_L \gamma_{L2,n}$	$\theta_T \gamma_{T3,n}$	$\theta_P \gamma_{P3,n}$	$\theta_L \gamma_{L3,n}$
User 103	71%	29%	0%	78%	22%	0%	100%	0%	0%
User 102	49%	51%	0%	57%	43%	0%	100%	0%	0%
User 59	27%	73%	0%	35%	65%	0%	100%	0%	0%
User 88	45%	55%	0%	54%	46%	0%	100%	0%	0%
User 87	13%	9%	78%	41%	22%	37%	100%	0%	0%
Average (186 users)	13%	17%	71%	37%	33%	30%	100%	0%	0%

*C= Classic; B = Bayesian; T = travel time (h); M = motorway percentage (0; 1); L = label time (0/1); the alternative chosen is reported in bold.

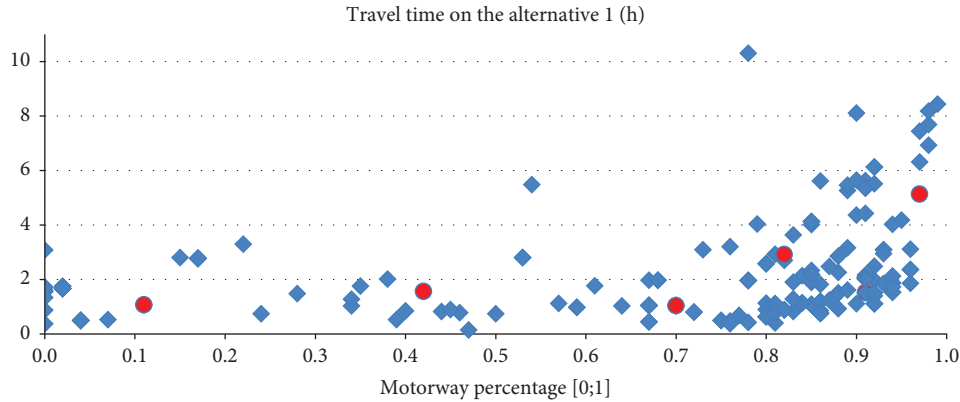


FIGURE 4: Travel time and percentage of motorway in the minimum travel time paths of the 5 users selected (red indicators).

TABLE 5: Probabilities for the five representative users (ref. parameters in Tables 1 and 2).

ID, approach*	Data**	1, C	2, C	3, B	4, B	5, C	6, C	7, B	8, B
		D	D	D	D	A	A	A	A
		Interviews	124	186	124	186	124	186	124
User 103	Time	25.9%	24.5%	25.4%	24.7%	33.9%	33.9%	30.0%	30.1%
	Motorway	53.0%	55.7%	54.2%	55.7%	35.8%	35.8%	44.4%	44.2%
	Cost	21.2%	19.7%	20.4%	19.6%	30.3%	30.3%	25.6%	25.7%
User 102	Time	35.2%	34.0%	35.9%	34.9%	45.0%	45.2%	40.1%	40.4%
	Motorway	50.7%	53.0%	51.6%	52.8%	36.7%	36.7%	43.7%	43.6%
	Cost	14.1%	13.1%	12.5%	12.3%	18.3%	18.1%	16.2%	16.0%
User 59	Time	42.7%	42.7%	43.8%	43.4%	43.9%	44.1%	43.6%	43.8%
	Motorway	45.8%	47.1%	46.3%	47.0%	37.7%	37.7%	41.9%	41.8%
	Cost	11.4%	10.2%	10.0%	9.6%	18.4%	18.2%	14.5%	14.4%
User 88	Time	45.4%	45.0%	46.7%	46.0%	50.7%	50.9%	48.0%	48.3%
	Motorway	47.5%	48.4%	47.9%	48.4%	42.2%	42.2%	44.9%	44.8%
	Cost	7.1%	6.6%	5.4%	5.6%	7.1%	6.9%	7.1%	6.9%
User 87	Time	100.0%	100.0%	99.8%	99.8%	99.9%	99.9%	99.8%	99.8%
	Motorway	—	—	—	—	—	—	—	—
	Cost	0.0%	0.0%	0.2%	0.2%	0.1%	0.1%	0.2%	0.2%

*C, Classic; B, Bayesian; **D, disaggregate; A, aggregate; the alternative chosen is reported in bold.

the disaggregate Bayesian approach. The results are compared considering the two samples. Considering the probability distribution, it is possible to evaluate the significance level for each interval. The variation reported in Table 3 is a measure of the interval estimation for a given significance level. The variance of the label variable is greater than the other variances; out of sense of completeness, values of *t*-Student significance are also reported.

