WORKING PAPER

ITLS-WP-18-15

Generalized multivariate extreme value

models for explicit route choice sets

By
Erik-Sander Smits[a,b], Adam J. Pel[a],
Michiel C.J. Bliemer[c] and Bart van Arem[a]

[a] Delft University of Technology, the Netherlands
[b] Arane Adviseurs in Verkeer en Vervoer, the Netherlands
[c] Institute of Transport and Logistics Studies (ITLS),
 The University of Sydney Business School, Australia

August 2018

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in

Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER:	Working Paper ITLS-WP-18-15

TITLE: **Generalized multivariate extreme value models for explicit route choice sets**

ABSTRACT: This paper analyses a class of route choice models with closed-form probability expressions, namely, Generalized Multivariate Extreme Value (GMEV) models. A large group of these models emerge from different utility formulas that combine systematic utility and random error terms. Twelve models are captured in a single discrete choice framework. The additive utility formula leads to the known logit family, being multinomial, path-size, paired combinatorial and link-nested. For the multiplicative formulation only the multinomial and path-size weibit models have been identified; this study also identifies the paired combinatorial and link-nested variations, and generalizes the path-size variant. Furthermore, a new traveller's decision rule based on the multiplicative utility formula with a reference route is presented. Here the traveller chooses exclusively based on the differences between routes. This leads to four new GMEV models. We assess the models qualitatively based on a generic structure of route utility with random foreseen travel times, for which we empirically identify that the variance of utility should be different from thus far assumed for multinomial probit and logit-kernel models. The expected travellers' behaviour and model-behaviour under simple network changes are analysed. Furthermore, all models are estimated and validated on an illustrative network example with long distance and short distance origin-destination pairs. The new multiplicative models based on differences outperform the additive models in both tests.

AUTHORS: *Smits, Pel, Bliemer and Arem*

CONTACT: INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (H73)

The Australian Key Centre in Transport and Logistics Management

The University of Sydney   NSW 2006   Australia

Telephone: +612 9114 1824

E-mail: business.itlsinfo@sydney.edu.au

Internet: http://sydney.edu.au/business/itls

DATE: August 2018

1 Introduction

Route choice is important in transport applications such as network equilibrium modelling, day-to-day route choice decisions, and route guidance. The literature describes various methods to model the route choice behaviour of travellers. They range from simple deterministic shortest route choice to sophisticated stochastic models in which random error terms capture travel time uncertainty and taste heterogeneity amongst travellers. Commonly, the route choice model is applied in an iterative process, for example, to reach equilibrium in a congested network or for en-route decisions. For large networks the number of times route choice has to be simulated is very high. Therefore, a good balance between realism and computational efficiency is required.

1.1 Random utility maximization

A common choice mechanism is that travellers consider multiple routes, assign a subjective utility to each route, and choose the route with the highest utility. The discrete choice framework based on random utility maximization (RUM) allows random components in the utility formulations. In this case the utility of each route follows some random distribution. Observe that a route consists of a set of links, and we assume that a route's utility is (among other things) determined by the characteristics of the corresponding links. Then the joint probability distribution of route utilities contains dependencies when routes have overlap, and thus share (dis)utility from the same link(s). Furthermore the distribution is heteroscedastic (i.e., the variability of utility differs amongst routes) since routes have different lengths. Section 2 provides an in-depth analysis of desired random route utility properties, that especially explores the distribution of foreseen travel times besides the usual analyst error. This gives new insights into the variance and covariance structure of route utility.

The Multinomial Probit (MNP) model for route choice (see Daganzo and Sheffi, 1977; Yai et al., 1997) can easily address correlation and heteroscedasticity, but no closed-form formulation of the route probabilities exists. The latter leads to computationally expensive simulations. On the other hand the Multinomial Logit (MNL) model for route choice (see Daganzo and Sheffi, 1977)[1] has an elegant closed-form formulation for the route probabilities; however, route utilities have to be independent and homoscedastic. Extensions to mixed logit models (Bekhor et al., 2002) are possible to overcome some of the limitations, however, these models again have to rely on numerical solution methods requiring simulations which are often infeasible on large scale networks. In addition, the thus far proposed covariance structures in MNP and mixed logit are not in line with revealed foreseen travel times; the *variance* has been assumed proportional to the mean, whereas instead, data shows a linear relation between *standard deviation* and mean (see Section 2.1).

Several adaptations to the MNL model exist to address correlation due to route overlap (Prato, 2009). They either insert a correction term into the utility, or exploit the more general Multivariate Extreme Value (MEV) distribution for the error term. In Section 3.2 an overview of these methods is presented. The heteroscedasticity is less addressed in the literature. In practice, the scale parameter of MNL is sometimes considered to be proportional to the distance between origin and destination, but this does not address heteroscedasticity within the origin-destination (OD) route set (see Chen et al., 2012). Less pragmatic is the solution by Castillo et al. (2008) who assume Weibull distributed utilities leading to the Multinomial Weibit (MNW) model. Only slightly different is the approach of Gálvez (2002); Fosgerau and Bierlaire (2009) who use a multiplicative error term instead of an additive error term. The probabilities of the earlier-known Kirchhoff distribution of routes as presented by Fellendorf and Vortisch

---

[1] Daganzo and Sheffi (1977, Eqn. 10) are the first to present MNL route choice probabilities, based on the method of Dial (1971)

(2010) also coincide with MNW. Recently, Kitthamkesorn and Chen (2013) included the path-size factor into the MNW model to correct for correlation between routes.

1.2 Generalized Multivariate Extreme Value models

In this paper the state-of-the-art of models with closed-form utility formulations is reviewed, and they are gathered in a framework with a single, general form for the choice probabilities. Two classes of MEV can be identified – each based on a different utility formula –, which are unified in what we call Generalized Multivariate Extreme Value (GMEV). The generalization gives rise to three undiscovered route choice models based on the multiplicative utility formula. Furthermore, we recall that equivalent MEV formulations exist for models in which a correction term is inserted in the utility formulation (e.g. Path-Size Logit). All these models can therefore be captured in the GMEV framework based on either additive or multiplicative utility formulas. Section 4 assesses these models qualitatively by analysing their distributions and comparing them with the desired properties formulated in Section 2.3. Additive models prove not to be able to capture random foreseen travel time, but they do capture the analyst error. On the other hand, multiplicative models can handle both errors, but based on the same distribution (i.e., the foreseen travel time and analyst error are completely dependent).

One of the desired properties cannot be fulfilled by any existing model. Namely, in a transport network model the route choice model is applied to many route sets, so it is important that the model gives realistic results for all route sets existing in a network. In particular, the results should remain realistic when the network and/or routes change. With respect to this, both MNL and MNW have an undesired property. In MNL, the route probabilities will remain equal if a constant is added to each route's utility. In MNW, the route probabilities will remain equal if each route's utility is multiplied with a constant. Both these properties are unrealistic. Therefore, Xu et al. (2015) provide a hybrid method where the choice probabilities have a MNL en MNW component; however, this hybrid method stems from a combined impedance function (i.e., it defines the choice probabilities based on the choice probability functions of logit and weibit). Although the model can be useful in practice, the model does not stem from a utility formulation with an additive and multiplicative error term (Section 3.1.3 discusses how the hybrid models would fit in the GMEV framework).

To partially overcome the shortcomings of MNW, we introduce new GMEV route choice models based on the multiplicative utility formula and a reference route. The probability of choosing the reference route only depends on the non-overlapping differences with other routes, since the overlap is explicitly removed. The models fit in the same framework as the existing models. Furthermore, they are designed to be suitable under changing networks and routes. A qualitative assessment and a numerical benchmark on a small network example of the existing and new models show that the new models can resemble expected behaviour under more types of network changes than existing closed-form models can. For example, the model performs well if all route costs are multiplied with a factor, but also if a constant cost is added to all routes.

1.3 Route set generation

This paper assumes that a relevant route set is explicitly given. For an overview of route set generation techniques, see Frejinger et al. (2009); Prato (2012). However, it should be noted that the generation and composition of the route set has a large influence on model outcomes (Bliemer and Bovy, 2008; Cascetta et al., 2002; Prato, 2012). Another disadvantage with explicit route sets is that a correct sample of routes is required to obtain unbiased parameter estimates (Frejinger et al., 2009). For discussions on sampling

routes based on distributions derived from route choice models see (Frejinger et al., 2009; Flötteröd and Bierlaire, 2013; Guevara and Ben-Akiva, 2013). Bierlaire et al. (2008) points out difficulties with selection bias in the estimation of additive MEV models from choice based samples. A recently revisited different approach – that overcomes route set generation/sampling related problems – is to implicitly generate routes as done by Dial (1971); Papola and Marzano (2013); Fosgerau et al. (2013a); Mai et al. (2015). However, these models either have a restricted route set (Dial, 1971; Papola and Marzano, 2013) (see Section 3.2.5), or contain all routes, even those with loops (Fosgerau et al., 2013a; Mai et al., 2015). In addition, no methods with implicit choice set generation that can handle non-additive travel costs are known to exist by the authors of this paper.

Since explicit finite route sets have practical advantages, they are commonly used in traffic assignment models to either avoid too computationally expensive (dynamic) shortest-path calculations, to allow dynamic network loading of path flows, or to allow for traffic assignment models with non-additive link costs. For example, Zhou et al. (2015) generates an a-priori route set in a static assignment context using the deterministic method of Bekhor et al. (2006), Bliemer et al. (2014) generates an a-priori route set for a quasi-dynamic model using the stochastic method of Fiorenzo-Catalano et al. (2004).

1.4 Contribution

The contribution of this paper is summarized as follows. First, we describe the desired properties of random route utility which includes random foreseen travel times, and empirically verify the linear relation between mean and standard deviation of foreseen travel time. Second, existing closed-form route choice models based on random utility maximization are presented in a novel GMEV framework and it is shown that all existing closed-form models with explicit choice sets fit in the GMEV framework. Third, the unexplored area of multiplicative MEV models gives rise to three new closed-form route choice models, and the proposed multiplicative MEV models based on a reference route give rise to another four new closed-form route choice models. The working of this model is illustrated with a small network example, and its behaviour with respect to network changes is compared to the behaviour of existing models. Fourth, we show that the additive models, contrary to multiplicative models, can not capture random foreseen travel time. Fifth, we show the strength of a unified GMEV model by providing its stochastic user equilibrium formulation. Sixth, all twelve models (of which seven are new) are estimated on a carefully constructed small network example with multiple choice sets for origin-destination pairs with both short and long distances. The estimation is done twice with two different synthetic datasets generated by MNP simulations based on the desired properties of random route utility, this allows benchmarking the models on the other dataset. The benchmark provides insight in the performance of the model if its applied on different networks, without re-estimation. Overall, when the models are estimated on one synthetic dataset and validated on the other synthetic dataset, the new models based on reference routes have a better validation than existing models. They can better approximate – without simulations – the MNP probabilities that take, amongst others, heteroscedasticity and correlation into account.

2 Random Route Utility Formulation

The random utility maximisation (RUM) framework is commonly used for describing route choice. Each route alternative is associated with some utility and the traveller is assumed to choose the route that provides the maximum expected utility. Since utility is not directly observable, analysts typically

assume a structure of utility that includes measurable components and random terms that describe unmeasurable components. This section provides a route utility formulation with random foreseen travel times, that is used in Section 4 to assess the GMEV models.

Let R denote the set of relevant routes between a certain origin-destination (OD) pair. The systematic utility V_r for a route $r \in R$ is decomposed in a part based on travel time τ_r, and 'other' systematic utility V_r^0, thus $V_r = V_r^0 + \beta \tau_r$, where τ_r is a random variate representing foreseen travel time, β is the travel time parameter. Then we write the random utility for route $r \in R$ as[2],

$$U_r = V_r + \varepsilon_r = V_r^0 + \beta \tau_r + \varepsilon_r, \tag{1}$$

where ε_r is the analyst error. V_r^0 is assumed to be deterministic and can contain all route attributes other than the foreseen travel time, such as travel distance, running cost, number of traffic lights encountered, number of left turns, travel time variability, etc. Typical random component ε_r includes attributes that are considered by travellers when choosing their routes, but that are not included in $V_r^0 + \beta \tau_r$. Since the analyst does not have perfect knowledge on how the traveller gathers information about its travel time, we explicitly consider foreseen travel time as a random distribution τ_r in this work. In addition, travellers can have different preferences with respect to travel time and they do not perfectly measure the actual values of the attribute levels, but rather have a subjective perception that deviates from the actual values. Denote the analyst's estimate of foreseen travel time with $\hat{\tau}_r$, which generally does not consider all conditions (e.g., exact departure time, weather, information sources) the traveller is aware of. The deviation between the actual and estimate of foreseen travel time is typically larger for longer trips (see next Section). For example, when asking a traveller about the travel time of a recent short trip, his or her expectation may be off by one or two minutes, while for a long trip the expectation may be off by 10 minutes or more. We would like to point out that the foreseen travel time distribution does not capture travellers' disproportionate resilience towards high travel time variability, which can be captured by an (un)reliability term in V_r^0.

The utility in Equation (1) is very general since the analyst can include any attribute in V_r^0, and the, in our opinion, two most important sources of randomness for route utility are captured. Extension for errors around other terms in the systematic utility can be made in a straightforward fashion. We focus in this paper on the foreseen travel time because it is the most important factor in route choice. In addition, data is available to support assumptions on its structure.

Both τ_r and ε_r are random variables and are described by probability distributions. A common assumption for analyst errors ε_r is that they are independently and identically distributed (i.i.d.). The next section discusses the foreseen travel time distribution in more detail. Since τ_r and ε_r capture different sources of randomness, they can be considered independent, i.e. $\mathrm{Cov}(\tau_r, \varepsilon_r) = 0$ for all $r \in R$.

2.1 Structure of foreseen travel time

Consider two similar road segments with lengths respectively 1 and 5 kilometres, and suppose the distribution of foreseen travel times for the 1-km road segment is known. There are two natural ways to determine the foreseen travel time distributions of the 5-km road segments by, which impose different relations between their means and variances. The first method – used in all previous Probit studies (e.g., Yai et al., 1997; Daganzo and Sheffi, 1977) – is based on link-additivity and postulates that the foreseen travel time of the 5-km road segment is equally distributed as the sum of foreseen travel times

---

[2]The OD pair and individual are not indexed since they are not relevant for this study and omitting them makes the formulas more readable.

of five independent 1-km road segments. This implies for each set of road segments for which the *additivity postulate* holds, that the ratio between the mean and the *variance* is equal for each segment. Link-additivity is convenient in transport networks, since overlap and heteroscedasticity of routes can automatically be captured when the normal distributed travel costs are drawn for each link and then summed per route. The second method is based on scaling and postulates that the foreseen travel time of the 5-km road segment is equally distributed as five times the foreseen travel time of one 1-km road segment. This implies for each set of road segments for which the *scaling postulate* holds, that the ratio between the mean and the *standard deviation* is equal for each segment. Furthermore, the convenience of link-additivity does not apply under the scaling postulate.

When an analyst wants to specify means and variances of foreseen travel times of routes, these two postulates each imply a different structure. Under the natural assumption that the mean equals the analyst's estimate $\hat{\tau}_r$, the variance is linear in $\hat{\tau}_r$ under the linear-additivity postulate and quadratic in $\hat{\tau}_r$ under the scaling postulate. In order to examine which postulate is more plausible, we first analyse the relationship between mean travel times and standard deviation of travel times in two datasets, and then discuss findings in the literature related to the postulates.

We have analysed household travel survey data from the Netherlands for the years 2010, 2011, and 2012 (Onderzoek Verplaatsingen in Nederland, OViN). In this survey, travellers are asked to state the travel times from trips they have made during a single day. We pooled data from trips with motorised private vehicles (i.e., car-driver, car-passenger, and motor bicycle) per 4-digit postal code and municipality[3]. After excluding OD pairs with less than 25 observations we obtained 686 municipality aggregated OD pairs and 216 postal code aggregated OD pairs containing 38,106 and 8,079 trips, respectively. Figure 1 shows the relationship between the average travel time and the standard deviation of the perceived travel times for each OD pair. The data shows a strong correlation and a linear relation between the average travel time and the standard deviation of travel time. Under the assumption that the traveller had complete information, the perceived travel time and foreseen travel time coincide, and the data is then in accordance with the scaling postulate. Clearly, the spread in reported travel times here is also the result of variations in traffic conditions and travel demand, routing, and aggregation of different households in each OD pair. Unfortunately, we do not have the exact data on the foreseen travel time per route in the household travel surveys; however, we do have data from another smaller study about route choice and information.

This second data set comes from a study where 32 commuters between The Hague and Delft in the Netherlands were asked about the used information sources (see Ramos et al., 2012; Ramos, 2015). Some of them were equipped with GPS devices for real-time information. After every trip they made, they were asked to indicate what their expected route and foreseen travel time for their next trip are. We aggregated this data per route per respondent, and determined the mean and standard deviation for all routes with at least 10 entries. In total 407 entries for 21 routes from 19 respondents[4] are used for the results in Figure 2. This dataset also indicates a linear relationship between the average and standard deviation. Thus both datasets indicate that the scaling postulate is more plausible than the additivity postulate.

Mahmassani et al. (2012) investigate travel time variability based on GPS and simulated trajectories and they conclude that a linear relation exists between the standard deviation of observed travel times and the average travel times. They furthermore reject a square root and quadratic relation. Mahmassani et al. (2013) provide the same conclusion at a network level, rather than just at link level. Fosgerau

---

[3]The four large municipalities Amsterdam, Utrecht, The Hague and Rotterdam are subdivided in respectively 7, 3, 5 and 5 districts.

