Jourquin *European Transport Research Review* (2026) 18:102
https://doi.org/10.1186/s12544-026-00863-9

European Transport Research Review

RESEARCH Open Access

From modal-choice estimation to freight OD-matrix assignment: a calibrated framework for European multimodal transport models

Bart Jourquin[1*]

Abstract

Freight modal-choice models are commonly judged by their ability to reproduce aggregate modal shares. For network assignment, however, this is not enough: the model must also reproduce where the tonnes are located in modal origin-destination (OD) matrices. This paper proposes a calibrated framework that links modal-choice estimation with freight matrix assignment. A behavioural multinomial logit (MNL) model is estimated on aggregate OD-commodity flows for road, inland waterway (IWW) transport and rail. The model uses variables expressed relative to road, commodity-group constants, mode-specific distance splines and a rail market-structure correction. Since sparse rail and inland-waterway matrices cannot be reproduced accurately by the behavioural model alone, a bounded base-year calibration layer is added to the non-road utilities. The framework is tested on three European datasets: a European NUTS2 matrix, a "Benelux+" NUTS3 matrix and a domestic German NUTS3 matrix. Performance is assessed at both the aggregate modal-share level and the OD-cell level using Weighted Absolute Percentage Error (WAPE). The detailed calibration results are interpreted as in-sample base-year reconstruction, while transferability is examined separately through repeated OD-pair holdout experiments used to distinguish base-year reconstruction from performance on unseen OD pairs. The results show that the behavioural layer provides a useful structure, but that detailed bounded calibration is required for base-year modal OD-matrix reproduction. The estimated model is then implemented in Nodus, a transport network modelling software designed for freight transport, allowing the calibrated framework to be used for assignments.

Keywords Freight transport, Modal choice, OD matrices, Multinomial logit, Calibration, Network assignment

1 Introduction

Freight transport policy increasingly relies on models that can compare road, rail and inland waterway options under different pricing, infrastructure or network-performance assumptions. In Europe, this issue is particularly relevant because modal-shift policies are often framed around the possibility of moving part of the traffic currently assigned to road towards rail or inland waterways. For such applications, global modal shares are useful, but they are not sufficient. Assignment models require modal OD matrices: they need to know not only how many tonnes use each mode, but also which OD pairs and commodity groups those tonnes belong to.

This paper starts from a practical difficulty encountered in freight modelling. A modal-choice model may reproduce the total rail share or the total inland-waterway share while allocating the corresponding tonnes to the wrong OD cells. Such a model is unsuitable for

*Correspondence:
Bart Jourquin
bart.jourquin@uclouvain.be
[1]Center for Operations Research and Econometrics (CORE), UCLouvain, Louvain-la-Neuve, Belgium

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

network assignments. The problem is especially acute for non-road modes. Rail and inland-waterway matrices are sparse: many OD-commodity cells are technically available, but carry no observed non-road flow, while a relatively small number of corridors and commodities account for a large share of the traffic.

Freight transport modelling has long had to balance behavioural detail, data availability and operational usability. Reviews of national and international freight models stress the need to combine aggregate flows, logistics structure, modal split and network assignment in a coherent modelling system [12, 13, 26]. More recent models have introduced richer logistics and transport-chain representations, especially when shipment-level data are available [5, 15]. Yet in many applications, including recent large-scale European work [16], the available data remain regional OD matrices rather than individual shipments.

Discrete-choice models provide a natural language for modal choice. The multinomial logit model and its extensions are well established in transport demand analysis [6, 21, 27], and logit-type specifications have been used for freight mode, route, shipment and transport-chain choices [2, 9, 15, 22, 23]. These studies show that costs, times, shipment characteristics, commodity type and logistics organisation all matter. The research problem can thus be stated as follows:

> *How can an aggregate freight modal-choice model be estimated, calibrated and evaluated so that it remains behaviourally interpretable while reproducing modal OD matrices accurately enough for network assignment?*

The difficulty is that the two objectives are not identical. A behavioural model aims to explain systematic modal-choice effects, such as the influence of cost, moving time, distance and commodity groups. A matrix-reproduction model aims to allocate tonnes to the correct OD cells. For sparse freight matrices, a behavioural layer alone will usually assign positive probabilities to all available modes, thereby spreading small amounts of non-road traffic across many technically feasible but empirically unused OD cells. This behaviour is consistent with the logit structure, but problematic for assignments.

The natural starting point towards a solution is the conventional multinomial logit model. Let the utility V of mode m be:

$$V_m = \alpha_m + \beta_C C_m + \beta_{T,m} T_m \quad (1)$$

where:

C_m is the cost of mode m;
T_m is the transit time of mode m;
α_m is a mode-specific constant;
β_C is a cost coefficient;
$\beta_{T,m}$ is a mode-specific time coefficient.

Road is used as the reference mode, so its alternative-specific constant is normalised to zero. In such a specification, the signs of the β estimators are expected to be negative, because a cost increase or a longer transit time for a mode has a negative impact on the probability of choosing it.

The immediate problem is that cost and time are highly correlated. This is not surprising. Longer and less direct routes tend to be both more expensive and slower. In this application, the problem is amplified because both cost and time are constructed from the same assignment system and network representation. As a result, the specifications described in Eq. 1 (or similar) produce unstable coefficients, active bounds, and identification problems. A second problem is that the dependent variable is not a discrete choice made by one decision-maker. Instead, each row contains modal tonnes and therefore modal shares. The likelihood must thus be adapted accordingly.

The framework proposed in this paper combines two layers. The first is a behavioural multinomial logit model, using alternative utility functions that provide a more stable and systematic structure.

The second is a base-year calibration layer that adds utility corrections to the non-road alternatives. This layer is not meant to replace the first layer. Its role is to account for the OD, corridor and commodity features that are not captured by the available explanatory variables, and to restore the sparse structure of observed modal matrices. It is related to two well-established modelling traditions:

- The family of pivot-point and incremental demand models, in which the base-year situation is not ignored: observed base matrices or base shares are used as an anchor, and the model is mainly used to predict changes around that reference situation. Early incremental logit formulations were already motivated by the need to represent the effect of policy changes without rebuilding a full demand model from scratch [20]. Later work on nested incremental logit and pivot-point procedures formalised this idea for practical travel-demand forecasting, emphasising the value of combining behavioural response functions with observed base-year demand [4, 10, 11].
- The OD-matrix adjustment methods in which an initial or target OD matrix is updated using additional information, most often traffic counts, while trying to preserve as much as possible of the prior matrix structure. Entropy maximisation, information minimisation, generalised least squares and gradient-based formulations all share this

general objective: the adjusted matrix should be closer to observed network evidence, but should not become an arbitrary table of fitted numbers [1, 8, 24, 29].

The approach presented in this paper follows the same general logic, but it differs in where the calibration is applied, as the model does not directly adjust modal OD tonnes outside the choice model. Instead, the correction enters the non-road utilities as bounded calibration constants. These constants can be interpreted as residual utility terms: they capture OD-, commodity- and corridor-specific effects that are not represented by the available explanatory variables, while leaving the behavioural part of the multinomial logit model intact. Bounding the constants is important because it prevents the adjustment from becoming an unconstrained "fitting device".

This makes the framework closer to a utility-based pivot around the observed base-year modal matrices than to a purely behavioural modal-choice model. The behavioural coefficients describe systematic responses to cost, moving time, distance and market-structure variables. The calibration constants then restore the observed base-year geography of rail and inland-waterway matrices. The two layers therefore have different roles: the first provides behavioural structure, while the second provides a controlled matrix-reproduction mechanism needed for assignment.

Further, the evaluation strategy adopted in this paper follows a second, equally important modelling tradition: the distinction between fitting a base-year model and demonstrating that the model is useful for application. Standard transport-modelling guidance has long emphasised that calibration and validation are not the same exercises. Calibration adjusts a model so that it reproduces known base-year conditions, whereas validation asks whether the model reproduces observations, patterns or relationships that were not simply used as direct fitting targets. This distinction is explicit in practical model-validation manuals and guidance documents, including the FHWA *Travel Model Validation and Reasonableness Checking Manual* [28], and ATAP guidance on travel-demand model development [3].

A detailed "mode-OD-group" calibration can reproduce the base-year modal matrices very closely, but this does not by itself prove that the model is transferable. For this reason, the paper reports several complementary diagnostics. Aggregate modal shares are used as a first check, but a model can reproduce the total rail share while assigning rail tonnes to the wrong OD pairs. The main assessment of base-year matrix reconstruction therefore takes place at the OD-cell level, where observed and predicted tonnes are compared directly. This is consistent with the broader assignment-modelling view that a model should be assessed not only on global totals, but also on the spatial distribution of flows and on the reasonableness of the matrices that are assigned to the network.

The use of the Weighted Mean Absolute Percentage Error (WMAPE, also referred to as WAPE in parts of the forecasting literature) and bias reflects the same concern [14, 17, 19]. Freight OD matrices, especially for rail and inland waterways, contain many zero or very small cells. In that context, ordinary percentage errors such as MAPE can be unstable or undefined, a problem well known in the forecasting literature for intermittent or sparse demand. WAPE is therefore used as the OD-cell error measure because it relates the sum of absolute cell errors to the observed modal tonnage, while bias is reported separately to systematically identify over- or under-predictions.

Finally, the paper includes two checks that are specifically intended to avoid presenting calibration as validation. First, the distribution of calibration constants is reported, both unweighted and weighted by observed modal tonnes. This shows whether the fit is obtained through a small number of extreme corrections or through a more interpretable pattern of residual utility terms. Second, an OD-pair holdout test is performed: the model is estimated on a subset of the OD pairs and evaluated on the remaining cells, providing a more realistic test of the model structure that can be applied to unseen OD pairs. In this sense, the evaluation framework separates three questions that are often mixed together: whether the behavioural layer gives plausible modal shares, whether the calibrated model reconstructs the base-year matrices, and how much of the estimated structure transfers to OD pairs not used for calibration.