3.2.4. *Clearly Right and Clearly Wrong Statistics.* Figure 3 shows the clearly right and clearly wrong statistics proposed in de Luca and Cantarella [34] for the full sample and the Classic and Bayes disaggregate approach for calibration. Clearly right (wrong) index is the percentage of users in the sample -shown on *y* axis- with a modelled probability for the chosen (not chosen) alternative greater than a predetermined threshold -shown on *x* axis-. The values are plotted with a threshold greater than 50%, considering that with lower values the statistic has no significance. The values are similar as expected since the two

approaches lead to similar models, but it is important to note that in the left area, the clear wrong of the Bayesian approach is greater (better) than the Classic approach; very often the clear right of the Bayesian approach is lower (better) than the Classic approach. They give the information that in this area, the Bayesian approach gives better results than the Classic approach with the adopted data.

3.2.5. *Analysis of Some Users.* Starting from the 186 users, 5 representative users are randomly selected with the characteristics reported in Table 4. They are sorted according to the percentage use of the motorway. The characteristics of these users are adopted for post-calibration analysis. These users are represented with a red dot in Figure 4.

In Table 4, the effect of the label variable in the Classic approach compared to the Bayesian approach is evident. To compare the percentage contribution of each attribute component (*j*) to each alternative (*k*), the percentage weight is evaluated considering the indicator:

$$w_{j,k,n} = 100 \cdot \frac{\text{abs}(\theta_k \cdot y_{j,k,n})}{\sum_{k'} \text{abs}(\theta_{k'} \cdot y_{j,k',n})}, \quad (25)$$

with θ_k the value of the k -th parameter and $y_{j,k,n}$ the value of the j -th attribute, for the alternative k , for the user n .

Considering the 4 users, numbered 103, 102, 88, and 59, the effect of the label variable is zero considering that the label attribute has a value equal to zero. Considering user numbered 87, the effect of the label variable is 78% in the Classic approach and 37% in the Bayesian approach. Also, considering the contribution for all users of the sample (all 186 users) and evaluating the average contribution, in the Classic vs. Bayesian approach, the contribution is 71% against 30%.

Considering all 186 users, the average contribution of each attribute component (j) to each alternative (k) in each expected value of the utility is as follows:

- (i) In the Classic approach, the contribution of each component is not balanced (13%, 17%, and 71% with three attributes and 41% and 59% with two attributes).
- (ii) In the Bayesian approach, the contribution of each component is better balanced (37%, 33%, and 30% with three attributes and 50% and 50% with two attributes).

This consideration shows that the unexplained information in the label variable in the Classical approach is greater than in the Bayesian approach.

This indicator does not aim to indicate greater or smaller accuracy of a method. It aims to check if the value of any parameter makes a high contribution in the utility calculation. The high contribution indicates that it alone could significantly influence the probability calculation. In practice, it is preferable to have models where the weight of the different contributions is balanced and the variation of the corresponding attributes significantly affects user choice.

Table 5 shows the probabilities computed for the 5 representative users considered with the optimized parameters reported in Tables 1 and 2. It can be observed that for the chosen alternative, the Bayes probabilities are higher than the other probabilities. In the specific case of user number 87, the disaggregate classic approach provides 100% probability for alternative 1, considering the high value of the label variable.

4. Conclusions

In this paper, path choice model calibration is tackled by comparing Bayesian (obtaining the posterior pdf of the parameters) and Classic (obtaining the maximum likelihood parameters) estimation methods and two different types of data: disaggregate (observing the choices actually made by a sample of users) and aggregate (observing the observed frequencies).

The Bayesian approach allows to obtain the posterior distribution of the parameters and therefore also allows to easily obtain the estimate by credible intervals. The posterior

distribution can be adopted for forecasting scenarios or new prior distribution if new data are available. Indeed, it is suitable for successive re-estimations when further information is collected in a dynamic context for decision making in changing environments.