[4]Two respondents reported two routes at least 10 times; 17 other respondents reported one route at least 10 times.
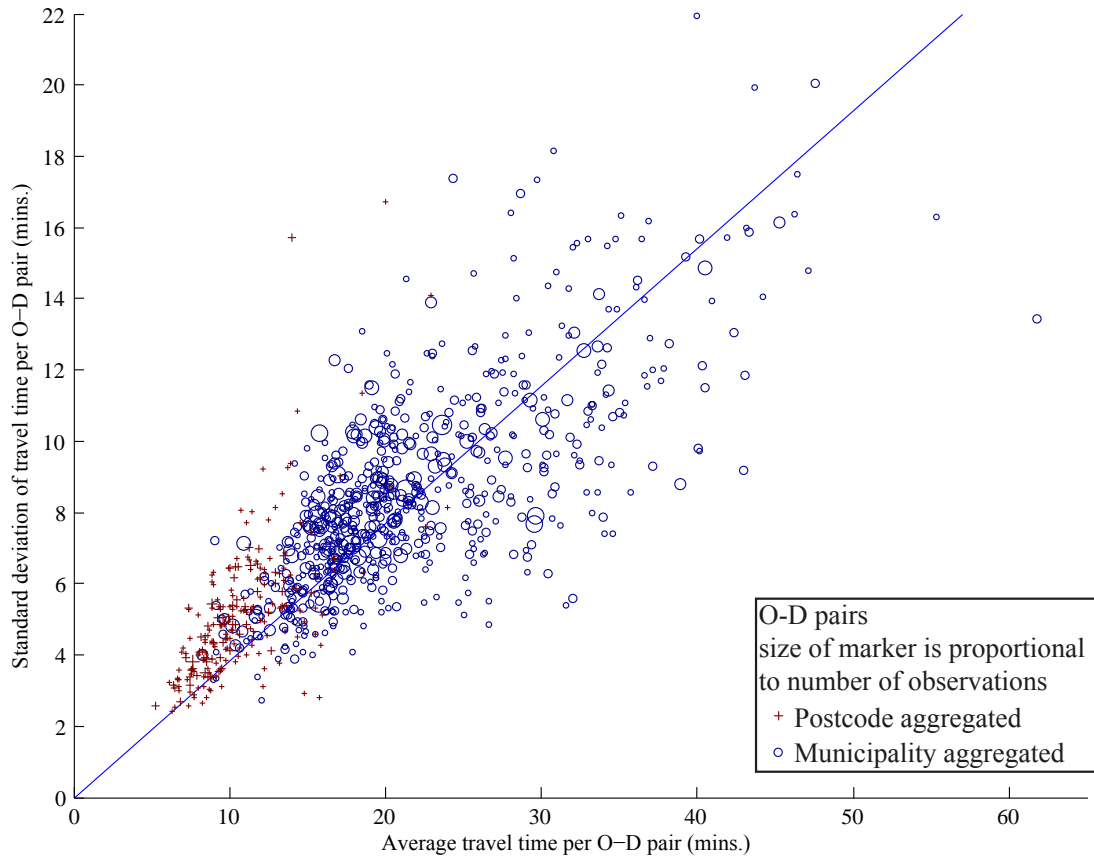
Figure 1: Relation between the average and the standard deviation of revealed perceived travel times for an OD pair, shown for 216 postcode OD pairs and 686 municipality OD pairs from Dutch national travel survey data. Linear regression between mean and standard deviation has a R-square of 0.9379. Linear regression between mean and variance has a R-square of 0.8016. The linear regression line has slope 0.3859.

Figure 2: Relation between the average and the standard deviation of foreseen travel times for a route from a respondent, shown for 21 routes from 19 respondents from trips between Delft and The Hague. The marker size is proportional to the number of entries per route. Linear regression between mean and standard deviation has an R-square of 0.9300. Linear regression between mean and variance has an R-square of 0.8401. The linear regression line has slope 0.1301.
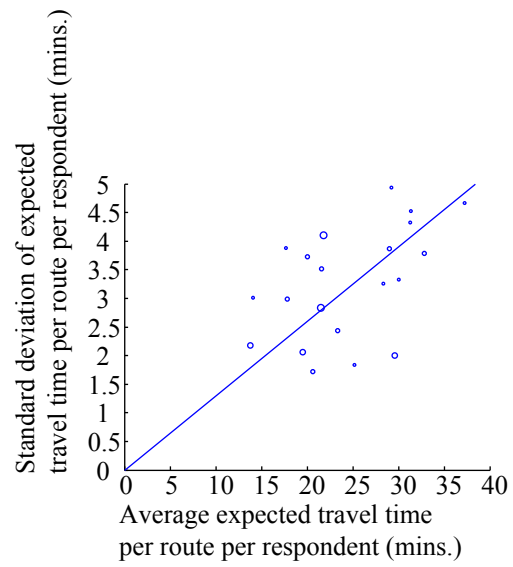
([2010](#)) and [Yildirimoglu et al. (2013)](#) use data on time-of-day dependent relations between mean and standard deviation to show hysteresis. Their plots also strongly suggest a linear relation. Furthermore, [Hellinga et al. (2012)](#) took a year of measurements from loop detector data on a motorway and aggregated each day into 15-minute averages. They also show a strong linear relationship between the standard deviation of the observed travel times and the average travel times. Despite that none of these studies analysed *foreseen* travel time distributions, they all indicate that the scaling postulate holds for their travel times. These results, the OViN data, and the actual foreseen travel time data, should convince that the scaling postulate is more plausible than the additivity postulate for distributions of foreseen route travel times.

Write $\sigma(\tau_r)$ as the standard deviation of foreseen travel time. Let this standard deviation be proportional to $\hat{\tau}_r$ with proportionality constant θ, then $\sigma(\tau_r) = \theta \hat{\tau}_r$

This finding contradicts the assumptions for MNP route choice models made by [Daganzo and Sheffi (1977)](#) and [Yai et al. (1997)](#). Their assumption of a linear relationship between the variance of travel times and $\hat{\tau}_r$, stems from the additivity postulate. In these models it is practical that route travel time can be readily obtained from the summation of normal distributed link travel times, but this seems to be unrealistic. It is possible to adjust the MNP-model to make it compatible with the scaling postulate by directly determining the (co)variance-matrix for all routes. However, this comes at the cost of link-additivity.

2.2 Covariances of foreseen travel times

Since routes may overlap, foreseen travel times are correlated. Route overlap induces a structure of covariances between travel times. This means that if two routes are largely overlapping, then also the foreseen travel time of both routes should be almost identical. Consider two routes, $r, s \in R$. Let $\hat{\tau}_r = \hat{\tau}_{rs} + \hat{\tau}_{r \setminus s}$ be the decomposed travel time estimates of route r, where $\hat{\tau}_{rs}$ is the overlapping part with route s, and $\hat{\tau}_{r \setminus s}$ the non-overlapping part. The covariance between τ_r and τ_s should capture the overlap. We propose two candidate formulations for the covariance. One based on a correlation formulation derived from the literature, and one based on variance distribution over links.

The approaches from [Daganzo and Sheffi (1977)](#); [Yai et al. (1997)](#); [Bekhor et al. (2002)](#) rely on the additivity postulate, and the additivity over links leads to a definition of covariances. Under the scaling postulate these covariances are not directly usable due to the different relation between mean and standard deviation. However, the corresponding correlation, $\hat{\tau}_{rs}/\sqrt{\hat{\tau}_r \hat{\tau}_s}$, is independent of any relationship between mean and standard deviation, and can be used together with $\sigma(\tau_r) = \theta \hat{\tau}_r$. Then from the definition of correlation it follows that

$$\text{Corr}(\tau_r, \tau_s) = \frac{\hat{\tau}_{rs}}{\sqrt{\hat{\tau}_r \hat{\tau}_s}} = \frac{\text{Cov}(\tau_r, \tau_s)}{\sigma(\tau_r)\sigma(\tau_s)} \tag{2}$$

$$\Downarrow$$

$$\text{Cov}(\tau_r, \tau_s) = \theta^2 \hat{\tau}_r \hat{\tau}_s \frac{\hat{\tau}_{rs}}{\sqrt{\hat{\tau}_r \hat{\tau}_s}} = \theta^2 \hat{\tau}_{rs} \sqrt{\hat{\tau}_r \hat{\tau}_s}. \tag{3}$$

An alternative formulation of the covariance can be obtained by using the definition $\text{Var}(\tau_r - \tau_s) = \text{Var}(\tau_r) + \text{Var}(\tau_s) - 2\text{Cov}(\tau_r, \tau_s)$ and an approximation of variance of the non-overlapping parts of travel time. Under the assumption that a route's variance is equally 'distributed' over the total travel time and that τ_r can be decomposed in independent parts, $\tau_r = \tau_{rs} + \tau_{r \setminus s}$, the variance associated with the

non-overlapping part equals

$$\mathrm{Var}(\tau_{r\backslash s}) = \frac{\hat{\tau}_{r\backslash s}}{\hat{\tau}_r}\mathrm{Var}(\tau_r) = \frac{\hat{\tau}_{r\backslash s}}{\hat{\tau}_r}\theta^2\hat{\tau}_r^2 = \theta^2\hat{\tau}_r\hat{\tau}_{r\backslash s}. \tag{4}$$

Then the covariance equals

$$\mathrm{Cov}(\tau_r, \tau_s) = \frac{1}{2}\left(\mathrm{Var}(\tau_r) + \mathrm{Var}(\tau_s) - \mathrm{Var}(\tau_r - \tau_s)\right) \tag{5}$$

$$= \frac{1}{2}\left(\mathrm{Var}(\tau_r) + \mathrm{Var}(\tau_s) - \mathrm{Var}(\tau_{r\backslash s} - \tau_{s\backslash r})\right) \tag{6}$$

$$= \frac{1}{2}\theta^2\left(\hat{\tau}_r^2 + \hat{\tau}_s^2 - \hat{\tau}_r\hat{\tau}_{r\backslash s} - \hat{\tau}_s\hat{\tau}_{s\backslash r}\right) \tag{7}$$

$$= \frac{1}{2}\theta^2\left(\hat{\tau}_r(\hat{\tau}_r - \hat{\tau}_{r\backslash s}) + \hat{\tau}_s(\hat{\tau}_s - \hat{\tau}_{s\backslash r})\right) \tag{8}$$

$$= \theta^2\hat{\tau}_{rs}\frac{\hat{\tau}_r + \hat{\tau}_s}{2}. \tag{9}$$

Notice that the covariance formulations of Equations (3) and (9) are respectively based on the geometric mean and arithmetic mean of travel times $\hat{\tau}_r$ and $\hat{\tau}_s$. Note that although we base the covariance on the travel time and its overlap, it is also possible to define it based on overlap of distance or total systematic utility.

2.3   Desired choice model properties

In summary, the assumptions on the error terms and the utility formulation in Equation (1) determine the expected value, variance and covariance of the utility.

- The expected value of utility of route $r \in R$ is

$$\mathbb{E}(U_r) = \mathbb{E}\left(V_r^0 + \beta\tau_r + \varepsilon_r\right), \tag{10}$$

$$= \mathbb{E}\left(V_r^0\right) + \mathbb{E}\left(\beta\tau_r\right) + \mathbb{E}(\varepsilon_r), \tag{11}$$

$$= V_r^0 + \beta\hat{\tau}_r + \mathbb{E}(\varepsilon_r). \tag{12}$$

- The variance of the utility of route $r \in R$ is

$$\mathrm{Var}(U_r) = \mathrm{Var}\left(V_r^0 + \beta\tau_r + \varepsilon_r\right), \tag{13}$$

$$= \mathrm{Var}\left(V_r^0\right) + \mathrm{Var}\left(\beta\tau_r\right) + \mathrm{Var}(\varepsilon_r), \tag{14}$$

$$= \theta^2\hat{\tau}_r^2 + \mathrm{Var}(\varepsilon_r). \tag{15}$$

- The covariance between routes $r, s \in R$, $\mathrm{Cov}(U_r, U_s) = \mathrm{Cov}(\tau_r, \tau_s)$, equals either Equation (3) or (9).

Furthermore, when the route choice model is part of a transport network model, the following properties are desired:

- applicability to and compatibility between all types of (changing) networks, including those with OD pairs with short and long distances, and including those with changing differences and ratios between routes;

9

- closed-form probabilities to avoid computationally expensive simulations;

- capture correlations due to overlap.

3 Random Route Utility Maximization models with Generalized Multivariate Extreme Value distributions

In this section an overview is given of choice models that are applicable for routes based on random utility maximization (RUM) with closed-form expressions for the probabilities. This implies that MNP, Mixed Logit, and models with an error correction component (e.g., Logit kernel) are not considered; we refer to (Prashker and Bekhor, 2004; Frejinger and Bierlaire, 2006) for an overview of these simulation based models. However, since the MNP model is capable of completely capturing both the analyst error and the random foreseen travel time (and thus heteroscedasticity and correlation)[5], we will use MNP simulated data as ground truth in the illustrative example in Section 7. The existing models are all based on a single random variable, which should approximate both the analyst error and travel time distribution; Section 4 answers to which degree this is possible.

In this section the family of distributions will be deduced from RUM models with separable systematic and random utility components (i.e., U_r can be written as a function of V_r and ε_r). This leads to two types of Extreme Value (EV) models. First consider the additive utility formula $U_r = V_r + \varepsilon_r, \forall r \in R$, where $\left(\varepsilon_1, \ldots, \varepsilon_{|R|}\right)$ is Multivariate Extreme Value (MEV) distributed (McFadden, 1978); these models fall in the EV type I (Gumbel) category. Second, consider the multiplicative utility formula $U_r = V_r \varepsilon_r, \forall r \in R$ and assume that $\left(-\ln \varepsilon_1, \ldots, -\ln \varepsilon_{|R|}\right)$ is MEV distributed (Gálvez, 2002; Fosgerau and Bierlaire, 2009); these models fall in the EV type III (reversed Weibull) category. The EV type II (Fréchet) is not suitable for route choice since Fréchet distributions have a lower bound on its support (i.e., there is a maximum on the route cost/travel time if such a model existed). The Generalized Multivariate Extreme Value (GMEV) models consist of both the additive and multiplicative MEV models. Note that in older literature the MEV models were also named GEV. GEV is a family of univariate distributions consisting of Gumbel, Fréchet and reversed Weibull. An MEV distribution is the joint distribution of multiple random variables with marginal distribution from one GEV type. MEV type I was discovered first by McFadden (1978) under the name 'The Generalized Extreme Value model'.

Other parametric approaches – not based directly on separable RUM – also exist. In (Castillo et al., 2008; Li, 2011; Nakayama, 2013; Nakayama and Chikaraishi, 2015; Chikaraishi and Nakayama, 2015) utilities $U_r, \forall r \in R$ are independently Generalized Extreme Value (GEV) distributed. Li (2011) captures even a larger class of models, with independent route utilities. Fosgerau et al. (2013b) and Mattsson et al. (2014) provide the largest class of parametric approaches, where dependencies can be included as well. Note that all these models cannot be decomposed in systemtic and random components[6]. They directly parametrize the utility's probability distributions which is less explainable in behavioural terms. The exception is the derivation of the revised q-generalized logit model by Chikaraishi and Nakayama (2015) from the q-product function.

---

[5]Note that we use an adjusted MNP (compared to the literature) that is in line with the (co)variances under the scaling postulate.

[6]Except for their specific cases where they fit in the GMEV type.

10

3.1 The two types of Multivariate Extreme Value distributions

MEV distributions are very flexible since they allow for correlation between their variables. Generating functions are the core of MEV distributions and unfortunately these are not easily interpretable, making the theory rather complex. It is well-known that MEV type I models capture the MNL, Paired Combinatorial Logit (PCL) and Link-Nested Logit (LNL) models, but it is less known that C-Logit, Path-Size Logit (PSL) and Path-Size Correction Logit belong to the MEV family. It is convenient that the latter models also fit in the generic framework since general results can be applied to them. All these mentioned 'logit based' models have an additive utility formula, so first the MEV derivation for this additive case is presented. Thereafter, the multiplicative MEV models are derived, and we show how they both fit in the GMEV framework.

3.1.1 Additive MEV models

The additive utility formula is $U_r^{\mathrm{A}} = V_r + \varepsilon_r^{\mathrm{A}}, \forall r \in R$. Define the probability of choosing route $r \in R$ as

$$P_r^{\mathrm{A}} := \mathrm{P}\left(U_r^{\mathrm{A}} \geq U_p^{\mathrm{A}}, \quad \forall p \neq r \right) \tag{16}$$

$$= \mathrm{P}\left(V_r + \varepsilon_r^{\mathrm{A}} \geq V_p + \varepsilon_p^{\mathrm{A}}, \quad \forall p \neq r \right). \tag{17}$$

Following McFadden (1978), assume that the joint distribution of $\left(\varepsilon_1^{\mathrm{A}}, \ldots, \varepsilon_{|R|}^{\mathrm{A}} \right)$ follows MEV type I distribution $\boldsymbol{\varepsilon}^{\mathrm{MEV}}$ with cumulative distribution function

$$F_{\boldsymbol{\varepsilon}^{\mathrm{MEV}}}(x_1, \ldots, x_{|R|}) = e^{-G(e^{-x_1}, \ldots, e^{-x_{|R|}})}, \tag{18}$$

where *generating function* $G : \mathbb{R}^{|R|} \to \mathbb{R}$ satisfies

- μ-homogeneity:

$$G(a z_1, \ldots, a z_{|R|}) = a^{\mu} G(z_1, \ldots, z_{|R|}), \quad \forall a > 0 \tag{19}$$

- Limit property:

$$\lim_{z_r \to \infty} G(z_1, \ldots, z_{|R|}) = \infty, \quad \forall r \in R \tag{20}$$

- Alternating signs:

$$\frac{\partial^{|\hat{R}|} G(z_1, \ldots, z_{|R|})}{\prod_{r \in \hat{R}} \partial z_r} (-1)^{|\hat{R}|-1} \geq 0, \quad \forall \hat{R} \subseteq R, \hat{R} \neq \emptyset. \tag{21}$$

The alternating signs property means that the sign of G switches every time an additional distinct partial derivative is taken. The generating function needs to satisfy the properties in order for $F_{\boldsymbol{\varepsilon}^{\mathrm{MEV}}}$ to be a well-defined multivariate cumulative distribution function. With this utility formulation and distribution the following closed-form choice probabilities can be derived:

$$
\begin{aligned}
P_r^{\mathrm{A}} &= \mathrm{P}\left(U_r^{\mathrm{A}} \geq U_p^{\mathrm{A}}, \quad \forall p \neq r \right) \\
&= \frac{y_r^{\mathrm{A}} G_r(y_1^{\mathrm{A}}, \ldots, y_{|R|}^{\mathrm{A}})}{\mu G(y_1^{\mathrm{A}}, \ldots, y_{|R|}^{\mathrm{A}})},
\end{aligned}
\tag{22}
$$

where $G_r(x_1, \ldots, x_{|R|}) = \frac{\partial G(x_1, \ldots, x_{|R|})}{\partial x_r}$ and $y_r^{\mathrm{A}} = e^{V_r}, \forall r \in R$. More details on the generating function and the derivation of the choice probabilities can be found in (McFadden, 1978). By using Euler's homogeneous function theorem Equation (22) can be rewritten as

$$P_r^{\mathrm{A}} = \frac{y_r^{\mathrm{A}} G_r(y_1^{\mathrm{A}}, \ldots, y_{|R|}^{\mathrm{A}})}{\sum_{p \in R} y_p^{\mathrm{A}} G_p(y_1^{\mathrm{A}}, \ldots, y_{|R|}^{\mathrm{A}})} \tag{23}$$

$$= \frac{e^{V_r} G_r(e^{V_1}, \ldots, e^{V_{|R|}})}{\sum_{p \in R} e^{V_p} G_p(e^{V_1}, \ldots, e^{V_{|R|}})} \tag{24}$$

$$= \frac{e^{V_r + \ln G_r(e^{V_1}, \ldots, e^{V_{|R|}})}}{\sum_{p \in R} e^{V_p + \ln G_p(e^{V_1}, \ldots, e^{V_{|R|}})}} \tag{25}$$

Since the denominator is independent of r and the formulation is similar to that of MNL, the probabilities provided in Equation (25) are the easiest to interpret. In the remainder of this paper the formulations of Equations (22) and (23) will be used because in the multiplicative case only the definition of the y_r^{A}'s will change, and these can then be substituted easily.