This distinction is important because the objective of the framework is not only to estimate plausible behavioural coefficients, but also to produce base-year modal OD matrices that are sufficiently accurate for assignment.

In summary, the contribution of this paper is threefold:

1. It proposes a relative-to-road multinomial logit specification for road, inland waterways and rail.
2. It introduces a bounded calibration procedure for base-year modal matrix fit and evaluates the model using OD-cell metrics.
3. It tests the same modelling architecture on three datasets with different spatial scales and market structures and shows how the estimated parameters can be used directly in a multimodal freight assignment model.

2 Data and modelling context

The framework is designed for annual OD-commodity matrices. Each observation corresponds to a flow between an origin, a destination and a commodity group. For each cell, annual tonnes are observed by road, inland waterway transport and rail, together with modal attributes such as cost, distance, moving duration and availability[1]. The dataset used in this study is derived from the ETISPlus European FP7 research programme [25]. It consists of estimated origin–destination (OD) matrices for the year 2010 at the NUTS-2 and NUTS-3 regional levels, covering road, rail, and inland waterway (IWW) transport. A detailed description of the dataset and its integration into a multimodal freight transport network assignment model is provided in Jourquin [18].

For an OD-commodity observation i, observed annual tonnes by mode are denoted by $q_{\text{road, i}}$, $q_{IWW,\,i}$ and $q_{rail,\,i}$. The total observed tonnes are simply the sum of the modal transported quantities:

$$Q_i = q_{road,i} + q_{IWW,i} + q_{rail,i} \tag{2}$$

and the observed modal share of mode m can be computed as:

$$s_{m,i} = \frac{q_{m,i}}{Q_i} \tag{3}$$

Only cells with strictly positive total tonnes Q_i are retained. Road is required to be technically available for all retained observations; cells for which no valid road path or attributes can be computed are excluded. Inland waterways and rail are available only when valid network-derived attributes exist. In this paper, modal availability refers initially to technical network availability. A non-road mode is treated as technically available when the assignment procedure identifies a valid path and the corresponding cost, distance and moving-time attributes are strictly positive. Technical availability does not necessarily imply that the mode belongs to the effective commercial choice set for a particular OD–commodity movement. Terminal access, service frequency, minimum shipment size, handling requirements, contractual arrangements and other logistics constraints may make a technically feasible alternative unattractive or practically unusable.

Observed non-use is also not equivalent to unavailability. A zero modal flow may indicate a structural absence from the effective choice set, but it may also be a realised zero for an otherwise feasible alternative. Since the available matrices provide one aggregate base-year observation for each OD–commodity cell and do not identify these mechanisms separately, the model cannot determine directly whether each observed zero is structural or incidental.

Intrazonal flows are not modelled. Each NUTS2 or NUTS3 region is represented by a single centroid, so an origin and destination located within the same region collapse to the same network node. Meaningful internal paths, distances, costs and moving times therefore cannot be constructed for these movements. Intrazonal OD cells are consequently excluded from the behavioural estimation, calibration and assignment, and the analysed sample contains interzonal flows only.

The framework is applied to three sub-datasets (Table 1). The first is a European NUTS2 dataset, used as a broad European benchmark that includes many longer-distance OD pairs. The second is a "Benelux+" NUTS3 dataset. The term *Benelux+* refers to the study area centred on Belgium, the Netherlands and Luxembourg, and extended to the adjacent surrounding regions included in the network model. More precisely, on a map using the Mercator projection, the area corresponds to a somewhat arbitrary rectangular area whose north-eastern limit is close to Friesland and whose south-western limit lies south of the Paris region. The "+" therefore indicates that the dataset is not limited strictly to the three Benelux countries, but also includes neighbouring areas required to represent the main freight corridors consistently. This region is of particular interest because IWW has a significant presence there. The third is a domestic German NUTS3 dataset, with detailed national/regional flows and substantial rail and inland-waterway structure. The purpose of using these three datasets is not to force a single coefficient vector onto all of them, but to test whether the same modelling architecture can be reused while keeping coefficients and calibration constants dataset specific.

Table 1 Dataset structure and observed modal shares

OD matrix	Granularity	Rows	Tonnes (millions)	Availability		Modal shares		
				IWW	Rail	Road	IWW	Rail
Europe	NUTS2	130,572	5,084	21.27%	99.46%	80.01%	6.18%	13.81%
Benelux+	NUTS3	328,664	2,587	49.98%	98.20%	79.72%	11.94%	8.34%
Germany	NUTS3	260,387	1,829	48.27%	99.10%	85.46%	2.45%	12.10%

All descriptive statistics and estimation results reported in Sects. 2, 3, 4, 5, and 6 refer to the same prepared estimation sample

[1] Not all transport modes are available between all origin-destination pairs.

The matrices distinguish ten commodity groups, indexed from 0 to 9[2]. Commodity information enters the model through commodity-specific constants and, in the detailed calibration, through commodity-specific correction factors. This is needed because modal choice in freight is strongly commodity dependent. Bulk flows, containerised goods, manufactured products and time-sensitive goods do not have the same relationship to rail or inland waterways.

The modal attributes are not observed prices or travel times contained directly in the ETISPlus OD matrices. ETISPlus provides the freight-demand matrices and the underlying digitised modal networks. These networks are prepared for assignment in Nodus, which computes a pure-mode path for each technically available mode, OD pair and commodity group. The resulting modal costs, distances and moving times are subsequently exported to the database used by the Python estimation framework. The Python script therefore reads the previously computed variables rather than recalculating them during estimation.

For mode m, OD–commodity observation i and commodity group g, total transport cost can be written schematically as

$$C_{m,i} = c_m^{\text{load}} + c_m^{\text{unload}} + \sum_{\ell \in R_{m,i}} c_{\ell,m}^{\text{move}} \tag{4}$$

where $R_{m,i}$ is the modal path (succession of links ℓ) generated by Nodus. Costs are expressed in euros per transported tonne, using cost functions in 2010 prices. The loading and unloading terms are fixed per-tonne costs that vary by mode and commodity group. The link-level moving costs depend on link length, mode or vehicle type, commercial speed and the corresponding unit operating cost. The underlying cost functions account for carrier-related components including labour, vehicle capital, fuel, maintenance, insurance and directly related operating and administrative costs. They also include the time-related component of the transported goods where applicable. Note that $C_{m,i}$ also depends on the group of commodities.

Vehicle loading assumptions are embedded in the Nodus cost functions. Load factors are exogenous and vary by commodity group and vehicle type. The implementation uses one representative truck type and one representative train type. For inland waterways, although six vessel classes are represented, for the sake of simplicity, the cost corresponding to CEMT2-type barges is used.

Loading and unloading costs are included in $C_{m,i}$, but they were not retained as separate explanatory variables in the final utility specification. Thus, the statement that handling costs were excluded from the behavioural model refers to their separate introduction alongside total cost, not to their removal from the total-cost measure itself. Loading and unloading durations are similarly not included in the moving-time variable used in the reference specification.

The model represents pure road, rail and inland-waterway alternatives rather than intermodal transport chains. Consequently, no transshipment cost is added between modes. Centroid connectors are included in the modal paths and their associated distance, time and moving cost, but they belong to the corresponding modal network and do not represent explicit road pre- or post-haulage for rail or inland-waterway transport.

The cost and moving-time variables are therefore not independently observed market data. They are constructed from modal paths generated within the same multimodal network representation. Path length, moving time and transport cost are therefore mechanically related: longer or less direct paths generally require more time and also generate higher calculated costs. This common construction creates substantial correlation between cost, moving time and distance, particularly when the available variation is mainly geographical rather than price driven.

The resulting coefficients should therefore be interpreted as conditional effects within the constructed network dataset. A coefficient reaching its bound may indicate that the corresponding variable provides little additional independent information after the other path-based attributes, distance splines, commodity constants and market-structure variables have been included. It does not establish that the underlying attribute is irrelevant to real freight decisions.

3 Behavioural modal-choice model specification

Since road is the dominant and virtually always available reference alternative, the non-road attributes are expressed as logarithmic ratios relative to road. This makes each non-road option comparable with the corresponding road movement for the same OD pair. A ratio of 1 maps to zero, values below 1 indicate that the non-road mode is cheaper or faster than road, and values above 1 indicate the opposite.

The specification therefore relies on a compact set of variables: relative inland-waterway cost, relative inland-waterway moving time, relative rail moving time, mode-specific distance splines, commodity-group constants and a rail market-structure correction based on inland-waterway availability.

The behavioural component is a multinomial logit model with the utility of road $V_{road,i}$ normalised to zero:

[2] Following the classical 10 main chapters of the NST/R nomenclature.

$$V_{road,i} = 0 \tag{5}$$

For inland waterways, the cost C and moving-time T variables are handled as:

$$c_{IWW,i} = \ln\left(\frac{C_{IWW,i}}{C_{road,i}}\right) \text{ and } t_{IWW,i} = \ln\left(\frac{T_{IWW,i}}{T_{road,i}}\right) \tag{6}$$

A similar ratio is used for rail moving time:

$$t_{rail,i} = \ln\left(\frac{T_{rail,i}}{T_{road,i}}\right) \tag{7}$$

All continuous explanatory variables are centred before estimation. This does not change the behavioural meaning of the coefficient, but it improves numerical conditioning and makes the alternative-specific constants easier to interpret: they correspond to the average observed conditions rather than to an arbitrary zero value. For inland waterways, for instance, the centred variables are:

$$\tilde{c}_{IWW,i} = \ln\left(\frac{C_{IWW,i}}{C_{road,i}}\right) - \overline{\ln\left(\frac{C_{IWW}}{C_{road}}\right)} \tag{8}$$

$$\tilde{t}_{IWW,i} = \ln\left(\frac{T_{IWW,i}}{T_{road,i}}\right) - \overline{\ln\left(\frac{T_{IWW}}{T_{road}}\right)} \tag{9}$$

Distance enters utilities through a piecewise-linear spline rather than through a single linear term. Let d_i denote the road distance of OD cell i. Three distance components are defined as:

$$d_{i1} = \min(d_i, 150) \tag{10}$$

$$d_{i2} = \min\{\max(d_i - 150, 0), 150\} \tag{11}$$

$$d_{i3} = \max(d_i - 300, 0) \tag{12}$$

These variables measure respectively the first 150 km, the distances between 150 and 300 km, and trips longer than 300 km. The thresholds were selected ex ante rather than estimated by maximising likelihood or minimising WAPE. They provide a common and interpretable segmentation across the three datasets: distances below 150 km represent relatively short regional movements, the 150–300 km interval represents an intermediate market in which non-road alternatives may begin to become competitive, and distances above 300 km represent longer-haul movements. The values were chosen to provide sufficiently broad segments across the different geographical scales represented by the three datasets[3]. Using the same thresholds across datasets preserves the common model architecture and avoids selecting dataset-specific knots that could improve in-sample fit mechanically. After scaling by 100 km and centring, they enter the utility of each non-road mode m as:

$$f_m(d_i) = \beta_{D,m,1}\tilde{d}_{i1} + \beta_{D,m,2}\tilde{d}_{i2} + \beta_{D,m,3}\tilde{d}_{i3} \tag{13}$$

where $\tilde{d}_{i1}, \tilde{d}_{i2}$ and $\tilde{d}_{i3}$ are the centred distance components.