As expected, from the Bayesian statistics theory, as the number of observations increases, the estimates of parameters asymptotically tend to the maximum likelihood. On the other hand, apart from very large sample, Bayesian estimators provide estimates more consistent with the sample size and allow to take into account preliminary knowledge within the prior distribution.

With reference to the adopted sample, the Bayesian and Classic estimation methods give quite similar values for the pseudo- ρ^2 statistical indicators. But, the comparison through several other indicators shows that the Bayesian estimation approach provides a better estimate than the Classic approach for several reasons, such as the calibrated parameters are more stable; the absolute values of the labels, which have a similar functioning to the specific constants of the alternatives, are lower; and the resulting models show better values of clearly right indicator.

The proposed method is rather general and can be applied and transferred to other contexts. The path choice model has general validity and with appropriate specifications can be tested in other contexts also to compare different estimation methodologies.

The results obtained refer to the sample used; in the future the proposed method must be tested with other samples, possibly referring to other contexts. In addition, the analyses should be extended by considering also choice models different from the Logit model, such as the Weibit model, as well as comparisons and integrations of econometric and psychological choice models. Application in a dynamic context is another issue worth of further research effort.

Data Availability

Data are not available due to privacy restrictions.

Disclosure

This study was performed as part of the employment of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Giulio Erberto Cantarella was responsible for supervision, conceptualization, methodology, formal analysis, and reviewing and editing. Antonino Vitetta was responsible for methodology, formal analysis, investigation, software, data curation, validation, visualization, and reviewing and editing.