For a route set R the combination of a generating function G and a *generating vector* $\mathbf{y} = (y_1, \ldots, y_{|R|})$ completely determines choice probabilities with Equation (22). In other words, the route set, the generating function and the generating vector completely specify a choice model. Therefore, we can write the choice probabilities as a function of the generating function and generating vector for all types of GMEV models. For a choice model X applied on route set R with μ-homogeneous generating function G^X and generating vector $\mathbf{y}^X$, we can write the probability of choosing route r as

$$P_r^X(G^X; \mathbf{y}^X) := \frac{y_r^X G_r^X(\mathbf{y}^X)}{\mu G^X(\mathbf{y}^X)}, \quad \forall r \in R. \tag{26}$$

Note that for the additive utility formula the generating vector is always equal to the exponentials of systematic utilities. Therefore define $\mathbf{y}^{\mathrm{A}} := (e^{V_1}, \ldots, e^{V_{|R|}})$ as the *additive generating vector*. This additive generating vector should be used for all MEV models with an additive utility form, thus the generating function then defines choice probabilities.

3.1.2 Multiplicative MEV models

Before we present examples of the generating function, we first examine the multiplicative utility formula. Fosgerau and Bierlaire (2009) are the first to explore this field. In their analysis, the multiplicative formulation is transformed into an equivalent additive formulation. The analysis presented here is slightly different because we allow more flexibility on the scale parameter. The models in (Castillo et al., 2008; Nakayama, 2013; Kitthamkesorn and Chen, 2013) yield similar results for specific instances of generating functions, but the utility is not purely the product of systematic utility and an error term; instead, the utility is directly parametrized.

The multiplicative utility formula is

$$U_r^{\mathrm{M}} = V_r \varepsilon_r^{\mathrm{M}}, \quad \forall r \in R, \tag{27}$$

where $V_r < 0$ and $\varepsilon_r^{\mathrm{M}} \geq 0$. This domain of the systematic utility is more restrictive than for additive MEV; however, in the route choice context it is natural that route costs are valued negatively, yielding negative

systematic utility. By applying a log-transformation to Equation (16) and some further algebra, we can derive the probability of choosing route $r \in R$ as:

$$P_r^{\mathrm{M}} = \mathrm{P}\left(V_r \varepsilon_r^{\mathrm{M}} \geq V_p \varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r \right) \tag{28}$$

$$= \mathrm{P}\left(-V_r \varepsilon_r^{\mathrm{M}} \leq -V_p \varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r \right) \tag{29}$$

$$= \mathrm{P}\left(\ln\left(-V_r \varepsilon_r^{\mathrm{M}} \right) \leq \ln\left(-V_p \varepsilon_p^{\mathrm{M}} \right), \quad \forall p \neq r \right) \tag{30}$$

$$= \mathrm{P}\left(\ln(-V_r) + \ln\left(\varepsilon_r^{\mathrm{M}} \right) \leq \ln\left(-V_p \right) + \ln\left(\varepsilon_p^{\mathrm{M}} \right), \quad \forall p \neq r \right) \tag{31}$$

$$= \mathrm{P}\left(-\ln(-V_r) - \ln\left(\varepsilon_r^{\mathrm{M}} \right) \geq -\ln\left(-V_p \right) - \ln\left(\varepsilon_p^{\mathrm{M}} \right), \quad \forall p \neq r \right) \tag{32}$$

Equation (32) has the same structure as the choice probabilities in Equation (17) of the additive choice model with systematic utilities $\tilde{V}_r = -\ln(-V_r), \forall r \in R$ and error terms $\tilde{\varepsilon}_r^{\mathrm{A}} = -\ln\left(\varepsilon_r^{\mathrm{M}} \right)$. Therefore, we can apply the theory of the additive MEV with these variable substitutions. So in the multiplicative case, $\left(-\ln \varepsilon_1^{\mathrm{M}}, \ldots, -\ln \varepsilon_{|R|}^{\mathrm{M}} \right)$ follows MEV type I distribution $\boldsymbol{\varepsilon}^{\mathrm{MEV}}$ (i.e., $(\varepsilon_1^{\mathrm{M}}, \ldots, \varepsilon_{|R|}^{\mathrm{M}})$ is not similar to $\boldsymbol{\varepsilon}^{\mathrm{MEV}}$!), this distribution of $(\varepsilon_1^{\mathrm{M}}, \ldots, \varepsilon_{|R|}^{\mathrm{M}})$ is called MEV type III. This is a multivariate distribution whose marginal distributions are of type reversed Weibull, and inherits the covariance structure from the generating function. So, it is the reversed Weibull equivalent of the model of McFadden (1978). Note that all generating functions can be applied on both types. The *multiplicative generating vector*, denoted with $\boldsymbol{y}^{\mathrm{M}}$, is derived from the additive generating vector and becomes

$$\boldsymbol{y}^{\mathrm{M}} = \left(e^{\tilde{V}_1}, \ldots, e^{\tilde{V}_{|R|}} \right) = \left(e^{-\ln(-V_1)}, \ldots, e^{-\ln(-V_{|R|})} \right) = \left(\frac{-1}{V_1}, \ldots, \frac{-1}{V_{|R|}} \right). \tag{33}$$

The probabilities of models with multiplicative utility formulas can be derived from Equation (26) by simply using the multiplicative generating vector $\boldsymbol{y}^{\mathrm{M}}$. By applying Euler's homogeneous function theorem again we can get

$$P_r = \frac{\frac{-G_r(-1/V_1, \ldots, -1/V_{|R|})}{V_r}}{\sum_{p \in R} \frac{-G_p(-1/V_1, \ldots, -1/V_{|R|})}{V_p}}, \tag{34}$$

but the form of Equation (26) using $\boldsymbol{y}^{\mathrm{M}}$ is preferred. The main advantage of the multiplicative utility formula is that the utilities are automatically heteroscedastic. The standard deviation of the utility is proportional with the systematic utility and thus even the multiplicative equivalent of MNL is heteroscedastic.

A well-known property of additive models is that the probabilities are invariant under addition of a constant to all systematic utilities. Therefore, only differences in utilities matter in the MNL model. Analogously, the multiplicative models are invariant under multiplying all systematic utilities with the same factor (mini-proof: If $\lambda V_r \varepsilon_r^{\mathrm{M}} \geq \lambda V_p \varepsilon_p^{\mathrm{M}}$ then $V_r \varepsilon_r^{\mathrm{M}} \geq V_p \varepsilon_p^{\mathrm{M}}$), and only ratios between utilities matter in the multiplicative equivalent of MNL. In Section 4 its properties are further analysed.

3.1.3 Hybrid approach

Xu et al. (2015) introduce a hybrid model directly derived from the logit and weibit choice probability functions. This model inherits characteristics of as well logit as weibit. The model uses two scale

parameters, one from logit (μ^A) and one from weibit (μ^M), and has choice probabilities equal to

$$P_r = \frac{e^{\mu^A V_r} \left(-\frac{1}{V_r}\right)^{\mu^M}}{\sum_{p \in R} e^{\mu^A V_p} \left(-\frac{1}{V_p}\right)^{\mu^M}}, \quad \forall r \in R. \tag{35}$$

Unfortunately, the model cannot be written as a RUM model with route r's utility in the form of $U_r = V_r \times \varepsilon^M + \varepsilon^A$ with a additive error term ε^A and multiplicative error term ε^M, which would have been a very neat solution. A straightforward hybrid approach in our framework would be to introduce generating vector $(-e^{V_1}/V_1, \ldots, -e^{V_{|R|}}/V_{|R|})$, i.e., the element-wise multiplication of $\boldsymbol{y}^A$ and $\boldsymbol{y}^M$. This will not lead to the approach of Xu et al. (2015), since only one scale parameter will be introduced when the vector is applied on a generating function.

However, the hybrid model of Xu et al. (2015) can be derived from either the A- or M-models from our framework by introducing an additional term in the systematic utility. In the additive form, the hybrid model of equation (35) is obtained with

$$U_r = V_r - \frac{\mu^M}{\mu^A} \ln(-x) + \varepsilon_r^A, \quad \forall r \in R, \tag{36}$$

where a logarithmic term is inserted to simulate the multiplicative model (Xu et al., 2015). In our framework, this can be represented with generating vector

$$\boldsymbol{y} = \left(e^{V_1 - \rho \ln(-V_1)}, \ldots, e^{V_{|R|} - \rho \ln(-V_{|R|})}\right) = \left(\frac{e^{V_1}}{(-V_1)^\rho}, \ldots, \frac{e^{V_{|R|}}}{(-V_{|R|})^\rho}\right), \tag{37}$$

where $\rho > 0$ is a parameter representing the ratio between the scales. In the multiplicative form, the hybrid model of equation (35) is obtained with

$$U_r = V_r \times e^{-\frac{\mu^A}{\mu^M} V_r} \times \varepsilon_r^M, \quad \forall r \in R, \tag{38}$$

where a exponential factor is inserted to simulate the additive model. In the framework again, this becomes generating vector

$$\boldsymbol{y} = \left(\frac{-1}{V_1 e^{-\frac{V_1}{\rho}}}, \ldots, \frac{-1}{V_{|R|} e^{-\frac{V_{|R|}}{\rho}}}\right) = \left(\frac{e^{\frac{V_1}{\rho}}}{-V_1}, \ldots, \frac{e^{\frac{V_{|R|}}{\rho}}}{-V_{|R|}}\right), \tag{39}$$

with the same interpretation for ρ as above. The difference between the two generating vectors lies at the estimated scale parameter. With the latter formulation μ^M is estimated directly from the model and μ^A is derived from ρ. While with equation (37) μ^A is estimated directly and μ^M is derived from ρ. There is no difference in model outcomes and/or flexibility. Using any of the two approaches, it is possible to derive hybrid path-size, paired combinatorial, and link-nested models (see next section). The covariance structure provided in the latter two model instances are then only applied on the main error structure (i.e., dependent on which of the two generating functions is chosen).

3.2   Generating functions and model instances

To create instances of the family of GMEV route choice models, the generating function has to be specified. Together with one of the two derived generating vectors, Equation (26) will provide the closed-form choice probabilities of the choice model. Any function that satisfies Equations (19)-(21) is a generating function. The cross-nested logit uses one of the most general forms of the generating function

14

and is analysed by Bierlaire (2006); Papola (2004). All presented generating functions here are also special cases of cross-nested logit models.

To address overlap on links, the presented model instances assume that each link is associated with one cost, namely systematic utility. This assumption is made for clarity reasons, differentiation into link length, travel and other cost can easily be done, also for the various factors and coefficients introduced below. Let $\mathcal{L}$ be a set of links and let the systematic utility of link $l \in \mathcal{L}$ be V_l. For each route $r \in R$, $\mathcal{L}_r$ is the set of links of which r consists and the systematic route utility equals $V_r = \sum_{l \in \mathcal{L}_r} V_l$.

3.2.1 Multinomial

The well-known MNL model (see McFadden, 1974) is based on multinomial generating function

$$G^{\mathrm{MN}}(\boldsymbol{z}) := \sum_{r \in R} z_r^{\mu}, \tag{40}$$

where $\mu > 0$ is the scale parameter. The traditional (additive) MNL model, denoted A-MN, has the following simple route choice probabilities,

$$P_r^{\mathrm{A\text{-}MN}}(G^{\mathrm{MN}}; \boldsymbol{y}^{\mathrm{A}}) = \frac{y_r^{\mathrm{A}} G_r^{\mathrm{MN}}(\boldsymbol{y}^{\mathrm{A}})}{\mu G^{\mathrm{MN}}(\boldsymbol{y}^{\mathrm{A}})} = \frac{e^{\mu V_r}}{\sum_{p \in R} e^{\mu V_p}}, \quad \forall r \in R. \tag{41}$$

The multiplicative counterpart is the multinomial weibit model (or Kirchhoff distribution), denoted with M-MN and has route choice probabilities

$$P_r^{\mathrm{M\text{-}MN}}(G^{\mathrm{MN}}; \boldsymbol{y}^{\mathrm{M}}) = \frac{y_r^{\mathrm{M}} G_r^{\mathrm{MN}}(\boldsymbol{y}^{\mathrm{M}})}{\mu G^{\mathrm{MN}}(\boldsymbol{y}^{\mathrm{M}})} = \frac{\left(-\frac{1}{V_r}\right)^{\mu}}{\sum_{p \in R} \left(-\frac{1}{V_p}\right)^{\mu}}, \quad \forall r \in R. \tag{42}$$

3.2.2 Path-Size

We will first introduce a Path-Size (PS) generating function to define a specific MEV type, and then we will show that models proposed in the literature are identical. For this PS generating function the path-size factor PS_r is required for each route $r \in R$; this factor depicts the amount of overlap with other routes. Define PS generating function

$$G^{\mathrm{PS}}(\boldsymbol{z}) := \sum_{r \in R} \mathrm{PS}_r^{\beta} z_r^{\mu}, \tag{43}$$

where $\mu > 0$ is the scale parameter and β is the path-size parameter. The additive Path-Size model, denoted A-PS, has choice probabilities

$$P_r^{\mathrm{A\text{-}PS}}(G^{\mathrm{PS}}; \boldsymbol{y}^{\mathrm{A}}) = \frac{y_r^{\mathrm{A}} G_r^{\mathrm{PS}}(\boldsymbol{y}^{\mathrm{A}})}{\mu G^{\mathrm{PS}}(\boldsymbol{y}^{\mathrm{A}})} = \frac{\mathrm{PS}_r^{\beta} e^{\mu V_r}}{\sum_{p \in R} \mathrm{PS}_p^{\beta} e^{\mu V_p}}, \quad \forall r \in R. \tag{44}$$

The path-size-like models in the literature have a logarithmic term included in the utility. The MNL model with scale μ and systematic route utility $V_r + \gamma \ln x_r$ leads to the same choice probabilities as the additive MEV model with generating function $G^{\mathrm{PS}}(\boldsymbol{z})$, generating vector $\boldsymbol{y} = (e^{V_1}, \ldots, e^{V_{|R|}})$, path-size factors $\mathrm{PS}_r = x_r, \forall r \in R$, and path-size parameter $\beta = \mu \gamma$. So, C-Logit, Path-Size Logit, and Path-Size Correction Logit can be translated into a A-PS model. C-Logit presented by Cascetta et al. (1996) adds

the term $-\beta^{\text{CF}}\text{CF}_r$ to the utility, with parameter β^{CF} and commonality factor CF_r for each route r. This is equivalent to A-PS with $\text{PS}_r = \text{CF}_r$ and $\beta = -\mu\beta^{\text{CF}}$. Path-Size Logit presented by Ben-Akiva and Bierlaire (1999) adds the term $\beta^{\text{PS}}\tilde{\text{PS}}_r$ to the utility, with parameter β^{PS} and path-size factor $\tilde{\text{PS}}_r$ for each route r. This is equivalent to A-PS with $\text{PS}_r = \tilde{\text{PS}}_r$ and $\beta = \mu\beta^{\text{PS}}$. Path-Size Correction Logit (see Bovy et al., 2008) is equal in the same way. For those familiar with general cross-nested logit: note that G^{PS} is an instance of cross-nested logit with parametrized non-normalized inclusion factors with a single nest per alternative.

Different choices for PS_r are compared in (Frejinger and Bierlaire, 2006); they conclude that the best formulation for the path-size factor is

$$\text{PS}_r = \frac{\sum_{l\in\mathscr{L}_r}\frac{V_l}{\#_l}}{\sum_{l\in\mathscr{L}_r}V_l} = \frac{\sum_{l\in\mathscr{L}_r}\frac{V_l}{\#_l}}{V_r}, \tag{45}$$

where $\#_l = |\{r\in R | l\in\mathscr{L}_r\}|$ is the number of routes using link l. If $\text{PS}_r = 1$ then r has no overlap with any other route. If $\text{PS}_r \to 0$ then r shares each link with many other routes.

The multiplicative counterpart of A-PS is denoted with M-PS and has route choice probabilities

$$P_r^{\text{M-PS}}(G^{\text{PS}};\boldsymbol{y}^{\text{M}}) = \frac{y_r^{\text{M}}G_r^{\text{PS}}(\boldsymbol{y}^{\text{M}})}{\mu G^{\text{PS}}(\boldsymbol{y}^{\text{M}})} = \frac{\text{PS}_r^{\beta}\left(-\frac{1}{V_r}\right)^{\mu}}{\sum_{p\in R}\text{PS}_p^{\beta}\left(-\frac{1}{V_p}\right)^{\mu}}, \quad \forall r\in R. \tag{46}$$

Path-size is also added to MNW by Kitthamkesorn and Chen (2013, 2014), but they assume that $\beta = 1$, which thus leads to a specific instance of M-PS.