Equation 13 represents the cumulative distance contribution to modal utility. The spline coefficients are marginal utility slopes within successive distance intervals and should not be interpreted as separate level effects. At any given distance, the implied effect is obtained by summing the contributions of all active spline segments. The function is continuous at the 150 km and 300 km thresholds, although its slope may change at these points. The contribution of distance to the utility is therefore allowed to have a different slope in each distance range and for each mode. For example, the marginal distance effect for inland waterways below 150 km is governed by $\beta_{D,IWW,1}$, between 150 and 300 km by $\beta_{D,IWW,2}$, and above 300 km by $\beta_{D,IWW,3}$. This is more flexible than a single linear distance coefficient and allows rail and inland waterways to become more or less competitive as distance increases.

Classically, for each OD-commodity cell, the probability of mode m for OD-cell i is

$$P_{m,i} = \frac{\exp(V_{m,i})}{\sum_{k \in A_i} \exp(V_{k,i})} \tag{14}$$

where A_i is the set of available alternatives. Predicted modal tonnes are obtained by multiplying probabilities by total tonnes:

$$\hat{q}_{m,i} = P_{m,i} Q_i \tag{15}$$

The inland-waterway utility is specified as:

$$\begin{aligned} V_{IWW,i} = ASC_{IWW,g} + \beta_{C,IWW}\tilde{c}_{IWW,i} \\ + \beta_{T,IWW}\tilde{t}_{IWW,i} + f_{IWW}(d_i) \end{aligned} \tag{16}$$

Here $ASC_{IWW,g}$ is a commodity-group g constant, $\tilde{c}_{IWW}$ is the relative centred log cost ratio, $\tilde{t}_{IWW}$ is the relative centred log moving-time ratio and $f_{IWW}(d)$ is the inland-waterway distance function.

The retained rail utility is:

$$V_{rail,i} = ASC_{rail,g} + \beta_{T,rail}\tilde{t}_{rail,i} + f_{rail}(d_i) + \gamma I_{IWW,i} \tag{17}$$

[3] A sensitivity analysis is provided in Sect. 6.2.

In this specification, $I_{IWW,i}$ is the rail market proxy: it is equal to one when inland waterways are available for the OD cell and zero otherwise. It is included as a market-structure proxy. OD pairs with inland-waterway availability often belong to river corridors, port hinterlands or industrial regions where non-road freight options are better developed. The coefficient should not be interpreted as a causal effect of inland waterways on rail, but as a correction for unobserved multimodal market conditions.

This model architecture is common to all datasets, but coefficients are estimated separately. This choice reflects differences in zoning scale, distance distributions, availability patterns and modal markets.

Note that a rail-cost term was tested during model development. With this term, the rail utility becomes:

$$V_{rail,i} = ASC_{rail,g} + \beta_{C,rail}\tilde{c}_{rail,i} + \beta_{T,rail}\tilde{t}_{rail,i} + f_{rail}(d_i) + \gamma I_{IWW,i} \quad (18)$$

where $\tilde{c}_{rail,i}$ is the centred rail cost ratio relative to road.

The rail-cost specification was estimated as a sensitivity test but was not retained in the reference model. This decision does not imply that rail prices are irrelevant to freight modal choice. It means that the constructed rail-cost ratio does not provide a stable independent signal once commodity constants, moving time, distance splines and market structure are controlled for. Since cost, time and distance are all derived from the same network representation, substantial collinearity remains possible. The coefficient stability and pre- and post-calibration performance of the two specifications are compared in Sect. 6.3.

The behavioural model is estimated from aggregate modal shares rather than from individual shipment choices. For each OD–commodity observation i, the contribution to the objective function is

$$\ell_i(\beta) = w_i \sum_m s_{m,i} \log P_{m,i}(\beta), \quad (19)$$

where $\ell_i(\beta)$ is the contribution of OD-commodity observation i to the weighted log-likelihood, β denotes the vector of coefficients to be estimated, $s_{m,i}$ is the observed share of mode m, $P_{m,i}$ is the corresponding model probability and w_i is the row weight.

In the reference specification, the weight is the total transported quantity in the row, $w_i = Q_i$. Since $Q_i s_{m,i} = q_{m,i}$, the weighted objective can also be written as

$$l(\beta) = \sum_i \sum_m q_{m,i} \log P_{m,i}(\beta). \quad (20)$$

The modal tonnes therefore enter the estimation algebraically as frequency weights. Equivalently, the estimator minimises the discrepancy, measured through weighted cross-entropy, between the observed and predicted modal shares, with greater importance given to OD–commodity cells carrying larger freight volumes.

This does not mean that each tonne is treated as a separate observed shipment decision in a behavioural sense. Tonnes within the same OD–commodity cell share the same explanatory variables, and the data contain no information on the number, size or organisation of individual shipments. The likelihood should therefore be regarded as a volume-weighted aggregate or pseudo-likelihood rather than as a conventional individual-choice likelihood.

The estimated coefficients describe systematic, volume-weighted associations between modal attributes and the allocation of tonnes across modes. Their signs and relative magnitudes remain useful for characterising the aggregate modal structure, but they should not be interpreted as individual shipper preference parameters. They may also reflect aggregation, omitted logistics characteristics, the construction of the network attributes and differences in effective modal availability.

Alternative weighting functions, including square-root and power weights, were examined during model development. Such weights reduce the dominance of the largest OD cells and may be preferable when the primary objective is behavioural interpretation. Raw tonnage weights were retained here because the operational objective gives particular importance to reproducing the modal allocation of large freight flows.

4 Base-year calibration layer

The behavioural logit model does not, structurally, reproduce the non-road matrices. Whenever a non-road alternative is available, the logit model assigns it a positive probability. If the OD-cell total flow is large, even a small probability can generate noticeable predicted tonnes in a cell where the observed non-road flow is zero. At the same time, the model can underpredict OD-commodity cells where rail or inland waterways are strongly established. To address this, utility correction constants are added to the non-road utilities:

$$V^{cal}_{IWW,i} = V_{IWW,i} + CAL_{IWW,c(i)} \quad (21)$$

$$V^{cal}_{rail,i} = V_{rail,i} + CAL_{rail,c(i)} \quad (22)$$

Calibrated probabilities are computed with the same logit formula:

$$P_{m,i}^{cal} = \frac{\exp\left(V_{m,i}^{cal}\right)}{\sum_{k\in A_i}\exp\left(V_{k,i}^{cal}\right)} \tag{23}$$

and calibrated modal tonnes are

$$\hat{q}_{m,i}^{cal} = P_{m,i}^{cal} Q_i \tag{24}$$

The calibration of cell $c(i)$ can be defined at several levels: mode, mode-group, mode-distance, mode-group-distance, mode-IWW-availability, mode-group-IWW-availability, mode-OD and mode-OD-group. The operational specification uses the "mode-OD-group" level, which varies by mode, OD pair and commodity group.

Each observation corresponds to one OD–commodity row. The implementation stores one bounded calibration constant for IWW and one for rail for every such row. The Europe dataset therefore contains 261,144 stored calibration constants for 130,572 observations, Benelux+ contains 657,328 constants for 328,664 observations, and Germany contains 520,774 constants for 260,387 observations. Across the three datasets, this represents 1,439,246 stored constants for 719,623 observations, or two constants per observation.

These constants are not additional behavioural parameters estimated through the likelihood function. They are bounded residual corrections derived iteratively from the same base-year modal matrices. Moreover, a constant associated with an unavailable mode is inactive because that alternative does not enter the logit denominator.

Nevertheless, for the available non-road alternatives, the detailed calibration operates at essentially the same OD–commodity level as the observations. Very low in-sample WAPE values are therefore expected and measure base-year matrix reconstruction rather than predictive accuracy.

For a calibration cell c and mode m, observed and calibrated predicted tonnes are:

$$Q_{m,c}^{obs} = \sum\nolimits_{i\in c} q_{m,i} \; and \; Q_{m,c}^{cal} = \sum_{i\in c} \hat{q}_{m,i}^{cal} \tag{25}$$

The calibration update is based on the sign and magnitude of the gap between observed and predicted tonnes within each cell. For a non-road mode m and cell c, the current prediction is compared with the observed tonnage through the log-ratio:

$$\Delta_{m,c} = \ln\left(\frac{Q_{m,c}^{obs}+\epsilon}{Q_{m,c}^{cal}+\epsilon}\right). \tag{26}$$

where $\epsilon > 0$ is a small constant used to avoid division by zero.