References

- [1] E. Cascetta, *Transportation Systems Engineering: Theory and Methods*, Springer, New York, NY, USA, 2009.
- [2] G. E. Cantarella, D. P. Watling, S. de Luca, and R. Di Pace, *Dynamics and Stochasticity in Transportation Systems: Tools for Transportation Network Modelling*, Elsevier, Oxford, UK, 2020.
- [3] R. B. Dial, "A probabilistic multipath traffic assignment model which obviates path enumeration," *Transportation Research*, vol. 5, no. 2, pp. 83–111, 1971.
- [4] C. Manski, "The structure of random utility models," *Theory and Decision*, vol. 8, no. 3, pp. 229–254, 1977.
- [5] M. Ben-Akiva, M. J. Bergman, A. J. Daly, and R. Ramaswamy, "Modelling inter urban route choice behaviour," in *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, pp. 299–330, VNU Science Press, Utrecht, Netherlands, July 1984.
- [6] F. Russo and A. Vitetta, "Networks and assignment models for the Italian national transportation system," in *Proceedings of the 7th WCTR*, Sydney, Australia, 1995.
- [7] T. Morikawa, "A hybrid probabilistic choice set model with compensatory and non-compensatory rules," in *Proceedings of the 7th WCTR*, Sydney, Australia, 1996.
- [8] E. Cascetta, F. Russo, F. A. Viola, and A. Vitetta, "A model of route perception in urban road networks," *Transportation Research Part B: Methodological*, vol. 36, no. 7, pp. 577–592, 2002.
- [9] T. A. Domencich and D. McFadden, *Urban Travel Demand: A Behavioural Analysis*, American Elsevier, New York, NY, USA, 1975.
- [10] M. E. Ben-Akiva and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA, USA, 1985.
- [11] V. Henn, "Path choice making under uncertainty: a fuzzy logic based approach," in *Fuzzy Sets-Based Heuristics for Optimization*, J. Verdegay, Ed., pp. 277–292, Springer-Verlag, New York, NY, USA, 2003.
- [12] A. Quattrone and A. Vitetta, "Random and fuzzy utility models for road route choice," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 6, pp. 1126–1139, 2011.
- [13] M. L. De Maio and A. Vitetta, "Path choice on road transport system: a fuzzy approach," *Journal of Intelligent and Fuzzy Systems*, vol. 28, no. 5, pp. 2015–2027, 2015.
- [14] G. E. Cantarella and V. Fedele, "Fuzzy utility theory for analysing discrete choice behaviour," in *Proceedings of ISUMA 2003: Fourth International Symposium on Uncertainty Modeling and Analysis*, September 2003.
- [15] G. E. Cantarella and V. Fedele, "Fuzzy vs. random utility models: an application to mode choice behaviour analysis," in *Proceedings of the European Transport Conference 2003*, Strasbourg, France, October 2003.
- [16] A. Vitetta, "A quantum utility model for route choice in transport systems," *Travel Behaviour and Society*, vol. 3, pp. 29–37, 2016.
- [17] D. Gangi and M. Vitetta, "A Quantum utility and random utility model for path choice modelling: specification and aggregate calibration from traffic counts," *J Choice Model*, vol. 2021, p. 40, 2021.
- [18] A. A. J. Marley and J. J. Louviere, "Some probabilistic models of best, worst, and best-worst choices," *Journal of Mathematical Psychology*, vol. 49, no. 6, pp. 464–480, 2005.
- [19] P. D. Kvam, A. A. J. Marley, and A. Heathcote, "A unified theory of discrete and continuous responding," *Psychological Review*, vol. 130, no. 2, pp. 368–400, 2023.
- [20] R. H. Ewing, "Psychological theory applied to mode choice prediction," *Transportation*, vol. 2, no. 4, pp. 391–409, 1973.
- [21] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.
- [22] A. A. J. Marley, R. Duncan Luce, and I. Kocsis, "A solution to a problem raised in Luce and Marley (2005)," *Journal of Mathematical Psychology*, vol. 52, no. 1, pp. 64–68, 2008.
- [23] E. Castillo, J. M. Menéndez, P. Jiménez, and A. Rivas, "Closed form expressions for choice probabilities in the Weibull case," *Transportation Research Part B: Methodological*, vol. 42, no. 4, pp. 373–380, 2008.
- [24] E. Cascetta, A. Nuzzolo, F. Russo, and A. Vitetta, "A new path choice logit model overcoming IIA problems: specification and some calibration results for interurban networks," in *Proceedings of the 13th International Symposium on Transportation Traffic Theory, Jean-Baptiste Lesort*, Pergamon Press, Oxford, UK, 1996.
- [25] M. E. Ben-Akiva and M. Bierlaire, "Discrete choice methods and their applications to short term travel decisions," in *Handbook of Transportation Science*, R. W. Hall, Ed., pp. 5–34, Kluwer, Dordrecht, Netherlands, 1999.
- [26] C. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transportation Science*, vol. 11, no. 3, pp. 253–274, 1977.
- [27] G. E. Cantarella and M. G. Binetti, "Stochastic assignment with Gammit path choice models," in *Transportation Planning*, M. Patriksson, Ed., pp. 53–67, Kluwer Academic Publisher, Dordrecht, Netherlands, 2002.
- [28] S. de Luca, "Appendix 1: Choice Modelling Theories," in *International Encyclopedia of Human Geography*, Cantarella, Ed., University of North Carolina at Charlotte, Charlotte, NC, USA, 2020.
- [29] H. Yu, S. Zhu, J. Yang, Y. Guo, and T. Tang, "A bayesian method for dynamic origin–destination demand estimation synthesizing multiple sources of data," *Sensors*, vol. 21, no. 15, p. 4971, 2021.
- [30] Z. Zhu, S. Zhu, Z. Zheng, and H. Yang, "A generalized Bayesian traffic model," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 182–206, 2019.
- [31] Z. Zhu, A. Mardan, S. Zhu, and H. Yang, "Capturing the interaction between travel time reliability and route choice behavior based on the generalized Bayesian traffic model," *Transportation Research Part B: Methodological*, vol. 143, pp. 48–64, 2021.
- [32] S. Washington, P. Congdon, M. G. Karlaftis, F. L. Mannering, and F. L. Mannering, "Bayesian multinomial logit: theory and path choice example. Transportation research record," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2136, no. 1, pp. 28–36, 2009.
- [33] W. J. McCausland and A. A. J. Marley, "Bayesian inference and model comparison for random choice structures," *Journal of Mathematical Psychology*, vol. 62–63, pp. 33–46, 2014.
- [34] S. de Luca and G. E. Cantarella, "Validation and comparison of choice models," in *Travel Demand Management and Road User Pricing: Success, Failure and Feasibility*, W. Saleh and G. Sammer, Eds., pp. 37–58, Ashgate publications, Farnham, UK, 2009.