3.2.3 Paired Combinatorial

The (additive) Paired Combinatorial Logit (PCL) model as presented and analysed by Chu (1989); Koppelman and Wen (2000); Pravinvongvuth and Chen (2005); Prashker and Bekhor (1998); Gliebe et al. (1999); Chen et al. (2003) captures correlation between each pair of routes. For each route couple a nest is defined with a fixed nest specific scale. This scale is determined by the similarity index, denoted with φ_{rp}, for all $r \neq p \in R$. Define paired combinatorial generating function

$$G^{\text{PC}}(\boldsymbol{z}) := \sum_{r\in R}\sum_{p\in R\setminus\{r\}}\left(z_r^{\frac{\mu}{1-\varphi_{rp}}} + z_p^{\frac{\mu}{1-\varphi_{rp}}}\right)^{1-\varphi_{rp}}, \tag{47}$$

where $\mu > 0$ is the scale parameter. Multiple definitions for the similarity index exist, but the most common one is

$$\varphi_{rp} := \frac{\sum_{l\in\mathscr{L}_r\cap\mathscr{L}_p}V_l}{\sqrt{V_r V_p}}, \quad \forall r,p\in R. \tag{48}$$

The traditional additive PCL model, denoted A-PC, has route choice probabilities

$$P_r^{\text{A-PC}}(G^{\text{PC}};\boldsymbol{y}^{\text{A}}) = \frac{y_r^{\text{A}}G_r^{\text{PC}}(\boldsymbol{y}^{\text{A}})}{\mu G^{\text{PC}}(\boldsymbol{y}^{\text{A}})} = \frac{\sum_{p\in R\setminus\{r\}}e^{\frac{\mu V_r}{1-\varphi_{rp}}}\left(e^{\frac{\mu V_r}{1-\varphi_{rp}}} + e^{\frac{\mu V_p}{1-\varphi_{rp}}}\right)^{-\varphi_{rp}}}{\sum_{r'\in R}\sum_{p\in R\setminus\{r'\}}\left(e^{\frac{\mu V_{r'}}{1-\varphi_{r'p}}} + e^{\frac{\mu V_p}{1-\varphi_{r'p}}}\right)^{1-\varphi_{r'p}}}, \quad \forall r\in R. \tag{49}$$

16

The multiplicative counterpart is denoted with M-PC and has route choice probabilities

$$P_r^{\text{M-PC}}(G^{\text{PC}}; \boldsymbol{y}^{\text{M}}) = \frac{y_r^{\text{M}} G_r^{\text{PC}}(\boldsymbol{y}^{\text{M}})}{\mu G^{\text{PC}}(\boldsymbol{y}^{\text{M}})} = \frac{\sum_{p \in R \setminus \{r\}} \left(-\frac{1}{V_r}\right)^{\frac{\mu}{1-\varphi_{rp}}} \left(\left(-\frac{1}{V_r}\right)^{\frac{\mu}{1-\varphi_{rp}}} + \left(-\frac{1}{V_p}\right)^{\frac{\mu}{1-\varphi_{rp}}}\right)^{-\varphi_{rp}}}{\sum_{r' \in R} \sum_{p \in R \setminus \{r'\}} \left(\left(-\frac{1}{V_{r'}}\right)^{\frac{\mu}{1-\varphi_{r'p}}} + \left(-\frac{1}{V_p}\right)^{\frac{\mu}{1-\varphi_{r'p}}}\right)^{1-\varphi_{r'p}}}, \quad \forall r \in R. \quad (50)$$

3.2.4 Link-Nested

[Vovsha and Bekhor](#) ([1998](#)) are the first who applied the Cross-Nested Logit model to route choice such that links represent nests. In this (additive) Link-Nested Logit (LNL) a nest is created for each link and all routes that use the link are included in the nest. The inclusion coefficient of each route in each link (or nest) is denoted with α_{lr} for all links $l \in \mathscr{L}$ and routes $r \in R$. Define link-nested generating function

$$G^{\text{LN}}(\boldsymbol{z}) := \sum_{l \in \mathscr{L}} \left(\sum_{r \in R} \alpha_{lr} z_r^{\mu_l}\right)^{\frac{\mu}{\mu_l}}, \quad (51)$$

where $\mu > 0$ is the scale parameter and $\mu_l > 0, \forall l \in \mathscr{L}$ are the link specific scale parameters. In general, it is difficult to estimate all link specific scale parameters in large networks. More on this topic can be found in ([Bierlaire](#), [2006](#)), where also the necessity of normalizing the inclusion coefficients is discussed. For this study the following normalized inclusion coefficients are applied

$$\alpha_{lr} = \begin{cases} \frac{V_l}{V_r} & \text{if } l \in \mathscr{L}_r \\ 0 & \text{otherwise} \end{cases}. \quad (52)$$

The traditional additive LNL model, denoted A-LN, has route choice probabilities

$$P_r^{\text{A-LN}}(G^{\text{LN}}; \boldsymbol{y}^{\text{A}}) = \frac{y_r^{\text{A}} G_r^{\text{LN}}(\boldsymbol{y}^{\text{A}})}{\mu G^{\text{LN}}(\boldsymbol{y}^{\text{A}})} = \frac{\sum_{l \in \mathscr{L}} \alpha_{lr} e^{\mu_l V_r} \left(\sum_{p \in R} \alpha_{lp} e^{\mu_l V_p}\right)^{\frac{\mu}{\mu_l}-1}}{\sum_{l \in \mathscr{L}} \left(\sum_{p \in R} \alpha_{lp} e^{\mu_l V_p}\right)^{\frac{\mu}{\mu_l}}}, \quad \forall r \in R. \quad (53)$$

The multiplicative counterpart is M-LN with route choice probabilities

$$P_r^{\text{M-LN}}(G^{\text{LN}}; \boldsymbol{y}^{\text{M}}) = \frac{y_r^{\text{M}} G_r^{\text{LN}}(\boldsymbol{y}^{\text{M}})}{\mu G^{\text{LN}}(\boldsymbol{y}^{\text{M}})} = \frac{\sum_{l \in \mathscr{L}} \alpha_{lr} \left(-\frac{1}{V_r}\right)^{\mu_l} \left(\sum_{p \in R} \alpha_{lp} \left(-\frac{1}{V_p}\right)^{\mu_l}\right)^{\frac{\mu}{\mu_l}-1}}{\sum_{l \in \mathscr{L}} \left(\sum_{p \in R} \alpha_{lp} \left(-\frac{1}{V_p}\right)^{\mu_l}\right)^{\frac{\mu}{\mu_l}}}, \quad \forall r \in R. \quad (54)$$

3.2.5 Joint Network

Recently, [Papola and Marzano](#) ([2013](#)) introduced a new generating function based on an generalization of cross-nested generating functions to networks (see [Daly and Bierlaire](#), [2006](#); [Newman](#), [2008](#)). Because their approach is based on links instead of routes, the generating function defines dependencies between links instead of routes. This is fundamentally different from the GMEV models presented in this work that is based on a route set and based on the error structure of random utility per route. The joint network formulation of [Papola and Marzano](#) ([2013](#)) is not a member of the GMEV models for route choice in this study since it cannot handle generic route sets as input. Their method has an implicit enumeration of routes. This has the advantage that the problem of choice set generation is dealt

with internally. Unfortunately, this does not allow a utility formulation per route. Furthermore, the link based approach also delimits the routes that will be found. For their model specifically, scale parameters of 'nodes' depend on the shortest path to the destination. Their definition requires that a route advances through nodes for which the shortest path towards the destination decreases in every step. In real networks, the remaining shortest path will often increase when you deviate from the shortest path. This makes the model restrictive. Although a natural connection between a road network and a (recursive) network based generating function seems to be promising, Marzano (2014) suggests that the achievable covariance structure is not more advanced than that of cross-nested generating functions.

4 Qualitative Assessment of the Models

This section analyses the utility distribution of the GMEV models and assesses them qualitatively based on the desired properties presented in Section 2.3. The desired route utilities have two random variables (i.e., the random foreseen travel time and the analyst error) per route, but the models contain only one (GMEV-distributed) variable per route. We show that this variable can only resemble the analyst error in the additive models, and that it can resemble both errors in the multiplicative models, however, with completely correlated foreseen travel times and analyst errors.

4.1 Utility distribution

The results in this section for the additive models summarize the findings of McFadden (1978); Daly and Bierlaire (2006), while the derivations for the multiplicative models are slightly different from those in (Fosgerau and Bierlaire, 2009) that uses two scales. From the additive utility formulation and Equation (18) follows the multivariate utility distribution of the additive models:

$$F_{U^A}(x_1, \ldots, x_{|R|}) = \Pr\left(U_r^A \le x_r, \quad \forall r \in R\right) \tag{55}$$

$$= \Pr\left(V_r + \varepsilon_r^A \le x_r, \quad \forall r \in R\right) \tag{56}$$

$$= \Pr\left(\varepsilon_r^A \le x_r - V_r, \quad \forall r \in R\right) \tag{57}$$

$$= F_{\boldsymbol{\varepsilon}^{MEV}}\left(x_1 - V_1, \ldots, x_{|R|} - V_{|R|}\right) \tag{58}$$

$$= e^{-G(e^{-x_1 + V_1}, \ldots, e^{-x_{|R|} + V_{|R|}})} \tag{59}$$

18

With some algebra, the multivariate utility distribution of the multiplicative models is written as

$$F_{U^M}(x_1,\ldots,x_{|R|}) = \Pr\left(U_r^M \le x_r, \quad \forall r \in R\right) \tag{60}$$

$$= \Pr\left(V_r \varepsilon_r^M \le x_r, \quad \forall r \in R\right) \tag{61}$$

$$= \Pr\left(-V_r \varepsilon_r^M \ge -x_r, \quad \forall r \in R\right) \tag{62}$$

$$= \Pr\left(\ln\left(-V_r \varepsilon_r^M\right) \ge \ln(-x_r), \quad \forall r \in R\right) \tag{63}$$

$$= \Pr\left(\ln(-V_r) + \ln\left(\varepsilon_r^M\right) \ge \ln(-x_r), \quad \forall r \in R\right) \tag{64}$$

$$= \Pr\left(-\ln\left(\varepsilon_r^M\right) \le \ln(-V_r) - \ln(-x_r), \quad \forall r \in R\right) \tag{65}$$

$$= F_{\boldsymbol{\varepsilon}^{MEV}}\left(\ln(-V_1) - \ln(-x_1),\ldots,\ln\left(-V_{|R|}\right) - \ln\left(-x_{|R|}\right)\right) \tag{66}$$

$$= F_{\boldsymbol{\varepsilon}^{MEV}}\left(\ln\frac{V_1}{x_1},\ldots,\ln\frac{V_{|R|}}{x_{|R|}}\right) \tag{67}$$

$$= e^{-G\left(e^{-\ln\frac{V_1}{x_1}},\ldots,e^{-\ln\frac{V_{|R|}}{x_{|R|}}}\right)} \tag{68}$$

$$= e^{-G\left(\frac{x_1}{V_1},\ldots,\frac{x_{|R|}}{V_{|R|}}\right)} \tag{69}$$

Then the marginal distribution for the utility of route $r \in R$ is

$$F_{U_r^A}(x_r) = \lim_{\{x_p \to \infty\}_{p \ne r}} F_{U^A}\left(x_1,\ldots,x_{|R|}\right) \tag{70}$$

$$= e^{-G\left(e^{-x_r + V_r}\mathbf{1}_r\right)} \tag{71}$$

$$= e^{-e^{\mu(-x_r + V_r)}G(\mathbf{1}_r)} \tag{72}$$

$$= e^{-e^{\mu(-x_r + V_r) + \ln G(\mathbf{1}_r)}} \tag{73}$$

$$= e^{-e^{\mu\left(-x_r + V_r + \frac{\ln G(\mathbf{1}_r)}{\mu}\right)}}, \tag{74}$$

for the additive case, which can be identified as a Gumbel distribution, and with notation $\mathbf{1}_r := (0,\ldots,0,1,0,\ldots,0)$ (1 on the r-th position).[7] For the multiplicative models the marginal distribution function of route r's utility is

$$F_{U_r^M}(x_r) = \lim_{\{x_p \to 0\}_{p \ne r}} F_{U^M}\left(x_1,\ldots,x_{|R|}\right) \tag{75}$$

$$= e^{-G\left(\frac{x_r}{V_r}\mathbf{1}_r\right)} \tag{76}$$

$$= e^{-\left(\frac{x_r}{V_r}\right)^{\mu}G(\mathbf{1}_r)} \tag{77}$$

$$= e^{-\left(\frac{x_r}{V_r G(\mathbf{1}_r)^{\frac{1}{\mu}}}\right)^{\mu}}, \tag{78}$$

which can be identified as a reversed Weibull distribution, where reversed means that $-U_r^M$ is Weibull distributed .

The expected maximum utility is an important measure of choice models; it can be used to formulate the corresponding user equilibrium. Furthermore, it is input to define duality gaps. Denote U^{*A}

---

[7] Note that for $\mu = 1$ these are reversed exponential distributions.

Table 1: Expected value and variance of the desired, multiplicative and additive utility formulations

	Desired	Additive MEV (γ is Euler's constant)		Multiplicative MEV ($\Gamma(\cdot)$ is the Gamma function)
Utility U_r	$V_r^0 + \beta\tau_r + \varepsilon_r$	$V_r + \varepsilon_r^{\mathrm{A}} = V_r^0 + \beta\hat{\tau}_r + \varepsilon_r^{\mathrm{A}}$		$V_r \times \varepsilon_r^{\mathrm{M}} = \left(V_r^0 + \beta\hat{\tau}_r\right) \times \varepsilon_r^{\mathrm{M}}$
Expected value	$\underbrace{V_r^0 + \beta\hat{\tau}_r + \mathbb{E}(\varepsilon_r)}_{\text{affine in } \hat{\tau}_r}$	$\underbrace{V_r^0 + \beta\hat{\tau}_r + \dfrac{\ln G(\mathbf{1}_r) + \gamma}{\mu}}_{\text{affine in } \hat{\tau}_r}$		$\underbrace{\dfrac{V_r^0 + \beta\hat{\tau}_r}{G(\mathbf{1}_r)^{1/\mu}}\Gamma\left(1 + \dfrac{1}{\mu}\right)}_{\text{affine in } \hat{\tau}_r}$
Variance	$\underbrace{\theta^2\hat{\tau}_r^2 + \mathrm{Var}(\varepsilon_r)}_{\text{quadratic in } \hat{\tau}_r}$	$\underbrace{\dfrac{\pi^2}{6}\left(\dfrac{1}{\mu}\right)^2}_{\text{constant}}$		$\underbrace{\dfrac{\left(V_r^0 + \beta\hat{\tau}_r\right)^2}{G(\mathbf{1}_r)^{2/\mu}}\left(\Gamma\left(1 + \dfrac{2}{\mu}\right) - \Gamma\left(1 + \dfrac{1}{\mu}\right)^2\right)}_{\text{quadratic in } V_r^0 + \beta\hat{\tau}_r}$
Expected $\max_{r \in R} U_r$	Not closed form	$\dfrac{\ln G\left(e^{V_1}, \ldots, e^{V_{|R|}}\right) + \gamma}{\mu}$		$-G\left(\dfrac{-1}{V_1}, \ldots, \dfrac{-1}{V_{|R|}}\right)^{\frac{-1}{\mu}}\Gamma\left(1 + \dfrac{1}{\mu}\right)$

as the maximum utility for the additive models, and $U^{*\mathrm{M}}$ as the maximum utility for the multiplicative models. Their distributions can be written as

$$F_{U^{*\mathrm{A}}}(x) = F_{U^{\mathrm{A}}}(x, \ldots, x) \tag{79}$$

$$= e^{-G\left(e^{-x+V_1}, \ldots, e^{-x+V_{|R|}}\right)} \tag{80}$$

$$= e^{-e^{-\mu x}G\left(e^{V_1}, \ldots, e^{V_{|R|}}\right)} \tag{81}$$

$$= e^{-e^{\mu\left(-x + \ln G\left(e^{V_1}, \ldots, e^{V_{|R|}}\right)/\mu\right)}}, \text{ and} \tag{82}$$

$$F_{U^{*\mathrm{M}}}(x) = F_{U^{\mathrm{M}}}(x, \ldots, x) \tag{83}$$

$$= e^{-G\left(\frac{x}{V_1}, \ldots, \frac{x}{V_{|R|}}\right)} \tag{84}$$

$$= e^{-(-x)^\mu G\left(\frac{-1}{V_1}, \ldots, \frac{-1}{V_{|R|}}\right)} \tag{85}$$

$$= e^{-\left(\frac{x}{-G\left(\frac{-1}{V_1}, \ldots, \frac{-1}{V_{|R|}}\right)^{\frac{-1}{\mu}}}\right)^\mu}, \tag{86}$$

which can be identified as Gumbel and reversed Weibull distributions again.

Table 1 shows the desired expected value, variance, and expected maximum utility together with the actual expected value, variance and expected maximum utility of the models. It is based on the identification of the marginal distributions as Gumbel and reversed Weibull, and on systematic utility specification $V_r = V_r^0 + \beta\tau_r$. Neither the additive nor the multiplicative formulation completely coincide with the desired result.

The expected value is not a problem for any model; any value can be achieved after normalization and identification, and the location of utility is not decisive for choice probabilities. On the other hand, the variance of the additive models is constant and the variance of the multiplicative models is affected by V_r^0, and for the latter $\mathrm{Var}(\varepsilon_r)$ is not directly represented. However, as the next section describes, the constant in the systematic utility of the multiplicative model does not have to be normalized. This constant will return as a constant term in the variance, and can thus resemble $\mathrm{Var}(\varepsilon_r)$. Despite that the

multiplicative models' variances can capture the variances from the foreseen travel time distribution and analyst error simultaneously, they are both represented by ε_r and thus completely dependent. This is not in line with the desired independence.