This quantity gives the direction in which the utility correction should move. If the model underpredicts a cell, then $Q_{m,c}^{obs} > Q_{m,c}^{cal}$, the ratio is greater than one, and $\Delta_{m,c}$ is positive. The calibration constant is therefore increased, which raises the utility of that mode in that cell and increases its predicted probability. Conversely, if the model overpredicts a cell, the ratio is below one, $\Delta_{m,c}$ is negative, and the constant is reduced. The actual update equation can be written as:

$$CAL_{m,c}^{new} = \text{clip}\left(CAL_{m,c}^{old} + \lambda\, \text{clip}(\Delta_{m,c}, -s_{\max}, s_{\max}), -K, K\right) \tag{27}$$

in which λ is the damping factor, $s_{\max}$ is the maximum update step allowed at each iteration, and K is the maximum absolute value allowed for the constant. Calibration constants are indeed bounded to avoid extreme utility corrections and uncontrolled overfitting. The cap keeps the calibration stable and interpretable, while still allowing large base-year adjustments.

The logarithmic form is useful because the utility correction constants are additive in utility but multiplicative in odds. Adding a constant $CAL_{m,c}$ to a utility multiplies the odds of choosing that mode by $\exp(CAL_{m,c})$. Therefore, a log-ratio between observed and predicted tonnes provides a natural update direction in utility space.

The update is damped[4] and clipped because changing one modal utility also changes the probabilities of the other modes through the logit denominator. The process therefore proceeds iteratively: constants are adjusted, probabilities are recomputed, new discrepancies are measured, and the loop continues until the updates become small or the maximum number of iterations is reached.

The calibration constants are obtained through an iterative proportional correction in utility space. For each calibration cell c and each calibrated non-road mode m, observed and predicted tonnes are compared after each iteration. The predicted tonnes are computed as the sum of the modal probabilities multiplied by the OD-commodity tonnage over all observations belonging to the same calibration cell. The update is applied only to IWW and rail; the road utility remains normalized to zero.

Let $Q_{m,c}^{\text{obs}}$ be the observed tonnes for mode m in calibration cell c, and $Q_{m,c}^{\text{pred},(r)}$ the corresponding predicted tonnes at iteration r. The raw update is[5]:

$$u_{m,c}^{(r)} = \ln\left(\frac{Q_{m,c}^{\text{obs}}+\epsilon}{Q_{m,c}^{\text{pred},(r)}+\epsilon}\right) \tag{28}$$

[4] The damping factor reduces the size of each calibration update. It is used to avoid abrupt changes in utility constants and to make the iterative calibration more stable.

[5] Eqs. 28–30 provide the iteration-indexed form of the compact calibration update in Eqs. 26–27.

The update is first clipped to the interval $[-s_{\max}, +s_{\max}]$:

$$\tilde{u}_{m,c}^{(r)} = \max\left[-s_{\max}, \min\left(s_{\max}, u_{m,c}^{(r)}\right)\right] \quad (29)$$

The calibration constant is then updated with a damping factor and bounded to the interval $[-K, +K]$:

$$\mathrm{CAL}_{m,c}^{(r+1)} = \max\left[-K, \min\left(K, \mathrm{CAL}_{m,c}^{(r)} + \lambda\, \tilde{u}_{m,c}^{(r)}\right)\right] \quad (30)$$

If both observed and predicted tons are effectively zero, the update is set to zero. The constants therefore increase when a mode is underpredicted in a calibration cell and decrease when it is overpredicted. The numerical settings used for the reported base-year calibration are given in Appendix A.

In the reported base-year runs, the detailed "mode-od-group" calibration stopped after 80 iterations rather than by the tolerance criterion. This is mainly due to sparse OD-commodity cells for which non-road alternatives are technically available but have zero or very small observed flows. These cells can continue to request clipped negative updates even when the aggregate calibration error has already become very small. For this reason, the maximum number of iterations and the bound on the constants are part of the calibration design. They prevent an unconstrained search for exact reproduction in sparse cells while still allowing the calibration layer to restore the main base-year structure of the modal OD matrices.

It is important to be aware that the constants are not behavioural coefficients. A positive constant means that the behavioural model underpredicts the corresponding flow in the cell; a negative constant means it overpredicts it. Large negative constants often identify cells where a mode is technically available but absent from the observed matrix. In this sense, the second layer restores matrix sparsity and absorbs OD, commodity, terminal, corridor and logistics effects that are not directly observed.

Both the behavioural layer and the (optional) base-year calibration are implemented in a single Python script, in which the model parameters were estimated using Biogeme [7], an open-source Python package designed for maximum-likelihood estimation of discrete-choice models.

5 Evaluation metrics

Model performance is evaluated at the modal OD-cell level. The resulting metrics are used for different purposes: in the detailed calibration they measure in-sample base-year reconstruction, whereas in the holdout experiment they provide evidence on performance for OD pairs not used during estimation and calibration. Predicted aggregate tonnes are computed as:

$$Q_m^{obs} = \sum\nolimits_i q_{m,i}, \quad Q_m^{pred} = \sum\nolimits_i \hat{q}_{m,i} \quad (31)$$

And the corresponding modal shares are:

$$S_m^{obs} = \frac{Q_m^{obs}}{\sum_k Q_k^{obs}}, \quad S_m^{pred} = \frac{Q_m^{pred}}{\sum_k Q_k^{pred}} \quad (32)$$

At the OD-cell level, the prediction error is

$$e_{m,i} = \hat{q}_{m,i} - q_{m,i} \quad (33)$$

The main OD-cell metric is WAPE (Eq. 34). WAPE is preferred to MAPE because modal freight matrices contain many zero or very small non-road cells. By summing absolute errors before dividing by total observed tonnes, WAPE remains well-defined and interpretable for sparse matrices.

$$\mathrm{WAPE}_m = \frac{\sum_i |\hat{q}_{m,i} - q_{m,i}|}{\sum_i q_{m,i}} \quad (34)$$

The following bias is also reported:

$$B_m = \sum_i (\hat{q}_{m,i} - q_{m,i}), \qquad B_m^{\%} = \frac{\sum_i (\hat{q}_{m,i} - q_{m,i})}{\sum_i q_{m,i}} \quad (35)$$

It captures the signed difference between predicted and observed tonnes and serves as a diagnostic indicator of directional bias, showing whether the model tends to over- or under-predict modal tonnes after calibration.

Calibration-constant distributions are used as an additional diagnostic. The unweighted distribution shows how many cells require correction. The weighted distribution shows where observed modal tonnes are located in that distribution. This distinction is useful because a large number of negative constants may simply correspond to empty non-road cells.

Finally, an OD-pair holdout experiment is performed and repeated three times using the random seeds 12,345, 1234, and 123. For each dataset and repetition, the unique directional OD pairs, defined by the origin and destination identifiers, are first identified. Twenty per cent of these OD pairs are then selected randomly without replacement for the test sample, while the remaining 80% form the training sample. All commodity-group observations associated with a given OD pair are assigned to the same sample, thereby preventing observations for the same OD relation from appearing in both training and test data.

The split is performed on the raw data before preprocessing and estimation. The centring constants, fallback

values used for unavailable alternatives, behavioural coefficients and mode–group–distance calibration constants are obtained exclusively from the training sample. The held-out observations are subsequently transformed using the training-sample quantities and are used only for evaluation.

6 Results and discussion

6.1 Behavioural layer

The coefficients in Table 2 describe the behavioural component of the model before the base-year calibration constants are added. The table is deliberately limited to the main cost, moving-time, distance and market-structure parameters in order to facilitate comparison across datasets. The corresponding estimation diagnostics, including robust standard errors, t-statistics, p-values, confidence intervals and active-bound information, are reported in Appendix B.

The statistical results should be interpreted in light of the aggregate and volume-weighted nature of the estimation. The robust standard errors and confidence intervals reported in Appendix B are calculated for the weighted pseudo-likelihood. Because annual tonnes enter as frequency weights, the numerical information content implied by the estimation criterion is very large, producing extremely small standard errors and very large t-statistics. These values are useful for detecting numerical instability and active bounds, but the associated p-values and confidence intervals should not be interpreted as shipment-level statistical significance or as classical sampling uncertainty based on independent choices. They are conditional diagnostics of the weighted aggregate pseudo-likelihood. In the discussion below, greater emphasis is therefore placed on coefficient signs, active bounds, stability across datasets and sensitivity results than on statistical significance alone.

Commodity-group constants are also an important part of the behavioural specification. They account for differences in the underlying suitability of each commodity group for inland waterways and rail after controlling for the other variables in the utility functions. The complete set of commodity-group constants and their implied odds multipliers relative to road are reported in Appendix C. The results show a clear and consistent commodity structure. Solid mineral fuels have the most favourable constants for both non-road modes, while fertilisers have a particularly favourable rail effect. Most other commodity groups have negative constants, although their magnitude varies considerably across datasets.

The most stable result is the negative coefficient of the IWW cost ratio. This coefficient has the expected sign in all three datasets: when inland-waterway cost increases relative to road, the utility of inland waterways decreases. Because the variable is a logarithmic ratio, the coefficient can be read in proportional terms. A 10% increase in relative IWW cost reduces the IWW odds by roughly 30% in the Europe dataset, 35% in the Benelux+ dataset and 18% in the Germany dataset, all else equal. Among the behavioural variables, this is the clearest and most consistent cost signal.