Closed form formulations of the covariance of GMEV models are not known. Marzano et al. (2013); Marzano (2014) present a more tractable expression and a method to calculate them for additive MEV. The covariances of the additive models are based on the constant variance, and will therefore not change when the systemic utility changes. However, the systematic utility enters the variance quadratically in the multiplicative model, and therefore its covariance will also in- and decrease together with it. Thus, the multiplicative models capture the covariances better.

The expected maximum utility for the A-MN model indeed return in the well-known log-sum that can by used for an hierarchical derivation of the other models. This was first identified by Ben-Akiva (1973).

4.2 Stochastic user equilibrium formulation

The derived GMEV route choice models all have an equivalent stochastic user equilibrium formulation. Mathematical programming formulation are known for all logit based models (Fisk, 1980; Bekhor and Prashker, 1999), for the M-MN and (simplified) M-PS models (Kitthamkesorn and Chen, 2013, 2014), for the q-generalized logit model (Chikaraishi and Nakayama, 2015), and for the hybrid logit-weibit model (Xu et al., 2015). It is straightforward to derive the variational inequality formulation – that we discuss – from a mathematical programming formulation. Chen (1999); Guo et al. (2010) describe a stochastic user equilibrium for the MNL model using a variational inequality. Zhou et al. (2012) provide two variational inequality formulations for the C-logit model. Their first formulation is specific for C-logit and a special case of the formulation we present below. Their second formulation is more generic and works with any choice probability formula, but it returns different generalised costs than the approach we present below.

Since the choice probabilities of each GMEV model can be written solely in terms of the generating function and generating vector, this also holds for the corresponding variational inequality formulation. Denote the demand with D, the flow for route $r \in R$ as f_r, and $\boldsymbol{f}$ as the vector of flows. Since, amongst other attributes, the travel time in the systematic utility depends on the flow, write that the generating vector now depends on $\boldsymbol{y}(\boldsymbol{f})$[8]. For clarity, but without loss of generality, only one route set (i.e., one OD pair) is considered.

Consider the following variational inequality (VI) formulation; find equilibrium flow $\boldsymbol{f}^* = (f_1^*, \ldots, f_{|R|}^*)$, such that

$$\sum_{r \in R} \left(-\ln\left(y_r(\boldsymbol{f}^*) G_r(\boldsymbol{y}(\boldsymbol{f}^*)) \right) + \ln(f_r^*) \right)(f_r - f_r^*) \geq 0, \quad \forall \boldsymbol{f} \in \Omega,$$

$$\text{where } \Omega = \left\{ \boldsymbol{f} \in \mathbb{R}^{|R|} \middle| f_r > 0, \sum_{r \in R} f_r = D \right\}. \tag{87}$$

Define the generalized stochastic cost for route $r \in R$ under flow $\boldsymbol{f}$ as $c_r(\boldsymbol{f}) := -\ln\left(y_r(\boldsymbol{f}) G_r(\boldsymbol{y}(\boldsymbol{f})) \right) + \ln(f_r)$[9]. The corresponding Karush-Kuhn-Tucker system (see e.g., Facchinei and Pang, 2003, Proposi-

---

[8]When the generating vector is substituted, just replace V_r with $V_r(\boldsymbol{f})$ to denote dependency of attributes on flow.

[9]Note that this cost is unit-less. For MNL, it is possible to normalize it such that the systemic utility appears as a term.

tion 1.2.1) is then to find multipliers κ and $\lambda_1, \ldots, \lambda_{|R|}$ for which

$$0 = -\ln\left(y_r(\boldsymbol{f}^*)G_r(\boldsymbol{y}(\boldsymbol{f}^*))\right) + \ln(f_r^*) + \kappa - \lambda_r, \quad \forall r \in R, \tag{88}$$

$$\sum_{r \in R} f_r = D, \tag{89}$$

$$0 \le \lambda_r, \quad \forall r \in R, \tag{90}$$

$$0 < f_r, \quad \forall r \in R, \text{ and} \tag{91}$$

$$0 = \sum_{r \in R} \lambda_r f_r \tag{92}$$

hold. Equations (90), (91) and (92) imply that $\lambda_r = 0$ for all $r \in \mathbb{R}$. Rewriting equation (88) then gives

$$\ln\left(y_r(\boldsymbol{f}^*)G_r(\boldsymbol{y}(\boldsymbol{f}^*))\right) = \ln(f_r^*) + \kappa, \quad \forall r \in R \tag{93}$$

$$y_r(\boldsymbol{f}^*)G_r(\boldsymbol{y}(\boldsymbol{f}^*)) = f_r^* e^\kappa, \quad \forall r \in R. \tag{94}$$

Summing over routes gives

$$\sum_{r \in R} y_r(\boldsymbol{f}^*)G_r(\boldsymbol{y}(\boldsymbol{f}^*)) = D e^\kappa, \quad \forall r \in R. \tag{95}$$

Finally, dividing equation (94) by equation (95) gives

$$\frac{y_r(\boldsymbol{f}^*)G_r(\boldsymbol{y}(\boldsymbol{f}^*))}{\sum_{s \in R} y_s(\boldsymbol{f}^*)G_s(\boldsymbol{y}(\boldsymbol{f}^*))} = \frac{f_r^*}{D}, \tag{96}$$

which actually coincides with the choice probability definition of Equation (23). Therefore, VI problem (87) describes the stochastic user equilibrium with the choice model that is defined by generating function G and generating vector $\boldsymbol{y}$.

Such a VI formulation is useful, since the corresponding theory can be used to analyse existence and uniqueness of stochastic user equilibria (as for example done by Nagurney (1998); Bliemer et al. (2014)). In addition, the VI formulation provided the generalized stochastic cost per route – which should all be equal in equilibrium –, that can be used in assignment algorithms to determine duality gaps. Namely, as $\hat{\boldsymbol{f}} \to \boldsymbol{f}^*$, then

$$\frac{\sum_{r \in R} \hat{f}_r \left(c_r(\hat{\boldsymbol{f}}) - \min_{s \in R} c_s(\hat{\boldsymbol{f}})\right)}{\sum_{r \in R} \hat{f}_r \min_{s \in R} c_s(\hat{\boldsymbol{f}})} \to 0. \tag{97}$$

4.3 Normalization, identification, and invariance

The constant in the systematic utility (i.e., the alternative specific constant) of the logit-based (i.e., additive) models has to be normalized because only differences in utility matter. Furthermore, one of the attribute parameters or the scale has to be normalized due to identification. If the cost parameter is normalized to one, all other attribute parameters can be interpreted as willingness to pay for its attribute.

The multiplicative models however, do not require the constant in the utility to be normalized. This constant is also multiplied with the error term and thus this term is not equal amongst alternatives. This allows the modeller to use an additional parameter. Similar as for the additive case, one of the attribute parameters or the scale has to be normalized due to identification.

For logit based additive models only differences between utilities matter for the choice probabilities. Similarly, only the ratios between utilities matter for the choice probabilities in the multiplicative models (Fosgerau and Bierlaire, 2009). This means that the additive models are invariant under addition with a constant and the multiplicative models are invariant under multiplication with a constant. As the next example shows, it is doubtful that these properties are realistic and they limit both models. For the additive case consider two route sets that only differ by a constant, one with route costs (e.g., travel time) $\{1,6\}$ and one with $\{100,105\}$; it is not expected that the choice probabilities are the same for these routes sets. Analogously, route costs $\{2,3\}$ and $\{40,60\}$ should neither reflect the same probabilities for the multiplicative case. Nevertheless, since the constant in the systematic utility does not have to be normalized, the property is less restrictive for the multiplicative case. Section 6 provides further analysis.

5 Multiplicative MEV Models with Explicit Removal of Overlap

We present an adjustment to the multiplicative model based on the decision rule that travellers only compare the non-overlapping part of routes compared to a reference route. This relaxes the invariance properties to a certain extent, and implies a different choice mechanism for travellers. Probabilities for switching to another route and staying at the reference route are determined. This different decision rule can be retrieved from an alternative utility formulation, which allows a qualitative analysis of the model.

Using the systematic utility per link, and the multiplicative utility formula, let

$$U_r^{\mathrm{M}} = \left(\sum_{l \in \mathscr{L}_r} V_l \right) \varepsilon_r^{\mathrm{M}}, \quad \forall r \in R.$$

The choice probability for alternative $r \in R$ is

$$P_r^{\mathrm{M}} = \mathrm{P}(U_r^{\mathrm{M}} \geq U_p^{\mathrm{M}}, \forall p \neq r) \tag{98}$$

$$= \mathrm{P}\left(\left(\sum_{l \in \mathscr{L}_r} V_l \right) \varepsilon_r^{\mathrm{M}} \geq \left(\sum_{l \in \mathscr{L}_p} V_l \right) \varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r \right) \tag{99}$$

$$= \mathrm{P}\left(\sum_{l \in \mathscr{L}_r} V_l \varepsilon_r^{\mathrm{M}} \geq \sum_{i \in \mathscr{L}_p} V_l \varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r \right) \tag{100}$$

$$= \mathrm{P}\left(\sum_{l \in \mathscr{L}_r \setminus \mathscr{L}_p} V_l \varepsilon_r^{\mathrm{M}} + \sum_{l \in \mathscr{L}_r \cap \mathscr{L}_p} V_l \varepsilon_r^{\mathrm{M}} \geq \sum_{l \in \mathscr{L}_p \setminus \mathscr{L}_r} V_l \varepsilon_p^{\mathrm{M}} + \sum_{l \in \mathscr{L}_r \cap \mathscr{L}_p} V_l \varepsilon_p^{\mathrm{M}}, \quad \forall p \neq r \right). \tag{101}$$

In this formulation, each part of random utility is associated with a part of systematic utility according to the scaling postulate. The systematic utility that is shared amongst a pair of alternatives appears at both sides of the inequality. The shared links for route pair r, p are contained in $\mathscr{L}_r \cap \mathscr{L}_p$. An individual will evaluate the links independent of the alternatives they belong to. Therefore we assume that the part of the random utility belonging to these shared links is fully correlated. This means that for all $l \in \mathscr{L}_r \cap \mathscr{L}_p$ the terms $V_l \varepsilon_r^{\mathrm{M}}$ and $V_l \varepsilon_p^{\mathrm{M}}$ in Equation (101) are equal and can be subtracted from both sides of the inequality.

23

This can be derived in an econometrical sound fashion by reconsidering the utility formula based on a reference route. Let r be the reference route, and denote the utility for each route $p \in R$ as

$$U_p^{\mathrm{M}\Delta,r} := \begin{cases} \sum_{l \in \mathscr{L}_r} V_l \varepsilon_r^{\mathrm{M}} & \text{if } p = r \\ \sum_{l \in \mathscr{L}_p \setminus \mathscr{L}_r} V_l \varepsilon_p^{\mathrm{M}} + \sum_{l \in \mathscr{L}_r \cap \mathscr{L}_p} V_l \varepsilon_r^{\mathrm{M}} & \text{otherwise.} \end{cases} \tag{102}$$

The utility that overlaps with the reference route, is multiplied with the error term of the reference route, while the remainder has its own error term. We refer to this type of probabilities as the *MΔ-case* with reference route r. The probability of choosing route $p \in R$, conditional on reference route r is

$$P_p^{\mathrm{M}\Delta,r} = \begin{cases} \mathrm{P}\left(\sum_{l \in \mathscr{L}_{p \setminus s}} V_l \varepsilon_p^{\mathrm{M}} \geq \sum_{l \in \mathscr{L}_{s \setminus p}} V_l \varepsilon_s^{\mathrm{M}}, \quad \forall s \neq p\right) & \text{if } p = r \\ \mathrm{P}\left(\sum_{l \in \mathscr{L}_{p \setminus r}} V_l \varepsilon_p^{\mathrm{M}} + \sum_{l \in \mathscr{L}_r \cap \mathscr{L}_p} V_l \varepsilon_r^{\mathrm{M}} \geq U_s^{\mathrm{M}\Delta,r}, \quad \forall s \neq p\right) & \text{otherwise,} \end{cases} \tag{103}$$

with notation $\mathscr{L}_{r \setminus p} = \mathscr{L}_r \setminus \mathscr{L}_p$, and the overlapping parts are subtracted for case $p = r$ (see Equation (101)). First, we analyse the case $p = r$. That choice probability reads that the utility of non-overlapping links of routes s with p is smaller than the utility of the non-overlapping links of route p with each respective s. Thus, it is the probability that the traveller does not benefit from swapping some links and change route. Now isolate ε_r on the left hand side;

$$\text{if } p = r, \text{ then } P_p^{\mathrm{M}\Delta,r} = \mathrm{P}\left(\left(\sum_{l \in \mathscr{L}_{p \setminus s}} V_l\right)\varepsilon_p^{\mathrm{M}} \geq \left(\sum_{l \in \mathscr{L}_{s \setminus p}} V_l\right)\varepsilon_s^{\mathrm{M}}, \quad \forall s \neq p\right) \tag{104}$$

$$= \mathrm{P}\left(\varepsilon_p^{\mathrm{M}} \leq \frac{\sum_{l \in \mathscr{L}_{s \setminus p}} V_l}{\sum_{l \in \mathscr{L}_{p \setminus s}} V_l} \varepsilon_s^{\mathrm{M}}, \quad \forall s \neq p\right). \tag{105}$$

With the same algebra that derives Equation (32) from Equation (28), this multiplicative formulation can be transformed to an additive formulation:

$$\text{if } p = r, \text{ then } P_p^{\mathrm{M}\Delta,r} = \mathrm{P}\left(\varepsilon_p^{\mathrm{M}} \leq \frac{\sum_{l \in \mathscr{L}_{s \setminus p}} V_l}{\sum_{l \in \mathscr{L}_{p \setminus s}} V_l} \varepsilon_s^{\mathrm{M}}, \quad \forall s \neq p\right) \tag{106}$$

$$= \mathrm{P}\left(\ln(\varepsilon_p^{\mathrm{M}}) \leq \ln\left(\frac{\sum_{l \in \mathscr{L}_{s \setminus p}} V_l}{\sum_{l \in \mathscr{L}_{p \setminus s}} V_l} \varepsilon_s^{\mathrm{M}}\right), \quad \forall s \neq p\right) \tag{107}$$

$$= \mathrm{P}\left(\ln(\varepsilon_p^{\mathrm{M}}) \leq \ln\left(\frac{\sum_{l \in \mathscr{L}_{s \setminus p}} V_l}{\sum_{l \in \mathscr{L}_{p \setminus s}} V_l}\right) + \ln\left(\varepsilon_s^{\mathrm{M}}\right), \quad \forall s \neq p\right) \tag{108}$$

$$= \mathrm{P}\left(-\ln(\varepsilon_p^{\mathrm{M}}) \geq -\ln\left(\frac{\sum_{l \in \mathscr{L}_{s \setminus p}} V_l}{\sum_{l \in \mathscr{L}_{p \setminus s}} V_l}\right) - \ln\left(\varepsilon_s^{\mathrm{M}}\right), \quad \forall s \neq p\right) \tag{109}$$

For the case $p \neq r$ we can show that;

$$\text{if } p \neq r \text{ then } P_p^{\mathrm{M}\Delta,r} = \mathrm{P}\left(\sum_{l \in \mathscr{L}_{p \setminus r}} V_l \varepsilon_p^{\mathrm{M}} + \sum_{l \in \mathscr{L}_r \cap \mathscr{L}_p} V_l \varepsilon_r^{\mathrm{M}} \geq U_s^{\mathrm{M}\Delta,r}, \quad \forall s \neq p\right) \tag{110}$$

$$= \mathrm{P}\left(\frac{\sum_{l \in \mathscr{L}_{p \setminus r}} V_l}{\sum_{l \in \mathscr{L}_{r \setminus p}} V_l} \varepsilon_p^{\mathrm{M}} \leq \varepsilon_r^{\mathrm{M}} \wedge \left(\frac{\sum_{l \in \mathscr{L}_{p \setminus s}} V_l}{\sum_{l \in \mathscr{L}_{s \setminus p}} V_l} \varepsilon_p^{\mathrm{M}} \leq \frac{\sum_{l \in \mathscr{L}_{s \setminus p}} V_l}{\sum_{l \in \mathscr{L}_{p \setminus s}} V_l} \varepsilon_s^{\mathrm{M}}, \forall s \neq r, p\right)\right) \tag{111}$$

24

This equality is not trivial, one should check that Equation (111) holds or fails for all six orderings of U_r, U_p, and U_s by using and/or substituting the condition in Equation (106). So, if route p is chosen under reference route r, then it is beneficial to switch from route r to route p, and this improvement is larger than to switch to any other route s. Equation (111) can be rewritten in an additive form similar to (109) by taking the logarithm transformation. Assume – similar as in the M-models – that in the derived additive forms $\left(-\ln(\varepsilon_1^{\mathrm{M}}), \dots, -\ln(\varepsilon_{|R|}^{\mathrm{M}})\right)$ follows MEV distribution $\boldsymbol{\varepsilon}^{\mathrm{MEV}}$, then we have another type of GMEV route choice models. The generating vector is reference route specific, denoted with $\boldsymbol{y}^{\mathrm{M}\Delta,r}$, and given by

$$y_p^{\mathrm{M}\Delta,r} = \begin{cases} 1 & \text{if } r = p \\ \frac{\sum_{l \in \mathscr{L}_{r\backslash p}} V_l}{\sum_{l \in \mathscr{L}_{p\backslash r}} V_l} & \text{otherwise} \end{cases}, \quad \forall r, p \in R, \tag{112}$$

where we used that $e^0 = 1$ and

$$e^{-\ln\left(\frac{\sum_{l \in \mathscr{L}_{p\backslash r}} V_l}{\sum_{l \in \mathscr{L}_{r\backslash p}} V_l}\right)} = \frac{\sum_{l \in \mathscr{L}_{r\backslash p}} V_l}{\sum_{l \in \mathscr{L}_{p\backslash r}} V_l}.$$

Given any generating function G, reference route r and $M\Delta$-*generating vector* $\boldsymbol{y}^{\mathrm{M}\Delta,r}$ the choice probabilities can be derived with Equation (26). By then applying Euler's homogeneous function theorem, the probability of choosing route $p \in R$ with reference route $r \in R$ is:

$$P_p^{\mathrm{M}\Delta,r} = \frac{y_p^{\mathrm{M}\Delta,r} G_p(\boldsymbol{y}^{\mathrm{M}\Delta,r})}{G_r(\boldsymbol{y}^{\mathrm{M}\Delta,r}) + \sum_{\{p \in R | p \neq r\}} \frac{\sum_{l \in \mathscr{L}_{p\backslash r}} V_l}{\sum_{l \in \mathscr{L}_{r\backslash p}} V_l} G_p(\boldsymbol{y}^{\mathrm{M}\Delta,r})}. \tag{113}$$

Note that if no overlap exists the MΔ-models collapse to the M-models (i.e., they return the same probabilities).

The previous analysis was based on one reference route, but for applications there is not always a (single) reference route available since the (current) reference of the travellers is not known. The final step is to handle this uncertainty. Denote $P^{\mathrm{ref}}(r)$ as the probability that route r is the reference route. The final choice probability for route p then becomes

$$P_p^{\mathrm{M}\Delta} = \sum_{r \in R} P_p^{\mathrm{M}\Delta,r} P^{\mathrm{ref}}(r). \tag{114}$$

We consider three ways to determine the $P^{\mathrm{ref}}(r)$. First, it is possible to assign every route as reference route with the same probability (i.e., $P^{\mathrm{ref}}(r) = 1/|R|, \forall r \in R$). This seems only realistic when no irrelevant routes exist in the choice set, since it is not plausible that travellers use an irrelevant route as reference. Second, it is very natural to set the probability that a route is chosen equal to the probability that a route is the reference route. Then Equation (114) becomes a system of equations: $P_p^{\mathrm{M}\Delta} = \sum_{r \in R} P_p^{\mathrm{M}\Delta,r} P_p^{\mathrm{M}\Delta}$ (where $P_p^{\mathrm{M}\Delta}, p \in R$ are the unknowns). Furthermore, $\sum_{p \in R} P_p^{\mathrm{M}\Delta} = 1$, thus this system of equations can be identified as a Markov chain. Since $P_p^{\mathrm{M}\Delta,r} > 0, \forall p, r \in R$, a steady state exists which can be found by solving the system of equations. A third possibility is to fix one route (e.g., the fastest in free-flow conditions) as the reference route r, and to only determine the choice probabilities of MΔ, r. This means that $P^{\mathrm{ref}}(r) = 1$ for exactly one $r \in R$. This might not be the realistic for standard equilibrium models since this creates asymmetry. However, such an approach would be very feasible for en-route decisions, where the current route serves as reference point. Also in day-to-day models, the previously chosen trip can be a natural reference route.

This new family of MΔ does not account automatically for all overlap with the multinomial generating function. Any network with overlap in which all routes have the same length, will give equal choice probabilities; the network structure does not have any influence on the choice probabilities. Also, the conditional choice situation with reference route r does account for overlap between r and all other alternatives, but cannot capture dependencies between all these other alternatives. So, a specific generating function that can handle these dependencies has to be specified per reference route.

5.1 Model instances

Using the four available generating functions we can derive several models based on multiplicative utility formulas based on reference routes. Here we only present the choice probabilities of route r conditional on r being the reference route, i.e., using $p = r$ in Equation (113)[10]. The derived choice probabilities seem to be rather complex; however, that is only due to the asymmetry of the generating vector. Basically, the choice probabilities of a MΔ, r-model are not more complex than those for a A- or M-model, and they can be derived easily for applications.

MΔ-MN The disadvantage of independence in basic MNL is inherited. Not all overlap is explicitly removed in the MΔ-case, dependencies between non-reference routes maintain. The choice probabilities are

$$P_r^{\text{MΔ-MN},r}(G^{\text{MN}}; \boldsymbol{y}^{\text{MΔ},r}) = \frac{1}{1 + \sum_{p \in R \setminus \{r\}} \left(\frac{\sum_{l \in \mathscr{L}_{r \setminus p}} V_l}{\sum_{l \in \mathscr{L}_{p \setminus r}} V_l} \right)^{\mu}}, \quad \forall r \in R. \tag{115}$$

MΔ-PS The path-size choice probabilities are

$$P_r^{\text{MΔ-PS},r}(G^{\text{PS}}; \boldsymbol{y}^{\text{MΔ},r}) = \frac{\text{PS}_r^{\beta}}{\text{PS}_r^{\beta} + \sum_{p \in R \setminus \{r\}} \text{PS}_p^{\beta} \left(\frac{\sum_{l \in \mathscr{L}_{r \setminus p}} V_l}{\sum_{l \in \mathscr{L}_{p \setminus r}} V_l} \right)^{\mu}}, \quad \forall r \in R. \tag{116}$$

As it is possible to convert the path-size factor back into the utility formulation for the A-PS and M-PS models (Kitthamkesorn and Chen, 2013), this is also possible for the MΔ-PS model. This will lead to:

$$U_p^{\text{MΔ-PS},r} := \begin{cases} \frac{\sum_{l \in \mathscr{L}_r} V_l \varepsilon_r^{\text{M}}}{\text{PS}_r^{\beta}} & \text{if } p = r \\ \frac{\sum_{l \in \mathscr{L}_{p \setminus \mathscr{L}_r}} V_l \varepsilon_p^{\text{M}}}{\text{PS}_p^{\beta}} + \frac{\sum_{l \in \mathscr{L}_r \cap \mathscr{L}_p} V_l \varepsilon_r^{\text{M}}}{\text{PS}_r^{\beta}} & \text{otherwise.} \end{cases} \tag{117}$$

Having a reference route allows one to revisit the used path-size factors. These can be obtained by excluding all links of the reference route. This will lead to a formulation with reference route specific path-size parameters. Similar to equation (45), the formulation for the reference route r specific path-size factor $\text{PS}_{p,r}$ for route p is

$$\text{PS}_{p,r} = \begin{cases} \frac{\sum_{l \in \mathscr{L}_{p \setminus r}} \frac{V_l}{\#_l}}{\sum_{l \in \mathscr{L}_{p \setminus r}} V_l}, & \text{if } p \neq r \\ 1 & \text{if } p = r. \end{cases} \tag{118}$$

---

[10]Under the Markov chain assumption, these are the probabilities for staying in the same state.

MΔ-PC The paired combinatorial choice probabilities for in the MΔ-case are

$$P_r^{\text{MΔ-PC},r}(G^{\text{PC}}; \boldsymbol{y}^{\text{MΔ},r}) =$$

$$\frac{\sum_{p\in R\setminus\{r\}}\left(1+\left(\frac{\sum_{l\in\mathscr{L}_{r\setminus p}} V_l}{\sum_{l\in\mathscr{L}_{p\setminus r}} V_l}\right)^{\frac{\mu}{1-\varphi_{rp}}}\right)^{-\varphi_{rp}}}{\sum_{r'\in R\setminus\{r\}}\left(1+\left(\frac{\sum_{l\in\mathscr{L}_{r\setminus r'}} V_l}{\sum_{l\in\mathscr{L}_{r'\setminus r}} V_l}\right)^{\frac{\mu}{1-\varphi_{rr'}}}\right)^{1-\varphi_{rr'}}+\sum_{p\in R\setminus\{r,r'\}}\left(\left(\frac{\sum_{l\in\mathscr{L}_{r\setminus r'}} V_l}{\sum_{l\in\mathscr{L}_{r'\setminus r}} V_l}\right)^{\frac{\mu}{1-\varphi_{r'p}}}+\left(\frac{\sum_{l\in\mathscr{L}_{r\setminus p}} V_l}{\sum_{l\in\mathscr{L}_{p\setminus r}} V_l}\right)^{\frac{\mu}{1-\varphi_{r'p}}}\right)^{1-\varphi_{r'p}}}, \quad \forall r\in R.$$

(119)

This formula looks complicated, as the cumbersome indexing is required because the non-symmetric generating vector $\boldsymbol{y}^{\text{MΔ},r}$ is substituted, and then some terms collapse to 1. Without this substitution the probabilities are

$$P_r^{\text{MΔ-PC},r}(G^{\text{PC}}; \boldsymbol{y}^{\text{MΔ},r}) = \frac{\sum_{p\in R\setminus\{r\}}\left(y_{rr}^{\text{MΔ}\frac{\mu}{1-\varphi_{rp}}}+y_{rp}^{\text{MΔ}\frac{\mu}{1-\varphi_{rp}}}\right)^{-\varphi_{rp}}}{\sum_{r'\in R}\sum_{p\in R\setminus\{r'\}}\left(y_{rr'}^{\text{MΔ}\frac{\mu}{1-\varphi_{r'p}}}+y_{rp}^{\text{MΔ}\frac{\mu}{1-\varphi_{r'p}}}\right)^{1-\varphi_{r'p}}}, \quad \forall r\in R.$$

(120)

MΔ-LN The link-nested choice probabilities in the MΔ-case are

$$P_r^{\text{MΔ-LN},r}(G^{\text{LN}}; \boldsymbol{y}^{\text{MΔ},r}) = \frac{\sum_{m\in\mathscr{L}}\alpha_{mr}\left(\alpha_{mr}+\sum_{p\in R\setminus\{r\}}\alpha_{mp}\left(\frac{\sum_{l\in\mathscr{L}_{r\setminus p}} V_l}{\sum_{l\in\mathscr{L}_{p\setminus r}} V_l}\right)^{\mu_m}\right)^{\frac{\mu}{\mu_m}-1}}{\sum_{m\in\mathscr{L}}\left(\alpha_{mr}+\sum_{p\in R\setminus\{r\}}\alpha_{mp}\left(\frac{\sum_{l\in\mathscr{L}_{r\setminus p}} V_l}{\sum_{l\in\mathscr{L}_{p\setminus r}} V_l}\right)^{\mu_m}\right)^{\frac{\mu}{\mu_m}}}, \quad \forall r\in R.$$

(121)

5.2 Model properties

Since for the MΔ case utilities are sums of two variates of the MEV distributions, it is not possible to obtain a convenient closed-form for the joint density function of the utilities. Therefore, we cannot derive the expected maximum utility. However, from the marginal distributions of $\boldsymbol{\varepsilon}^{\text{M}}$ it is possible to obtain the expected value and variance of the MΔ utilities conditional on a reference route, which are:

$$\mathbb{E}(U_p^{\text{MΔ},r}) = \begin{cases} \frac{\sum_{l\in\mathscr{L}_r} V_l}{G(\mathbf{1}_r)^{1/\mu}}\Gamma\left(1+\frac{1}{\mu}\right) & \text{if } p=r \\ \left(\frac{\sum_{l\in\mathscr{L}_p\setminus\mathscr{L}_r} V_l}{G(\mathbf{1}_p)^{1/\mu}}+\frac{\sum_{l\in\mathscr{L}_r\cap\mathscr{L}_p} V_l}{G(\mathbf{1}_r)^{1/\mu}}\right)\Gamma\left(1+\frac{1}{\mu}\right) & \text{otherwise} \end{cases}, \text{ and,}$$

(122)

$$\text{Var}(U_p^{\text{MΔ},r}) = \begin{cases} \frac{\left(\sum_{l\in\mathscr{L}_r} V_l\right)^2}{G(\mathbf{1}_r)^{2/\mu}}\left(\Gamma\left(1+\frac{2}{\mu}\right)-\Gamma\left(1+\frac{1}{\mu}\right)^2\right) & \text{if } p=r \\ \left(\frac{\left(\sum_{l\in\mathscr{L}_p\setminus\mathscr{L}_r} V_l\right)^2}{G(\mathbf{1}_p)^{2/\mu}}+\frac{\left(\sum_{l\in\mathscr{L}_r\cap\mathscr{L}_p} V_l\right)^2}{G(\mathbf{1}_r)^{2/\mu}}\right)\left(\Gamma\left(1+\frac{2}{\mu}\right)-\Gamma\left(1+\frac{1}{\mu}\right)^2\right) & \text{otherwise.} \end{cases}$$

(123)

Furthermore, when the reference route is fixed (i.e., $P^{\text{ref}}(r)=1$ for exactly one $r\in R$), then SUE formulation (87) also holds for the MΔ, r-case.
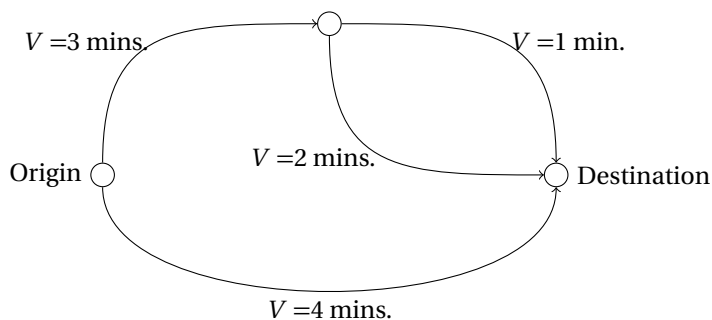
Figure 3: Simple three route overlap network.

Table 2: Choice probabilities for the MΔ-MN model on the simple network. The probabilities conditional on reference routes, as well as the two solution methods are provided.

	Upper	Middle	Lower
Ref. route = Upper	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
Ref. route = Middle	$\frac{8}{17}$	$\frac{4}{17}$	$\frac{5}{17}$
Ref. route = Lower	$\frac{5}{14}$	$\frac{4}{14}$	$\frac{5}{14}$
Equal $P^{\text{ref}}(\cdot)$ solution	≈ 0.409	≈ 0.240	≈ 0.350
Markov chain solution	≈ 0.401	≈ 0.239	≈ 0.359

5.3 Simple network

To provide more insight in the working of MΔ-models, the MΔ-MN probabilities are provided for a simple network. Figure 3 shows an origin and destination with three routes in between. Systematic utility is assumed to equal foreseen travel time, and no route specific constant is included. The upper and middle routes have overlap, and the upper route is faster. The lower route has no overlap with the other two, and is equally fast as the upper route. Table 2 shows the choice probabilities for each reference route based on $\mu = 1$, as well as, the solutions based on equal reference route probabilities and the Markov chain approach.

As expected, the slowest middle route has lowest choice probability. This choice probability is significantly smaller than it would be under A- and M-models, which is realistic since changing from the upper to the middle route means that the non-overlapping travel time doubles. Also, the two different solution methods for dealing with multiple reference routes do not differ that much.

On the other hand, there is a higher preference for the upper route than for the lower route while they have equal travel times. Based on the theory in Section 2, one would argue that the opposite should be true. Having the upper or lower route as the reference route, causes no difference in choice probabilities between the upper and lower route. Only because the middle route exists, and since it is more 'profitable' (i.e. the factor 2 between the non-overlapping parts between middle and lower is higher than the factor 4/3 between middle and lower route) to switch to the upper route, there is a slight preference for the upper route. This means that by having routes included as reference routes, the final

probabilities change (similar to what Bliemer and Bovy (2008) show).

So, this simple example provides insight in the working of the MΔ-models, but does not show its full potential. Clearly, the MΔ-models have different behaviour than all other known route choice models. The next two sections provide more insight in the differences between the A-, M-and MΔ-models – and their suitability for traffic assignment applications –. First, we analyse their basic behaviour under simple changes in the network configuration, and second, we analyse all models under more complex network variations. The latter quantitative test shows which model can best approximate the generic utility formulation of Section 2. These Sections provide the full potential of the MΔ-models.

6   Basic model behaviour under simple network changes

In this section we discuss the change in choice behaviour under three simple network adjustments. Consider the four networks depicted in the first column of Table 3. Network A is the basis and has 2 routes with different lengths of which the final parts overlap. In network B the overlapping part is extended, while in networks C1 and C2 the non-overlapping parts are extended by respectively adding a constant length and by multiplying their lengths. The models are so simple that, regardless the *exact* choice behaviour, the trend in choice probabilities for the two routes is known. The trend can either be that the probabilities remain equal, converge, or diverge. It is desired that choice models can reproduce the expected trend for each network change. However, not all trends can be reproduced by all choice models. Therefore, we analyse the behaviour of the three basic (-MN) models. We take advantage of having only two routes for the MΔ-MN model here, which avoids dealing with reference routes. However, similar results are obtained when extending to more than two routes, as is analysed in the next section.

Consider the switch from network A to network B. Since the non-overlapping parts remain equal, the choice probabilities for both routes obviously also remain equal. In the A-MN model only the difference between the routes matters, and since this difference does not change, the probabilities also remain equal for that model. For the M-MN model however, only the ratio between the routes matters, and this ratio changes. Because the non-overlapping part increases, the (lower:upper)-ratio decreases, and the choice probabilities will converge. The expected behaviour can not be reproduced by any M-MN model instance. The MΔ-MN model in the end, merely considers the ratio of the difference between routes, i.e., the ratio of the non-overlapping parts. Since the non-overlapping part does not change, the choice probabilities do not change. This holds for all MΔ-MN model instances

Consider the switch from network A to network C1. When a constant length is added to both routes, they become more similar the choice probability of the shortest route increases. So, it is expected that the route probabilities will converge[11]. The difference between the two routes will not change, therefore no A-MN model instance can reproduce the expected behaviour. On the other hand, the M-MN and MΔ-MN models will reproduce the expected behaviour since the (lower:upper)-ratio between routes (in- and excluding the overlapping part) decreases. The probabilities converge for all M-MN and MΔ-MN model instances.

Consider the switch from network A to network C2. When a both routes are multiplied by the same factor, the absolute detour of the longest route will increase, and will therefore be chosen less. So, it is expected that route probabilities will diverge[12]. The difference between the two routes will increase,

---

[11]As an easy example, consider routes with length x and $x + 1$; the choice probabilities will become 50 %-50% if $x \to \infty$ and 100%-0% if $x \to 0$

[12]As an easy example, consider routes with length x and $2 \times x$; the choice probabilities will become 50 %-50% if $x \to 0$ and

Table 3: **Behaviour under changing networks.** Networks B, C1 and C2 are slight variations on network A. Each network has two routes; the table shows the trend of the route probabilities when one switches for network A to any of the other three networks. They can either converge, diverge, or remain equal (depicted with arrows). The expected trend and achievable trends for three models are provided (see main text).