The moving-time coefficients are less stable. In the Europe NUTS2 dataset, both IWW and rail moving-time effects are negative, as expected, with a stronger effect for rail. In the two NUTS3 datasets, however, the moving-time coefficients reach the zero bound. This should not be read as evidence that transport time is irrelevant for freight decisions. A more plausible interpretation is that, at the finer spatial resolution, moving time does not add much independent information once relative costs, distance splines, commodity constants and availability structure are already included. Moving time is strongly correlated with distance and network geography, and its separate contribution is therefore difficult to identify in

Table 2 Key behavioural coefficients across datasets

Estimator	Europe	Benelux+	Germany	Interpretation
B_COST_IWW	-3.746	-4.507	-2.023	IWW cost effect relative to road
B_MOVE_IWW	-0.508	0.000	0.000	IWW moving-time effect
B_MOVE_RAIL	-1.928	0.000	0.000	Rail moving-time effect
B_RAIL_IWW_AVAIL	0.599	0.212	0.298	Rail correction for IWW-availability
B_DIST_IWW_1	-0.942	1.031	0.946	
B_DIST_IWW_2	0.439	0.153	1.443	IWW distance profile
B_DIST_IWW_3	0.033	-0.021	2.134	
B_DIST_RAIL_1	-0.269	0.302	0.251	
B_DIST_RAIL_2	0.934	1.356	2.221	Rail distance profile
B_DIST_RAIL_3	0.046	0.123	2.695	

Table 3 Main results of the behavioural layer

Dataset	Road			IWW			Rail		
	obs. %	est. %	WAPE	obs. %	est. %	WAPE	obs. %	est. %	WAPE
Europe	80.01	80.01	0.233	6.18	6.18	1.176	13.81	13.81	1.088
Benelux+	79.72	79.72	0.219	11.94	11.94	1.039	8.34	8.34	1.151
Germany	85.46	85.46	0.102	2.45	2.45	1.375	12.10	12.10	0.627

these matrices. Since the moving-time coefficients were constrained to be non-positive, an estimate equal to zero means that the optimum is reached at the upper bound. It should therefore be interpreted as an absence of an additional identifiable negative moving-time effect in that dataset, not as evidence that transport time has no practical relevance.

The distance coefficients are important, but they are clearly dataset-specific. In the European dataset, the first IWW distance segment is negative, while the second and the third are positive, suggesting that, at the broad European level, inland waterways are relatively disadvantaged for shorter movements. Rail shows a similar pattern, with a negative short-distance effect and a stronger positive effect between 150 and 300 km. In the Benelux + and Germany datasets, the distance coefficients are mostly positive, and particularly large in Germany for longer-distance segments. This confirms that distance does not act as a simple linear variable but that its effect depends on spatial considerations.

The positive coefficient associated with IWW availability in the rail utility is also informative. Its value is largest in the European dataset and smaller, but still positive, in the two NUTS3 datasets. In odds terms, the coefficient increases rail odds by a factor of about 1.82 in the Europe dataset, 1.24 in the Benelux+ dataset and 1.35 in the Germany dataset. As discussed in Sect. 3, it acts as a market-structure indicator. The variable therefore helps rail utility account for unobserved multimodal corridor conditions.

Overall, Table 2 supports the modelling strategy presented in this paper: the same specification can be applied to the three datasets, but the estimated coefficients should remain dataset specific. The IWW cost effect is robust, whereas moving time and distance effects depend more strongly on the spatial scale and regional freight transport network structure.

The pre-calibration results presented in Table 3 show a clear contrast: at the aggregate level, the behavioural model already reproduces modal shares quite well but, at OD-cell level, the WAPE values remain high, especially for rail and inland waterways.

To assess whether the richer behavioural specification improves matrix reproduction, it was compared with a simpler model estimated within exactly the same framework. The benchmark model follows a conventional logarithmic cost-and-time specification. For mode m, observation i and commodity group g, the underlying systematic utility is

$$V_{m,i} = \alpha_{m,g} + \beta_C \ln C_{m,i} + \beta_{T,m} \ln T_{m,i} \tag{36}$$

where $V_{m,i}$ is the systematic utility of mode m for observation i, $\alpha_{m,g}$ is the commodity-group-specific constant for mode m and commodity group g, β_C is the logarithmic transport-cost coefficient common to all modes, and $\beta_{T,m}$ is the mode-specific logarithmic moving-time coefficient. Road is used as the reference alternative, and its utility is normalised to zero when the model is expressed in relative form.

Table 4 WAPE comparison between the conventional cost-and-time model and the preferred behavioural specification

Dataset	Mode	Conventional cost–time model	Preferred behavioural model	Relative reduction (%)
Europe	Road	0.241	0.233	3.4
	IWW	1.213	1.176	3.1
	Rail	1.085	1.088	−0.2
Benelux+	Road	0.233	0.219	6.2
	IWW	1.099	1.039	5.4
	Rail	1.191	1.151	3.3
Germany	Road	0.128	0.102	20.4
	IWW	1.450	1.375	5.2
	Rail	0.775	0.627	19.1

The benchmark uses the same observations, availability rules, annual-tonnage weights, likelihood formulation and evaluation metrics as the reference model. Differences in WAPE (Table 4) can therefore be attributed to the utility specification rather than to differences in data processing or evaluation[6].

The reference specification produces lower OD-cell WAPE in eight of the nine dataset–mode comparisons. The improvements are moderate in most cases, ranging from approximately 3% to 6%, but are substantially larger for road and rail in Germany, where WAPE is reduced by approximately 20%. European rail is the only exception. In that case, WAPE changes from 1.085 under the conventional specification to 1.088 under the preferred specification, a difference of only 0.2% that should be regarded as negligible.

[6] With this simple specification, the estimated IWW and/or Rail time parameters are equal to zero (bounded) for both NUTS 3 models.

The comparison indicates that logarithmic cost and moving time alone do not capture all the systematic spatial variation in the modal matrices. The mode-specific distance profiles and the rail market-structure indicator generally add explanatory information, although their contribution varies across modes and datasets. The results do not imply that every additional term improves every individual case.

The comparison also confirms that the richer behavioural specification remains insufficient to reproduce the sparse non-road OD matrices closely. Even when it improves on the conventional benchmark, pre-calibration IWW and rail WAPE remains high. This provides further motivation for treating the bounded calibration layer as a separate base-year matrix-reconstruction component rather than expecting the behavioural model alone to reproduce the detailed geography of non-road flows.

6.2 Sensitivity to the distance-spline thresholds

As explained in Sect. 3, the thresholds of 150 and 300 km were selected ex ante to provide a common and interpretable distinction between relatively short, intermediate and longer-distance freight movements.

To examine the sensitivity of the results to this modelling choice, the behavioural model was re-estimated with lower thresholds of 100 and 250 km and higher thresholds of 200 and 400 km. All other elements were unchanged, including the observations, technical-availability rules, annual-tonnage weights, explanatory variables and coefficient constraints. Table 5 reports pre-calibration OD-cell WAPE, since this is the stage at which changes in the behavioural distance specification should be evaluated.

The WAPE results are relatively insensitive to the location of the knots. Across the nine dataset–mode combinations, the largest difference between specifications is approximately 0.031, for European rail. In relative terms, the complete range remains below approximately 3% of the reference WAPE in every case. No alternative threshold pair consistently dominates the others. The higher thresholds improve some European results, whereas the reference thresholds perform slightly better for several Benelux + and German results.

Table 5 Pre-calibration OD-cell WAPE under alternative distance-spline thresholds

Dataset	Mode	100–250 km	150–300 km	200–400km
Europe	Road	0.236	0.233	0.230
	IWW	1.187	1.176	1.190
	Rail	1.085	1.088	1.057
Benelux+	Road	0.224	0.219	0.223
	IWW	1.040	1.039	1.038
	Rail	1.160	1.151	1.154
Germany	Road	0.103	0.102	0.104
	IWW	1.373	1.375	1.373
	Rail	0.631	0.627	0.633

The main coefficient conclusions are also preserved. The IWW cost coefficient remains negative in every dataset, and the rail market-structure coefficient remains positive. The European moving-time coefficients remain negative, while both German moving-time coefficients remain at their zero bounds. The main exception concerns IWW moving time in Benelux+, which is negative with the 100–250 km thresholds but reaches its zero bound with the reference and higher thresholds. This indicates some sensitivity in the allocation of explanatory power between moving time and distance, rather than a material change in matrix-reproduction performance.

The 150 and 300 km thresholds are therefore retained as an interpretable common segmentation. The sensitivity analysis does not indicate that they represent uniquely optimal or precisely estimated economic breakpoints, but it shows that the main behavioural and matrix-reproduction conclusions do not depend materially on reasonable changes in their location.

6.3 Sensitivity to the rail-cost term

Table 6 compares the reference specification with an otherwise identical model that includes the centred logarithm of rail cost relative to road. The comparison is made before and after the detailed base-year calibration.

Table 6 Sensitivity of the rail results to the inclusion of rail cost

Dataset	Rail cost	Rail-cost coefficient	Pre-calibration rail WAPE	Pre-calibration rail bias (%)	Post-calibration rail WAPE	Rail constants at −8 (%)
Europe	No	-	1.088	|bias| < 0.01	0.006	85.54
Europe	Yes	−0.2856	1.073	|bias| < 0.01	0.006	85.54
Benelux+	No	-	1.151	|bias| < 0.01	0.011	89.07
Benelux+	Yes	0.0000	1.151	|bias| < 0.01	0.011	89.07
Germany	No	-	0.627	|bias| < 0.01	0.009	80.13
Germany	Yes	0.0000	0.627	|bias| < 0.01	0.009	80.13

The rail-cost coefficient was constrained to be non-positive. An estimate of zero therefore indicates that the optimum is located at the upper bound. Bias is the difference between predicted and observed aggregate rail tonnes divided by observed rail tonnes. The calibration results use the "mode–OD–group" level and an absolute constant bound of 8

The pre-calibration results are particularly relevant because this is where an additional behavioural variable should improve the explanatory structure of the model.

The coefficient provides some additional information in the Europe NUTS2 dataset, where it is negative and estimated away from its bound. The weighted log-likelihood improves slightly, from approximately -2.228334×10^9 to -2.228260×10^9. Pre-calibration rail WAPE also decreases from 1.088 to 1.073, a modest reduction of approximately 1.4%. This improvement is limited to the European dataset and does not persist after detailed calibration, for which rail WAPE remains unchanged at the reported precision.

More importantly, the coefficient reaches its zero bound in both NUTS3 datasets. The variable therefore does not provide a stable cost response across the three empirical settings. This instability may reflect the construction of the cost variable and its correlation with distance, moving time and network geography, rather than an absence of a genuine price effect in rail freight markets.