Network	Expected choice behaviour	A-MN	M-MN	MΔ-MN
A (1, 1, 2)				
B (1, 2, 2)	=	= ✓	↓↑ ✗	= ✓
C1 (2, 1, 3)	↓↑	= ✗	↓↑ ✓	↓↑ ✓
C2 (2, 1, 4)	↑↓	↑↓ ✓	↑↓ ✓	↑↓ ✓

therefore the probabilities in the A-MN models will diverge, which coincides with the expected behaviour. The ratio of the non-overlapping parts of the routes remains equal, but due to the overlapping part, the ratio between the whole routes changes. This increase in the (lower:upper)-ratio will lead to the desired diverging probabilities in the M-MN model. Of course, the divergence 'speed' depends on the length of the overlapping part. However, this can be adjusted by using the constant in the systematic utility specification. This constant is also the reason that the diverging probabilities are obtainable in the MΔ-MN model; here it also completely determines the divergence 'speed', which is an advantage compared to M-MN. All M-MN model instances diverge for this network example, but the behaviour remains dependent on the overlapping length. Almost all MΔ-MN model instances will have the expected behaviour, except for those with a constant equal to zero.

This analysis, summarized in Table 3, on the most simple and basic network variation shows that only the MΔ-MN model can reproduce all expected behaviours. We believe no simple network change exists of which the expected behaviour can be captured by A-MN or M-MN, but not by MΔ-MN. Since real networks do not consist of merely simple network changes, the next section analyses the competitiveness of all models on more comprehensive networks based on the route utility formulation of section 2, including the PS, PC, and LN variants, and including more than two routes which requires the use of reference routes for the MΔ-models.

100%-0% if $x \to \infty$

7 Network Example

The route choice models are applied on a network with three routes to show the advantage of the MΔ-models compared to the others. Bliemer and Bovy (2008) compare route choice models on different route sets. They demonstrate that existing closed-form route choice models (except for MNL) are sensitive to irrelevant route alternatives. Hence, if the route sets in application are different from the route sets in estimation, the results may be poor. This network example considers very different choice situations and the models are estimated and validated on different subsets of the data.

In Figure 4 the network with link travel times and routes is depicted. This network has been carefully constructed to put the choice models under stress, namely with and without overlap, and with short and long distance OD pairs. Also, the relative influence of the analyst and random foreseen travel time changes. The travel times of links 1, 2, 4 and 5 depend on variable x and by varying x different choice situations are created. Each value of x can represent a different OD pair in a transport network. The graph shows how route travel times increase with x; for $x = 0$ the difference (respectively ratio) between the slowest and fastest routes is four minutes (respectively 0.67), and this increases to eight minutes (respectively 0.90) for $x = 40$. Furthermore, overlap occurs on links 1 and 5.
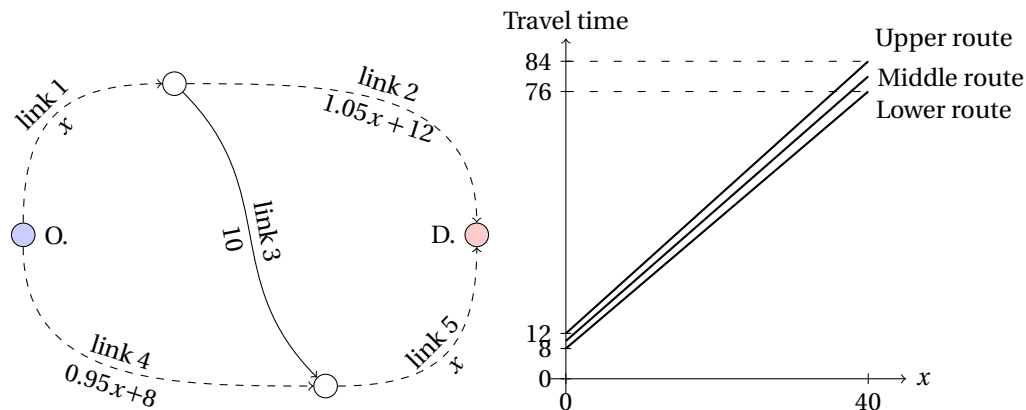

Figure 4: Network and travel times. Dashed links have variable costs.

For the experiment, the probabilities from the MNP model – that can properly handle both the analyst error and random foreseen travel time, and the desired variance-covariance structure based on the scaling postulate – are the ground truth. Thus, the route utilities are jointly distributed following a multivariate normal distribution. This distribution is specified in line with the random route utility of Section 2. Assume that there is no systematic utility other than travel time, then identification leads to $V_r^0 = 0$ for the A-models, and $V_r^0 = c$ for the M- and MΔ-models (see Section 4.3). For the normalization set the expected analyst error $\mathbb{E}(\varepsilon_r)$ to 0 and the travel time parameter β to -1. For the proportionality parameter (regarding the standard deviation of the foreseen travel time), linear regression on the OViN data of Figure 1 leads to $\theta = 0.3859$, and linear regression on the route survey leads to $\theta = 0.1301$; however, the first value is too high (see Section 2.1) and the latter too low (since it is the response of only one traveller), thus we assume $\theta = 0.2$. The covariances between the random foreseen travel times are assumed to be based on the arithmetic mean, see Equation (9). Note that these definitions of the (co)variances are different from those in the literature that assume proportionality between variance

31

and mean. The standard deviation of the analyst error is set to $\sigma(\varepsilon_r) = 10$ (minutes). This leads to the following multivariate normal utility distribution:

$$
\mathcal{N}\left(-\underbrace{\begin{pmatrix} 2.05x + 12 \\ 2x + 10 \\ 1.95x + 8 \end{pmatrix}}_{\mathbb{E}(\boldsymbol{U})}, 0.2^2 \underbrace{\begin{pmatrix} (2.05x+12)^2 & 2.025x^2 + 11x & 0 \\ 2.025x^2 + 11x & (2x+10)^2 & 1.975x^2 + 9x \\ 0 & 1.975x^2 + 9x & (1.95x+8)^2 \end{pmatrix}}_{\text{covariance matrix of } \tau} + \underbrace{\mathrm{diag}\begin{pmatrix} 100 \\ 100 \\ 100 \end{pmatrix}}_{\text{variance matrix of } \boldsymbol{\varepsilon}} \right).
$$

(124)

While x increases, not only the distance between origin and destination increases, but also the influence of the randomness from foreseen travel time compared to the randomness from the analyst error increases. For $x = 0$ only 3.8 percent of the variance of utility is due to the foreseen travel time; however, for $x = 40$, 76.4 percent of the variance is due to the foreseen travel time. Furthermore, the part of utility with overlap increases with x. In the development of this network example, the link travel times were chosen given the assumptions on the errors, and such that the route choice probabilities are more or less stable.
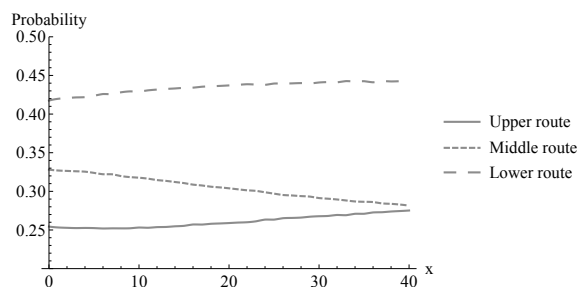

Figure 5: Ground truth probabilities based on a million multivariate normal samples for each $x \in \{0, \ldots, 40\}$.

For all $x \in \{0, 1, \ldots, 40\}$ the choice situation is simulated a million times by sampling from the multivariate normal distribution (Equation (124)), see Figure 5. The models are estimated twice; once on the dataset $x = \{5, \ldots, 15\}$, and once on dataset $x = \{25, \ldots, 35\}$ using log-likelihood maximization. So, eleven million 'observations' are used to estimate each model. For the MΔ-PS the path-size formulation of equation (45) is chosen[13]. For the LN models link specific scales are estimated for links 1 and 5; the other links – read nests – contain only one route and thus 'collapse': $(y^{\mu_l})^{\mu/\mu_l} = y^\mu$. The Markov chain approach for reference routes is chosen for MΔ-MN and MΔ-PS, and the 'equal probability reference route' approach is chosen for MΔ-PC and MΔ-LN[14]. For the validation, the models' log-likelihood on the other dataset is determined, so the parameters from the estimation on $x = \{5, \ldots, 15\}$ are applied to $x = \{25, \ldots, 35\}$, and vice versa.

Table 4 shows the estimated parameters.[15] The standard errors are very low due to the large artifi-

---

[13]This is chosen since equation (118) leads to much more path-size factors

[14]For these two models the software (Wolfram Mathematica) couldn't find analytical soltions of the choice probabilities using the Markov chain with unknown x. General applications do not have this parametrization on x, and thus shouldn't be problematic.

[15]The q-generalized logit model by Nakayama (2013) that captures both A-MN and M-MN has also been estimated; the results are not presented since the resulting model was always equivalent to the M-MN model (i.e., not to A-MN)

Table 4: Parameter estimates for every model for the two datasets

Model	Parameter estimates					
	Dataset $x = \{5,\ldots,15\}$			Dataset $x = \{25,\ldots,35\}$		
	Scale $\mu \approx$	Constant $c \approx$	Other	Scale $\mu \approx$	Constant $c \approx$	Other
A-MN	0.107			0.0699		
A-PS	0.107		$\beta \approx 0.182$	0.0681		$\beta \approx 0.501$
A-PC	0.0935			0.0576		
A-LN	0.0438		$\mu_1 \approx 0.095$ $\mu_5 \approx 0.818$	0.0225		$\mu_1 \approx 0.0225$ $\mu_5 \approx 1.091$
M-MN	11.518	-77.635		5.593	=0	
M-PS	12.932	-91.038	$\beta \approx 0.173$	8.690	-47.403	$\beta \approx 0.490$
M-PC	12.983	-108.293		7.809	-55.479	
M-LN	0.490	-142.935	$\mu_1 \approx 73.916$ $\mu_5 \approx 1175.47$	0.319	-0.258	$\mu_1 \approx 18.235$ $\mu_5 \approx 118.725$
MΔ-MN	7.475	-43.589		4.619	=0	
MΔ-PS	10.613	-72.889	$\beta \approx 0.145$	7.381	-41.176	$\beta \approx 0.432$
MΔ-PC	14.328	-126.223		7.396	-62.199	
MΔ-LN	0.475	-111.01	$\mu_1 \approx 58.367$ $\mu_5 \approx 7518.45$	0.569	=0	$\mu_1 \approx 15.265$ $\mu_5 \approx 61.098$

cial dataset and therefore not reported. Figure 6 shows the log-likelihoods of the estimation and validation results. Probabilities of the models are found in Figures 7 and 8. In every graph two instances of one model and the ground truth are shown, the blue lines depict the route probabilities for the model estimated on $x = \{5,\ldots,15\}$, the red lines depict the route probabilities for the model estimated on $x = \{25,\ldots,35\}$, and the grey lines depict the MNP (i.e., ground truth) probabilities.
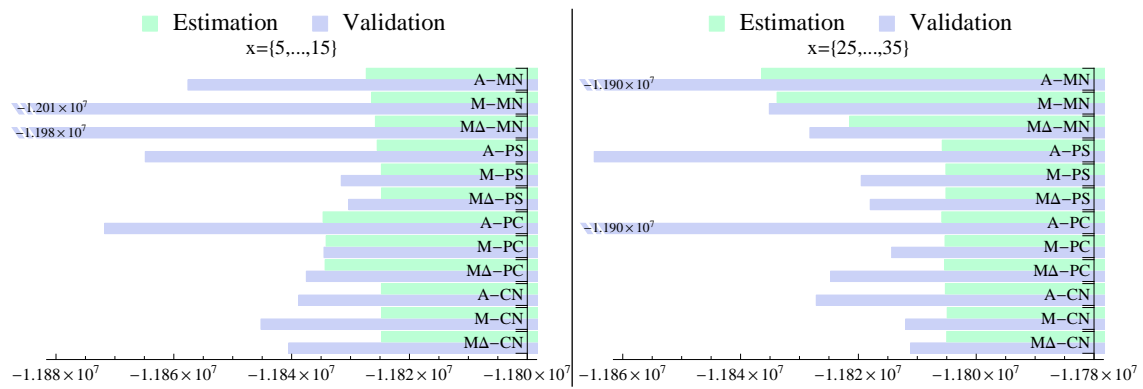
Figure 6: Log-likelihoods of model instances for estimation and validation. Parameter estimates of the left estimation are used for the validation in the right, and vice versa.
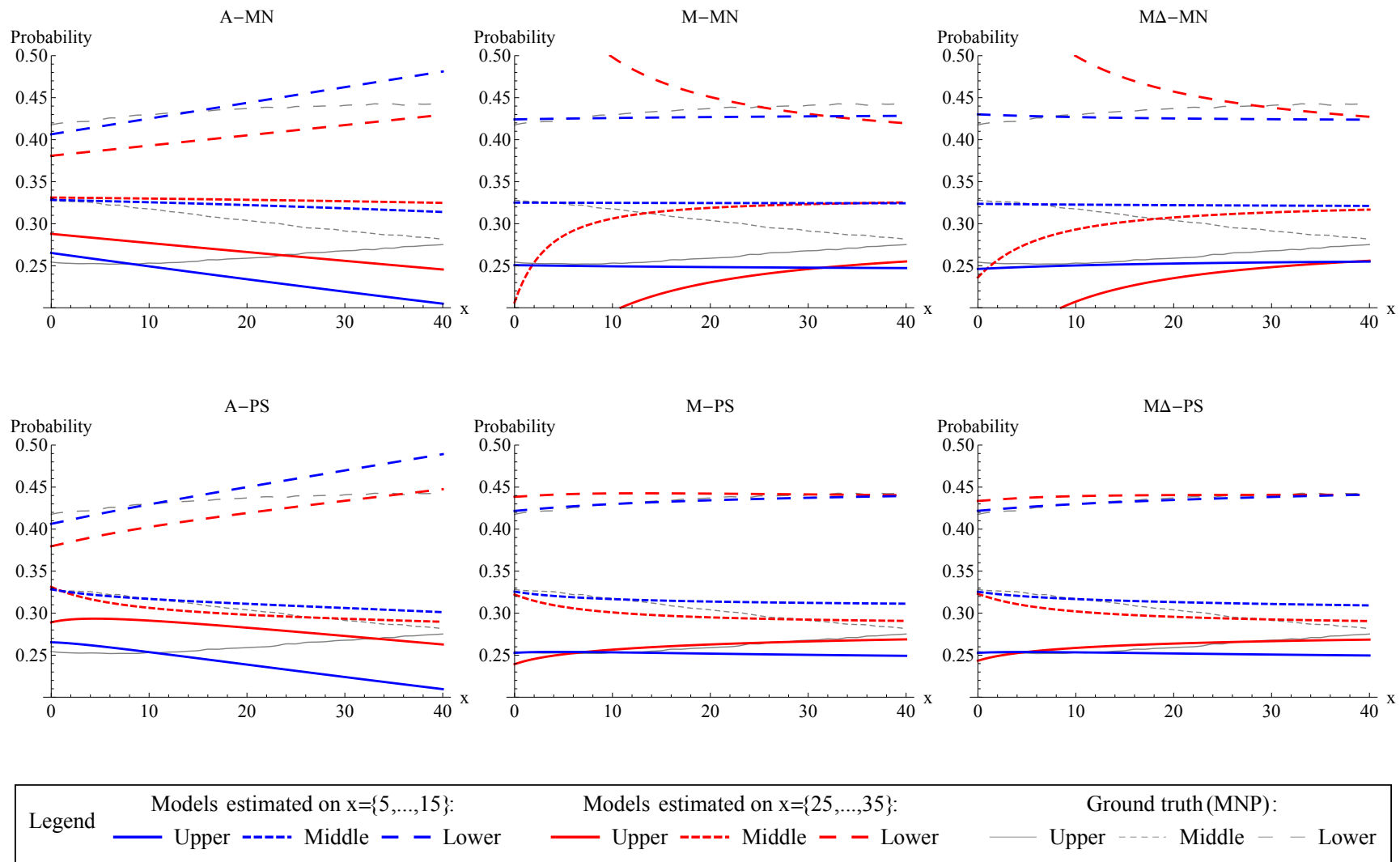
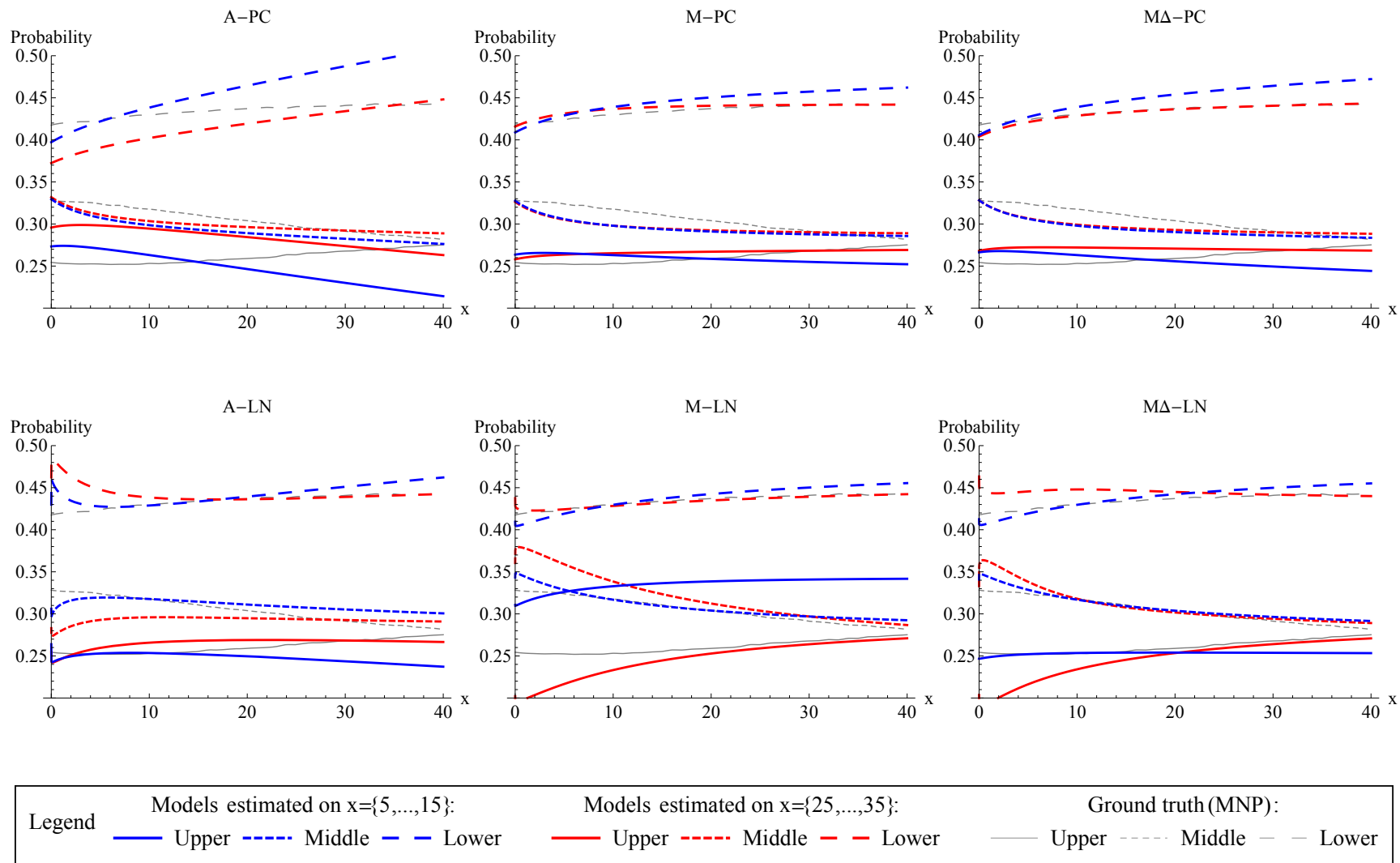Figure 7: Choice probabilities of the MN- and PS-models.

Figure 8: Choice probabilities of the PC- and LN-models.

The estimation results show that the MN models have the poorest log-likelihoods of all on the $\{25,\ldots,35\}$-data; these models cannot address the overlap of routes properly. The validation results of the MN models are also worse than the other types, so they cannot be transferred between short distance and long distance OD pairs. Another remark is that the PC models, which are designed to capture overlap, perform poorest on the $\{5,\ldots,15\}$-data; this is due the fact that the covariance between routes is low (because the analyst error is dominating), while the PC model always imposes dependencies.

In all cases, the A-models have the worst estimation result compared to the M- and MΔ-models. Remarkably, the multiplicative models also outperform the additive models on the $\{5,\ldots,15\}$-data, where the influence of the foreseen travel time is relatively small. The MΔ-models have better results on all four MN and PS estimations, while the M-models have better results on three out of four PC and LN estimations. This is because the additional parameter, the constant c, for the multiplicative models, which leads to a better fit.

The multiplicative errors capture the analyst error with constant c in the systematic utility. This constant is indeed larger for the models estimated on the $\{5,\ldots,15\}$-data, where the analyst error is dominant. Furthermore, note that all scale parameters μ are smaller for the $\{25,\ldots,35\}$-data (i.e., the variance is higher); this reflects the heteroscedasticity of route utility.

We would like to mention that the PC and LN estimation were problematic on other network configurations that we have tried. For the multiplicative models extremely high scales occurred, which can lead to numerical problems. They also generated very unrealistic probabilities outside the area they were estimated on. As mentioned earlier, it is infeasible to estimate all link-specific nest-scales in large networks for LN models.

8 Conclusions and Discussion

This paper presented twelve route choice models – of which seven are new – in a single framework, and assessed them qualitatively and quantitatively. Choice probabilities for all models have the same closed form expression, namely Equation (26), based on a generating function and a generating vector. The generating function determines how route overlap is captured, which is either multinomial (MN), path-size (PS), paired combinatorial (PC), or link-nested (LN). The generating vector determines the utility formula, which is either additive RUM (A), multiplicative RUM (M), or the newly presented multiplicative RUM based on reference routes that only considers differences between routes (MΔ).

For the qualitative assessment a basic structure of utility with random foreseen travel time was presented (see Section 2). We base our analysis on two postulates on random travel times of road segments that each lead to a different structure of randomness. Empirical evidence provides the new insight that the foreseen travel time distribution's mean and standard deviation have a linear relationship, contrary to a linear relationship between its mean and variance. The homoscedastic additive models are not able to capture the random foreseen travel time, but multiplicative models do allow for this. Furthermore, differences in normalization, identification, and invariance are pointed out. The constant in systematic utility in multiplicative models does not have to be normalized. This allows more degrees of freedom and a better fit on the data, but makes it more difficult to compare models from the different paradigms directly. One main advantage of the generic GMEV framework is that it can be analysed as a whole, we show this by providing the equivalent stochastic user equilibrium formulation for all models.

To show the distinctiveness of the MΔ-models, each model's behaviour under basic network changes was analysed. Only the MΔ-models can reproduce realistic behaviour when the characteristics of parallel and serial links change.

To test the models' potential on real networks, and to test whether they can be applied on datasets on which they are not estimated, a carefully constructed network example was presented. Based on our analyses, we expect good performance of the M-PS and M∆-PS models for route choice on real networks. They can capture overlap sufficiently, and they can handle random foreseen travel time. Also Fosgerau and Bierlaire (2009) and Chikaraishi and Nakayama (2015) have both compared additive with multiplicative formulations on multiple datasets, and they found that the multiplicative models have a better fit for all datasets.[16] However, additional empirical estimation and validation is required to conclusively assess all models. Finally, the PC and LN models are problematic to estimate on some other networks we tried, and they can cause numerical problems.

This paper does not discuss the route generation or sampling related to the explicit route sets of the models. As pointed out earlier, a correct sample of routes is required to obtain unbiased parameter estimates (Frejinger et al., 2009). For econometrically sound applications, sampling techniques as presented by (Frejinger et al., 2009; Flötteröd and Bierlaire, 2013; Guevara and Ben-Akiva, 2013) have to be adapted for the new models. Models with implicit route sets (Dial, 1971; Papola and Marzano, 2013; Fosgerau et al., 2013a; Mai et al., 2015) do not have this issue, but they might lead to unrealistic routes. The connection between the link-based MEV model of (Papola and Marzano, 2013) with the route-based GMEV framework does not seem feasible due to the different base units, but if it exists, it might lead to new GMEV model instances. On the other hand, Prato (2012) points out conceptual and empirical reasons that plea for the explicit approach.

9 Acknowledgements

References

Bekhor, S., Ben-Akiva, M., and Ramming, M. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144(1):235–247.

Bekhor, S., Ben-Akiva, M., and Scott Ramming, M. (2002). Adaptation of logit kernel to route choice situation. *Transportation Research Record: Journal of the Transportation Research Board*, 1805:78–85.

Bekhor, S. and Prashker, J. N. (1999). Formulations of extended logit stochastic user equilibrium assignments. In *14th International Symposium on Transportation and Traffic Theory*.

---

[16]Note that Chikaraishi and Nakayama (2015) state their model equals logit when $q = 0$ and to weibit when $q = 1$; however, they do not fully use the flexibility of weibit. When they would have estimated a constant for the weibit models (similar to the approach in this paper), the q-generalized logit models with $0 < q < 1$ will also become weibit.

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In Hall, R., editor, *Handbook of Transportation Science*, volume 23 of *International Series in Operations Research & Management Science*, pages 5–33. Springer US.

Ben-Akiva, M. E. (1973). *Structure of passenger travel demand models.* PhD thesis, Massachusetts Institute of Technology.

Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model. *Annals of Operations Research*, 144(1):287–300.

Bierlaire, M., Bolduc, D., and McFadden, D. (2008). The estimation of generalized extreme value models from choice-based samples. *Transportation Research Part B: Methodological*, 42(4):381 – 394.

Bliemer, M. and Bovy, P. (2008). Impact of route choice set on route choice probabilities. *Transportation Research Record: Journal of the Transportation Research Board*, 2076:10–19.

Bliemer, M. C., Raadsen, M. P., Smits, E.-S., Zhou, B., and Bell, M. G. (2014). Quasi-dynamic traffic assignment with residual point queues incorporating a first order node model. *Transportation Research Part B: Methodological*, 68(0):363–384.

Bovy, P., Bekhor, S., and Prato, C. (2008). The factor of revisited path size: Alternative derivation. *Transportation Research Record: Journal of the Transportation Research Board*, 2076:132–140.

Cascetta, E., Nuzzolo, A., Russo, F., and Vitetta, A. (1996). A modified logit route choice model overcoming path overlapping problems. sepcification and some calibration results for interurban networks. In *Proceedings of the 13 th International Symposium on Transportation and Traffic Theory*.

Cascetta, E., Russo, F., Viola, F. A., and Vitetta, A. (2002). A model of route perception in urban road networks. *Transportation Research Part B: Methodological*, 36(7):577–592.

Castillo, E., Menéndez, J. M., Jiméez, P., and Rivas, A. (2008). Closed form expressions for choice probabilities in the weibull case. *Transportation Research Part B: Methodological*, 42(4):373–380.

Chen, A., Kasikitwiwat, P., and Ji, Z. (2003). Solving the overlapping problem in route choice with paired combinatorial logit model. *Transportation Research Record: Journal of the Transportation Research Board*, 1857(-1):65–73.

Chen, A., Pravinvongvuth, S., Xu, X., Ryu, S., and Chootinan, P. (2012). Examining the scaling effect and overlapping problem in logit-based stochastic user equilibrium models. *Transportation Research Part A: Policy and Practice*, 46(8):1343–1358.

Chen, H.-K. (1999). *Dynamic travel choice models: A variational inequality approach.* Springer.

Chikaraishi, M. and Nakayama, S. (2015). How well drivers can estimate travel time? estimating travel-time-dependent variance based on a q-generalized logit model. In *Proceedings of the 14th International Conference on Travel Behaviour Research, Windsor, UK.*

Chu, C. (1989). A paired combinatorial logit model for travel demand analysis. In *Proceedings of the Fifth World Conference on Transportation Research*, volume 4, pages 295–309.

Daganzo, C. F. and Sheffi, Y. (1977). On stochastic models of traffic assignment. *Transportation Science*, 11(3):253–274.

Daly, A. and Bierlaire, M. (2006). A general and operational representation of generalised extreme value models. *Transportation Research Part B: Methodological*, 40(4):285–305.

Dial, R. B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5(2):83–111.

Facchinei, F. and Pang, J.-S. (2003). *Finite-dimensional variational inequalities and complementarity problems*, volume 1. Springer.

Fellendorf, M. and Vortisch, P. (2010). Microscopic traffic flow simulator vissim. In Barceló, J., editor, *Fundamentals of Traffic Simulation*, volume 145 of *International Series in Operations Research & Management Science*, pages 63–93. Springer New York.

Fiorenzo-Catalano, S., Van Nes, R., and Bovy, P. H. (2004). Choice set generation for multi-modal travel analysis. *European journal of transport and infrastructure research EJTIR, 4 (2)*.

Fisk, C. (1980). Some developments in equilibrium traffic assignment. *Transportation Research Part B: Methodological*, 14(3):243–255.

Flötteröd, G. and Bierlaire, M. (2013). Metropolis–hastings sampling of paths. *Transportation Research Part B: Methodological*, 48(0):53–66.

Fosgerau, M. (2010). On the relation between the mean and variance of delay in dynamic queues with random capacity and demand. *Journal of Economic Dynamics and Control*, 34(4):598–603.

Fosgerau, M. and Bierlaire, M. (2009). Discrete choice models with multiplicative error terms. *Transportation Research Part B: Methodological*, 43(5):494–505.

Fosgerau, M., Frejinger, E., and Karlstrom, A. (2013a). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, 56:70–80.

Fosgerau, M., McFadden, D., and Bierlaire, M. (2013b). Choice probability generating functions. *Journal of Choice Modelling*, 8:1–18.

Frejinger, E. and Bierlaire, M. (2006). Capturing correlation in large-scale route choice models. Technical report, École polytechnique fédérale de Lausanne. RO-060106.

Frejinger, E., Bierlaire, M., and Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, 43(10):984–994.

Gálvez, T. E. (2002). Elmodelo powit. *Urbana*, 7(30).

Gliebe, J., Koppelman, F., and Ziliaskopoulos, A. (1999). Route choice using a paired combinatorial logit model. In *78th TRB Meeting, Washington DC*.

Guevara, C. A. and Ben-Akiva, M. E. (2013). Sampling of alternatives in multivariate extreme value (mev) models. *Transportation Research Part B: Methodological*, 48(0):31–52.

Guo, X., Yang, H., and Liu, T.-L. (2010). Bounding the inefficiency of logit-based stochastic user equilibrium. *European Journal of Operational Research*, 201(2):463–469.

Hellinga, B., van Lint, H., and Hofman, F. (2012). The use of exogenously defined standard deviation versus mean travel time relationships for estimating the impact of policy measures on reliability. In *91st Annual Meeting of the Transportation Research Board, Washington DC, USA*.

Kitthamkesorn, S. and Chen, A. (2013). A path-size weibit stochastic user equilibrium model. *Transportation Research Part B: Methodological*, 57:378–397.

Kitthamkesorn, S. and Chen, A. (2014). Unconstrained weibit stochastic user equilibrium model with extensions. *Transportation Research Part B: Methodological*, 59:1–21.

Koppelman, F. S. and Wen, C.-H. (2000). The paired combinatorial logit model: properties, estimation and application. *Transportation Research Part B: Methodological*, 34(2):75–89.

Li, B. (2011). The multinomial logit model revisited: A semi-parametric approach in discrete choice analysis. *Transportation Research Part B: Methodological*, 45(3):461–473.

Mahmassani, H., Hou, T., and Dong, J. (2012). Characterizing travel time variability in vehicular traffic networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2315:141–152.

Mahmassani, H., Hou, T., and Saberi, M. (2013). Connecting network-wide travel time reliability and the network fundamental diagram of traffic flow. *Transportation Research Record: Journal of the Transportation Research Board*, In Press:1–22.

Mai, T., Fosgerau, M., and Frejinger, E. (2015). A nested recursive logit model for route choice analysis. *Transportation Research Part B: Methodological*, 75:100 – 112.

Marzano, V. (2014). A simple procedure for the calculation of the covariances of any generalized extreme value model. *Transportation Research Part B: Methodological*, 70(0):151–162.

Marzano, V., Papola, A., Simonelli, F., and Vitillo, R. (2013). A practically tractable expression of the covariances of the cross-nested logit model. *Transportation Research Part B: Methodological*, 57:1–11.

Mattsson, L.-G., Weibull, J. W., and Lindberg, P. O. (2014). Extreme values, invariance and choice probabilities. *Transportation Research Part B: Methodological*, 59:81–95.

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

McFadden, D. (1978). Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial Interaction Theory and Planning Models*. North Holland.

Nagurney, A. (1998). *Network economics: A variational inequality approach*, volume 10. Springer.

Nakayama, S. (2013). q-generalized logit route choice an network equilibrium model. *Procedia - Social and Behavioral Sciences*, 80:753–763. 20th International Symposium on Transportation and Traffic Theory (ISTTT 2013).

Nakayama, S. and Chikaraishi, M. (2015). A unified closed-form expression of logit and weibit and its application to a transportation network equilibrium assignment. *Transportation Research Procedia (of ISTTT21)*, 7:59–74.

Newman, J. P. (2008). Normalization of network generalized extreme value models. *Transportation Research Part B: Methodological*, 42(10):958–969.

Papola, A. (2004). Some developments on the cross-nested logit model. *Transportation Research Part B: Methodological*, 38(9):833–851.

Papola, A. and Marzano, V. (2013). A network generalized extreme value model for route choice allowing implicit route enumeration. *Computer-Aided Civil and Infrastructure Engineering*, 28(8):560–580.

Prashker, J. and Bekhor, S. (1998). Investigation of stochastic network loading procedures. *Transportation Research Record: Journal of the Transportation Research Board*, 1645(-1):94–102.

Prashker, J. N. and Bekhor, S., S. (2004). Route choice models used in the stochastic user equilibrium problem: A review. *Transport Reviews*, 24(4):437–463.

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, 2:65–100.

Prato, C. G. (2012). Meta-analysis of choice set generation effects on route choice model estimates and predictions. *Transport*, 27(3):286–298.

Pravinvongvuth, S. and Chen, A. (2005). Adaptation of the paired combinatorial logit model to the route choice problem. *Transportmetrica*, 1(3):223–240.

Ramos, G. M. (2015). *Dynamic Route Choice Modelling of the Effects of Travel Information using RP Data*. PhD thesis, Delft University of Technology.

Ramos, G. M., Frejinger, E., Daamen, W., and Hoogendoorn, S. (2012). A revealed preference study on route choices in a congested network with real-time information. In *Proceedings of the 13th International Conference on Travel Behaviour Research-Toronto*, pages 15–20.

Vovsha, P. and Bekhor, S. (1998). Link-nested logit model of route choice: Overcoming route overlapping problem. *Transportation Research Record: Journal of the Transportation Research Board*, 1645:133–142.

Xu, X., Chen, A., Kitthamkesorn, S., Yang, H., and Lo, H. K. (2015). Modeling absolute and relative cost differences in stochastic user equilibrium problem. *Transportation Research Part B: Methodological*, pages –.

Yai, T., Iwakura, S., and Morichi, S. (1997). Multinomial probit with structured covariance for route choice behavior. *Transportation Research Part B: Methodological*, 31(3):195–207.

Yildirimoglu, M., Koymans, A., and Geroliminis, N. (2013). Exploring the properties of mean-variance relations in freeway travel times. In *Transportation Research Board 92nd Annual Meeting*.

Zhou, B., Bliemer, M. C. J., Bell, M. G. H., and He, J. (2015). Two new methods for solving the path-based stochastic user equilibrium problem. *Computer-Aided Civil and Infrastructure Engineering*, pages n/a–n/a.

Zhou, Z., Chen, A., and Bekhor, S. (2012). C-logit stochastic user equilibrium model: formulations and solution algorithm. *Transportmetrica*, 8(1):17–41.