Including rail cost has virtually no effect on aggregate bias, post-calibration WAPE or the proportion of rail calibration constants at the negative bound. It therefore does not improve post-calibration OD-matrix reproduction or reduce the burden placed on the calibration layer. Effects on the road and IWW results are similarly negligible.

On this basis, the more parsimonious specification without rail cost is retained as the common reference model. The rail-cost version is treated as a sensitivity test showing that the main matrix-reproduction results do not depend on its exclusion. The model should consequently not be interpreted as providing a transferable rail-price elasticity; applications focused specifically on rail-pricing policies would require additional data, external elasticity evidence or a dedicated re-estimation.

6.4 Base-year calibration layer

After calibration, the WAPE values fall sharply for all three datasets and modes, as shown in Table 7. This demonstrates that the detailed "mode-OD-group" correction is effective at restoring the base-year geography and sparsity of the modal matrices.

These results are in-sample base-year reconstruction statistics rather than measures of predictive performance. The calibration stores two bounded constants per OD–commodity observation, one for each non-road mode, although constants associated with unavailable alternatives are inactive. The correction is therefore defined at almost the same level as the observed matrix. The very low post-calibration WAPE values are consequently an expected result of the detailed calibration design and should be interpreted as evidence of successful in-sample base-year reconstruction.

The purpose of Table 7 is to assess whether the framework can generate assignment-ready base-year matrices. It does not show that the detailed constants can be transferred unchanged to new OD pairs, another base year or a policy scenario.

Transferability is examined separately in Sect. 6.6 through repeated OD-pair holdout experiments, where OD-specific constants cannot be used and a coarser "mode-group-distance" calibration is applied.

The bias values add another useful check. While WAPE measures the size of the OD-cell errors, bias indicates whether a mode is still over- or under-predicted overall. After calibration, the remaining biases are very small. The slight tendency is for road to be marginally over-predicted and for the two non-road modes to be marginally under-predicted, but the differences are negligible compared with the observed modal tonnes. This suggests that the calibration layer of the model has removed almost all systematic modal imbalances, while still leaving the calibration bounded and transparent. The use of bounded constants, the analysis of their distribution, and the separate holdout test help clarify this distinction and provide an assessment of transferability.

6.5 Calibration constants

The distribution of calibration constants (Table 8) is useful because it shows what the layer is doing. At the detailed mode–OD–group level, the implementation stores one calibration entry for each non-road mode and each retained OD–commodity observation. The number reported in the table therefore equals the number of observations for both inland waterways and rail. This is a bookkeeping convention and should not be interpreted as indicating that both modes are technically available or actively calibrated in every observation. When a non-road mode is unavailable, its observed and predicted tonnes are both zero, no calibration update is applied,

Table 7 In-sample base-year OD-matrix reproduction after detailed calibration

Dataset	Road			IWW			Rail		
	WAPE		Bias (%)	WAPE		Bias (%)	WAPE		Bias (%)
	pre-calib	post-calib		pre-calib	post-calib		pre-calib	post-calib	
Europe	0.233	0.001	0.13	1.176	0.005	-0.44	1.088	0.006	-0.56
Benelux+	0.219	0.002	0.20	1.039	0.007	-0.62	1.151	0.011	-1.06
Germany	0.102	0.001	0.13	1.375	0.005	-0.48	0.627	0.009	-0.80

Table 8 Calibration-constant distribution by dataset and mode

Dataset	Mode	# stored entries	Unweighted			Weighted		
			median	% at -8	% at +8	median	% at -8	% at +8
Europe	IWW	130,572	0.00	18.19	0.00	2.61	0.00	0.00
	rail	130,572	-8.00	85.54	0.01	2.44	0.00	0.01
Benelux+	IWW	328,664	0.00	43.94	0.00	2.63	0.00	0.00
	rail	328,664	-8.00	89.07	0.01	3.24	0.00	0.07
Germany	IWW	260,387	0.00	40.64	0.04	2.62	0.00	0.13
	rail	260,387	-8.00	80.13	0.03	2.52	0.00	0.10

The count includes one stored mode–OD–group entry for each non-road mode and retained OD–commodity observation. Entries corresponding to technically unavailable alternatives are inactive and do not affect utilities or predicted modal probabilities

and the corresponding stored entry has no effect on the utility or modal probabilities. Only entries associated with technically available alternatives can generate an effective calibration correction. The first results shown in the table are the strong contrast between the unweighted and weighted distributions. If all cells are counted equally, many constants are very negative, especially for rail. This pattern reflects the sparsity of non-road matrices: many OD-commodity cells are technically available for rail or inland waterways but carry no observed flow. The effect is strongest for rail, where 80–89% of constants are at the lower bound, and less pronounced for IWW, where the corresponding share ranges from about 18% to 44%.

The high proportion of rail constants at the lower bound should be interpreted in relation to the distinction between technical and effective availability. Rail is technically available in almost all retained observations because a network path can be constructed, but the observed matrices indicate that it is used in only a much smaller subset of OD–commodity relations. The calibration therefore assigns very negative residual utilities to many cells in which the behavioural logit model would otherwise produce a small positive rail probability.

These lower-bound constants should not be interpreted as estimated availability parameters or as proof that the corresponding alternatives are structurally impossible. They indicate that, given the available explanatory variables, the base-year matrix contains no or negligible non-road flow in those cells. The calibration layer consequently approximates several mechanisms that the data do not distinguish, including effective choice-set exclusion, unobserved logistics constraints and realised zero demand.

A two-stage or hurdle specification would provide a more explicit representation. In such a model, a first component would estimate whether a non-road mode enters the effective choice set or carries positive flow, and a second component would model the conditional modal share or tonnage when it does. A structural-zero or zero-inflated approach could similarly distinguish cells that are effectively excluded from cells that are technically available but happen to record zero flow.

Such approaches would be attractive conceptually, but their identification would require additional information or repeated observations. With a single aggregate base-year matrix, an observed zero cannot reliably be classified as structural or incidental. Estimating a participation model from the same zero pattern that is later reproduced could also shift, rather than resolve, the risk of overfitting. For the present assignment-oriented application, the bounded utility correction is retained as a transparent operational approximation. Its role is limited to base-year reconstruction, and the lower-bound constants are not given a behavioural or causal interpretation.

Weighted by observed modal tonnes, almost no traffic is associated with constants at the negative cap. The many constants equal to the cap of -8 therefore mainly correspond to empty or negligible OD-commodity cells, rather than to suppressed observed non-road flows. They are not suppressing large observed non-road flows; they mostly remove the small positive probabilities that the logit model would otherwise assign to unused alternatives.

For the cells that actually carry non-road tonnes, the utility corrections are generally positive. The weighted medians are between about 2.4 and 3.2 for both IWW and rail. This means that observed non-road flows tend to occur in cells where the behavioural model alone underestimates the attractiveness of the corresponding mode. The second layer of the model therefore reinforces established non-road corridors and commodity flows, while reducing spurious traffic in unused cells.

It is also worth noting that very little observed tonnage is associated with constants at the positive cap. The weighted percentage at +8 remains close to zero in all cases. This is reassuring because it suggests that the calibration layer is not driven by a small number of extreme positive corrections. The calibration constants are often strong, but they are not simply forcing large flows through the upper bound.

The detailed calibration naturally raises concerns about overfitting. This concern is legitimate, since corrections

defined by OD pair, mode, and commodity group may reproduce a substantial part of the base-year matrix structure. For this reason, the calibrated fit should not be interpreted as evidence that the behavioural model alone can predict the detailed geography of non-road freight. It should rather be seen as a constrained base-year matrix-reproduction mechanism designed to support assignment.

6.6 OD-pair holdout and transferability diagnostic

The holdout exercise is used here as a diagnostic test rather than as a full cross-validation design. As expected, the holdout results (Table 9) give a more cautious view of the model than the base-year fitted results. In the former case, the detailed "mode-OD-group" constants are allowed to correct each observed OD-commodity cell. In the holdout test, this is no longer possible, because 20% of OD pairs are deliberately kept outside the sample. The calibration therefore has to rely on a coarser "mode-group-distance" level, which can be applied to unseen OD pairs but cannot capture OD-specific corridor effects.

The first result is that the post-calibration WAPE is lower than the pre-calibration one in all three datasets and for all three modes in this holdout split. The improvement is not dramatic. This suggests that the coarser calibration may capture some transferable structure, especially differences by mode, commodity group and distance range. In that sense, the second layer does more than reproduce individual OD pairs; part of it reflects more general regularities in the matrices.

The size of the improvement, however, remains limited. For Europe, IWW WAPE decreases from 1.451 to 1.382 and rail WAPE from 1.056 to 1.032. For Benelux+, the reduction is somewhat clearer, especially for rail, where WAPE falls from 1.102 to 1.000. Germany shows only small changes, particularly for road, where the WAPE is already low and remains almost unchanged. These values indicate that the model can improve predictions on unseen OD pairs, but not nearly as strongly as in the detailed base-year calibration.

The mode-specific patterns are also informative. Road has much lower WAPE than rail and inland waterways in all datasets. This reflects the fact that the road matrix is much denser and dominant in volume terms.

To examine the sensitivity of the results to the random partition, the OD-pair holdout experiment was repeated three times using the random seeds 12,345, 1234 and 123. As described in Sect. 5, the split was performed at the level of directional OD pairs, with all commodity-group observations associated with a given OD pair kept in the same sample. Each repetition used 80% of the OD pairs for training and 20% for testing.

The random OD-pair holdout remains an interpolation-oriented test. Training and test observations are drawn from the same dataset and therefore share the same regional system, network representation and broad corridor structure. The exercise tests whether the behavioural model and the coarser calibration can be applied to OD pairs that were not used during estimation, but it does not establish transferability to an entirely unseen region, port hinterland or freight corridor.

A spatially blocked or corridor-based holdout would provide a more demanding extrapolation test. Such an experiment would require a defensible and comparable definition of spatial or functional blocks across the European NUTS2, Benelux+ NUTS3 and German NUTS3 datasets. It would also be sensitive to the sparse concentration of rail and inland-waterway flows, since withholding one major corridor could remove a substantial share of the observed non-road market from the training sample. The present random holdout is therefore retained as a first transferability diagnostic, while spatially blocked validation is identified as a useful extension of the framework.

Table 10 reports, for each dataset and mode, the mean pre-calibration WAPE, the mean post-calibration WAPE, the sample standard deviation of the post-calibration WAPE and its range across the three runs.

The three repeated random holdouts show the same directional pattern as the initial split. Across the three datasets, three modes and three repetitions, thus 27 comparisons in total, post-calibration WAPE is lower than the corresponding pre-calibration WAPE in every comparison. Given the limited number of repetitions, these results should be interpreted as repeated random holdout diagnostics rather than as evidence of statistical stability or robustness. The observed gains are generally modest and vary across splits. Road remains the easiest mode to reproduce, with low WAPE and limited variation across splits. Rail and inland waterways show higher WAPE values and larger variation, especially where observed flows are sparse and concentrated in a small number of

Table 9 Holdout OD-cell WAPE using a "mode-group-distance" calibration

Dataset	IWW		Rail		Road	
	pre-calib	post-calib	pre-calib	post-calib	pre-calib	post-calib
Europe	1.451	1.382	1.056	1.032	0.254	0.250
Benelux+	1.049	0.991	1.102	1.000	0.223	0.215
Germany	1.363	1.340	0.633	0.624	0.096	0.096

Table 10 Holdout OD-cell WAPE across three random OD-pair splits

Dataset	Mode	Mean pre-calib. WAPE	Mean post-calib. WAPE	Std. dev. post-calib.	Range post-calib.	Mean reduction
Europe	IWW	1.200	1.165	0.195	1.005–1.382	0.035
Europe	rail	1.086	1.047	0.065	0.991–1.118	0.039
Europe	road	0.244	0.238	0.014	0.224–0.250	0.005
Benelux+	IWW	1.032	0.980	0.031	0.946–1.005	0.052
Benelux+	rail	1.195	1.075	0.077	1.000–1.155	0.120
Benelux+	road	0.224	0.214	0.007	0.207–0.221	0.010
Germany	IWW	1.383	1.347	0.095	1.255–1.446	0.036
Germany	rail	0.630	0.616	0.042	0.571–0.653	0.015
Germany	road	0.107	0.107	0.012	0.096–0.120	0.001

corridors. This is expected: changing the random OD split can include or exclude a few important non-road OD pairs.

These results therefore support a cautious interpretation of the holdout exercise. The coarser calibration appears to capture some transferable structure by mode, commodity group and distance band, but it does not reproduce the detailed base-year geography in the same way as the full "mode-od-group" calibration. The holdout test should therefore be read as a diagnostic of transferability, not as a substitute for the detailed base-year calibration used for assignment.

These results suggest that the framework should be understood as having two operating modes. For base-year assignment, the detailed "mode-OD-group" calibration is the relevant setting: it uses the behavioural model as a structural basis but corrects the OD-commodity geography of rail and inland-waterway flows so that the matrices can be assigned realistically. For diagnostic tests or scenario applications, a coarser calibration level is more appropriate, because it does not rely on OD-specific corrections that may not transfer beyond the base year. In that case, the model should be interpreted as a pivot around a calibrated reference situation.

6.7 Implications for policy-scenario applications

The sensitivity results also define the limits within which the framework can be used for policy scenarios. Scenario calculations should be implemented as pivots around the calibrated base year: the behavioural attributes and utilities are recomputed under the scenario, while the detailed base-year calibration constants are held fixed. Changes in predicted modal quantities then result from changes in the behavioural component rather than from recalibrating the model to the scenario.

The reference specification provides a robust direct response to changes in IWW cost in all three datasets. For rail, however, the situation is more restrictive. Rail cost is not included in the reference model, so a rail-price change has no direct effect on rail utility. Rail moving time has a negative estimated coefficient in the European dataset, but the coefficient reaches its zero bound in the Benelux + and German datasets. In those two models, a change in calculated rail moving time therefore produces no direct marginal response through the moving-time term.

Rail infrastructure changes may still affect predictions when they alter the existence or validity of a rail path, modal availability or another attribute entering the utility with a non-zero coefficient. The model should not, however, be assumed to represent the full effect of rail pricing or performance policies when the relevant coefficient is excluded or located at its bound. In particular, improvements that only reduce rail cost or moving time would have little or no direct effect in the two NUTS3 models.

For applications centred on rail pricing, the tested rail-cost specification could be used as a sensitivity case for the European dataset, where the coefficient is negative and estimated away from its bound. It would be less defensible to use the zero-bound estimates from the NUTS3 datasets as transferable rail-price responses. Such applications would require additional temporal or cross-sectional price variation, external elasticity evidence, a constrained coefficient based on prior information, or a dedicated re-estimation. The present framework is therefore better suited to base-year assignment and local pivot analysis than to unrestricted forecasting of rail-price policies.

7 Implementation and software-consistency test in Nodus

Nodus was used to compute the explanatory variables required by the model. It also provides an open API through which user-defined modal-choice plugins can be integrated transparently into the assignment workflow. The two-layer model discussed in this paper was thus implemented in a plugin to connect estimation with assignment. Estimated coefficients, centring constants, distance thresholds, calibration settings and constants are stored in a database key-value table and the plugin reads the table and computes calibrated logit probabilities.

The observed shares reported in Table 11 are calculated for the set of OD–commodity flows retained in the Nodus assignment run. This scope is slightly different

Table 11 Software consistency of aggregate modal shares between the Python model and the Nodus assignment

Model	Mode	Observed share (raw)	Python predicted share	Nodus plugin share
Europe	IWW	6.17%	6.13%	5.84%
	Rail	13.95%	13.89%	13.72%
	Road	79.88%	79.98%	80.44%
Benelux+	IWW	12.15%	11.10%	11.17%
	Rail	8.38%	8.33%	8.23%
	Road	79.48%	79.68%	80.59%
Germany	IWW	2.53%	2.44%	2.49%
	Rail	12.07%	12.01%	11.87%
	Road	85.40%	85.55%	85.64%

The observed shares in this table refer to the OD–commodity cells retained in the Nodus assignment run. Tables 1 and 3 refer to the prepared Python estimation sample

from the prepared estimation sample used in Tables 1 and 3. In the Python estimation procedure, observations are retained after checking positive total tonnage and the validity of the network-derived cost, distance and moving-time attributes. In Nodus, modal paths and availability are reconstructed within the assignment environment, and a small number of OD–commodity cells may consequently be treated differently.

The underlying freight matrices and reference year are the same. The small differences in observed modal shares therefore do not result from the modal-choice model or from the calibration procedure, but from the slightly different set of cells entering the estimation and assignment calculations. Indeed, the input data used by the Python script are derived from a network assignment in which IWW and rail routes are considered valid only if their length does not exceed 2.5 times that of the corresponding route for the reference mode. This rule is intended to exclude unrealistic modal alternatives. It affects a small number of OD pairs for which the original OD matrices report positive IWW or rail flows, although the corresponding alternative was excluded during the assignment.

Tables 11 and 12 are to be considered as software-consistency tests. They examine whether the estimated coefficients, centring constants and bounded calibration constants can be transferred from the prepared Python modelling environment to the operational Nodus assignment environment while producing broadly comparable modal shares and OD-cell results. The comparison does not assess agreement with independent observations of assigned road, rail or inland-waterway traffic and should therefore not be interpreted as an external validation.

The comparison is less close for German inland waterways than for the other dataset-mode combinations, but Tables 11 and 12 must be interpreted together. In the assignment scope, the observed German IWW share is 2.53%, compared with 2.44% in the Python prediction and 2.49% in the Nodus assignment. Nodus is therefore closer to the observed aggregate share than Python: its difference is about – 0.04 percentage points, compared with about – 0.09 percentage points for Python. Nevertheless, the Nodus IWW WAPE is 0.047, whereas the internal Python WAPE is 0.005.

This apparent contrast follows from the different information conveyed by the two indicators. The aggregate modal share is based on the sum of all predicted IWW tonnes, so overprediction in some OD-commodity cells can offset underprediction in others. WAPE instead sums the absolute cell-level errors before dividing by observed IWW tonnes. The German results therefore indicate that Nodus reproduces the total IWW volume reasonably well, but distributes part of that volume differently across OD-commodity cells. The difference between 0.005 and 0.047 cannot be attributed primarily to an aggregate modal-share bias.

Because German IWW traffic represents only about 2.5% of total freight and is concentrated on a limited number of navigable corridors, relatively small cell-level implementation differences can noticeably increase WAPE while leaving the aggregate modal share almost unchanged. A brief decomposition of the German IWW discrepancy indicates that it is not driven by a single OD relation or by only a few extreme cells. Rather, the additional absolute error is distributed across a larger set of OD–commodity relations and is concentrated in a few commodity groups, particularly petroleum products, chemicals, and minerals and building materials. These three groups account for about 81% of the additional absolute error in the implementation comparison. The higher Nodus WAPE should therefore be interpreted as a cell-level implementation discrepancy rather than as an aggregate modal-share bias. This remaining cell-level discrepancy constitutes a limitation of the present Nodus implementation-consistency test. Its precise operational source would require a more detailed comparison of

Table 12 Software consistency of OD-cell WAPE between the Python model and the Nodus assignment

Dataset	Road		IWW		Rail	
	Python	Nodus	Python	Nodus	Python	Nodus
Europe	0.001	0.002	0.005	0.011	0.006	0.009
Benelux+	0.002	0.002	0.007	0.007	0.011	0.011
Germany	0.001	0.002	0.005	0.047	0.009	0.009

paths, availability indicators, calibration keys and predicted tonnes between the two implementations.

8 Conclusion

This paper developed a calibrated freight modal-choice framework linking econometric estimation, base-year OD-matrix calibration and network assignment. The results show that a relative-to-road MNL provides a useful behavioural structure but cannot by itself reproduce the matrices. For freight assignment, aggregate modal shares are not enough; base-year matrix reproduction must also be assessed at the OD-cell level.

The behavioural results also provide several consistent findings. The relative IWW-cost coefficient is negative in all three datasets, while the rail market-structure indicator is positive. A rail-cost term was tested but not retained in the reference specification: it was negative and identified away from its bound only in the European NUTS2 dataset, where it produced a modest pre-calibration WAPE reduction, but reached its zero bound in the two NUTS3 datasets and produced no improvement after detailed calibration.

Moving-time and distance effects are likewise more dependent on spatial scale and dataset structure, with the moving-time coefficients reaching their zero bounds in the two NUTS3 applications. The commodity-specific constants further confirm substantial differences in modal propensity across freight categories. These findings support a common model architecture, but not a single transferable coefficient vector across datasets.

A bounded calibration layer is introduced for that purpose. The preferred operational "mode-OD-group" level substantially improves base-year matrix reproduction by adding bounded constants to inland-waterway and rail utilities. Calibration diagnostics show that many negative capped constants correspond to technically available but unused non-road cells, while observed non-road tonnes are generally not concentrated at the negative cap. This supports the interpretation of calibration as a transparent mechanism for restoring sparsity and corridor specificity.

The calibration does not distinguish structural non-availability from observed non-use: lower-bound constants may represent effective choice-set exclusion, unobserved logistics constraints or incidental zero flows. A two-stage or hurdle formulation would require additional information capable of identifying these mechanisms separately.

The repeated random holdout diagnostics provide limited evidence that coarser calibration levels can improve performance on unseen OD pairs within the same geographical datasets, but detailed OD-specific utility corrections remain necessary for base-year matrix fit. The holdout results concern unseen OD pairs within the same geographical datasets; transfer to entirely unseen regions or freight corridors remains to be assessed through spatially blocked or corridor-based experiments. The comparison across Europe NUTS2, Benelux+ NUTS3 and Germany NUTS3 also shows that the model should be general in architecture but dataset specific in coefficients and calibration constants.

The Nodus implementation demonstrates that the framework can be deployed operationally. The complete workflow connects data extraction, Biogeme estimation, bounded calibration, OD-cell evaluation and network assignment.

The framework's policy-response capabilities are also attribute dependent. The reference model contains no direct rail-cost response, and the rail moving-time coefficient is not identified away from zero in the two NUTS3 datasets. The framework should therefore be used as a calibrated pivot model, with base-year constants held fixed, and not as a general source of transferable rail-price or rail-time elasticities. Policy applications focused on these effects require additional evidence or dedicated re-estimation.

A remaining limitation is that the empirical assessment is based on the modal OD matrices themselves rather than on independent observations of assigned network flows. The comparison between the Python model and the Nodus implementation shows that the calibrated model can be transferred consistently to the assignment environment, but it does not replace a full external validation against rail corridor counts, inland-waterway traffic observations, port hinterland flows or road freight counts. Such a validation would require additional harmonized data and is therefore left for future work. It would be a natural next step to assess how well the calibrated OD matrices reproduce observed flows once assigned to the multimodal network.

The main conclusion is that freight modal-choice modelling for assignment requires both an interpretable behavioural structure and an explicit calibration mechanism capable of reproducing the sparse geography of freight flows.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12544-026-00863-9.

Supplementary Material 1: Appendix A

Supplementary Material 2: Appendix B

Supplementary Material 3: Appendix C

Acknowledgements

NA.

Author contributions

Bart Jourquin is the sole author of the manuscript.

Data availability
The datasets, Python scripts and Java source code for the Nodus plugin supporting the conclusions of this article are available in the « CalibratedODModalChoice » Zenodo repository https://doi.org/10.5281/zenodo.20493251.

Declarations

Ethics approval and consent to participate
Not applicable.

Generative AI use
During the preparation of this manuscript, the author used ChatGPT as an editorial and drafting assistant. The tool was used to help structure the manuscript, improve clarity and concision, reformulate selected passages, and assist with the preparation of explanatory text. It was not used to generate empirical results, estimate the models, alter the data, or produce figures. All modelling choices, computations, interpretations, references and conclusions were made, checked and approved by the author, who remains fully responsible for the content of the manuscript.

Competing interests
The authors declare no competing interests.

Received: 1 June 2026 / Accepted: 23 October 2026
Published online: 29 October 2026

References

1. Abrahamsson, T. (1998). *Estimation of origin-destination matrices using traffic counts - A literature survey*.
2. Arencibia, A. I., Feo-Valero, M., García-Menéndez, L., & Román, C. (2015). Modelling mode choice for freight transport using advanced choice experiments. *Transp Res Part Policy Pract*, *75*, 252–267. https://doi.org/10.1016/j.tra.2015.03.027
3. Australian Transport Assessment and Planning Steering Committee. (2016). *T1 travel demand modelling (Australian transport assessment and planning guidelines)*. Transport and Infrastructure Council.
4. Bates, J., Ashley, D., & Hyman, G. (1987). The nested incremental logit model: Theory and application to modal choice. In *Proceedings of the 15th PTRC Summer Annual Meeting*. University of Bath, England.
5. Ben-Akiva, M., & de Jong, G. (2013). The aggregate-disaggregate-aggregate (ADA) Freight model system. In M. Ben-Akiva, H. Meersman, & E. Van de Voorde (Eds.), *Freight transport modelling* (pp. 69–90). Emerald Group Publishing Limited. https://doi.org/10.1108/9781781902868-004
6. Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. MIT Press. MIT Press series in transportation studies.
7. Bierlaire, M. (2023). *A short introduction to Biogeme (Report No. TRANSP-OR 230620)*. Transport and Mobility Laboratory, School of Architecture. Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne.
8. Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator. *Transp Res Part B Methodol*, *18*, 289–299. https://doi.org/10.1016/0191-2615(84)90012-2
9. Cullinane, K., & Toy, N. (2000). Identifying influential attributes in freight route/mode choice decisions: a content analysis. *Transp Res Part E Logist Transp Rev*, *36*, 41–53. https://doi.org/10.1016/S1366-5545(99)00016-2
10. Daly, A., Fox, J., Patruni, B., & Milthorpe, F. (2011). Pivoting in travel demand models. In *European Transport Conference*.
11. Daly, A., Fox, J., & Tuinenga, J. G. (2005). *Pivot-point procedures in practical travel demand forecasting*.
12. de Jong, G., Gunn, H., & Ben-Akiva, M. (2004). A meta-model for passenger and freight transport in Europe. *Transport Policy*, *11*, 329–344. https://doi.org/10.1016/j.tranpol.2004.03.001
13. De Jong, G., Vierth, I., Tavasszy, L., & Ben-Akiva, M. (2013). Recent developments in national and international freight transport models within Europe. *Transportation*, *40*, 347–371. https://doi.org/10.1007/s11116-012-9422-9
14. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal Of Forecasting*, *22*, 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001
15. Jensen, A. F., Thorhauge, M., de Jong, G., & Rich, J. (2019). A disaggregate freight transport chain choice model for Europe. *Transp Res Part E Logist Transp Rev*, *121*, 43–62. https://doi.org/10.1016/j.tre.2018.10.004
16. Jourquin, B. (2025). Direct and cross cost elasticity estimations for freight transport in Europe using constructed dependent variables. *Res Transp Econ*, *111*, 101566. https://doi.org/10.1016/j.retrec.2025.101566
17. Jourquin, B. (2023). Accessibility Measures in Modal Choice Models: A Proof of Concept Applied to Freight Transport. *J Transp Technol*, *13*, 38–60. https://doi.org/10.4236/jtts.2023.131003
18. Jourquin, B. (2017). Estimation of travel time elasticities with a multimodal freight transport model calibrated and validated on publicly available regional data. In *14th International Nectar Conference*. Madrid.
19. Jourquin, B. (2016). Calibration and Validation of Strategic Freight Transportation Planning Models with Limited Information. *J Transp Technol*, *06*, 239–256. https://doi.org/10.4236/jtts.2016.65023
20. Kumar, A. (1980). Use of incremental form of logit models in demand analysis. *Transportation Research Record*, *775*, 21–27.
21. McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press New York.
22. Rich, J., Holmblad, P. M., & Hansen, C. O. (2009). A weighted logit freight mode-choice model. *Transp Res Part E Logist Transp Rev*, *45*, 1006–1019. https://doi.org/10.1016/j.tre.2009.02.001
23. Samimi, A., Kawamura, K., & Mohammadian, A. (2011). A behavioral analysis of freight mode choice decisions. *Transp Plan Technol*, *34*, 857–869. https://doi.org/10.1080/03081060.2011.600092
24. Spiess, H. (1990). *A gradient approach for the O-D matrix adjustment problem (No. Publication 693)*. *Centre de recherche sur les transports*. Université de Montréal.
25. Szimba, E., Kraft, M., Ihrig, J., Schimke, A., Schnell, O., Kawabata, Y., Newton, S., Breemersch, T., Versteegh, R., Meijeren, J. V., Jin-Xue, H., Stasio, C. D., & Fermi, F. (2012). ETISplus database content and methodology. https://doi.org/10.13140/RG.2.2.16768.25605.
26. Tavasszy, L. A., & de Jong, G. (2014). *Modelling freight transport*. Elsevier.
27. Train, K. E. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge University Press.
28. United States Federal Highway Administration (Ed.), (2010). *Travel model validation and reasonableness checking manual - Second Edition*.
29. Van Zuylen, H. J., & Willumsen, L. G. (1980). The most likely trip matrix estimated from traffic counts. *Transp Res Part B Methodol*, *14*, 281–293. https://doi.org/10.1016/0191-2615(80)90008-9

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.