A logit model for shipment size choice with latent classes – empirical findings for Germany

Raphael Piendl [a,1], Gernot Liedtke [b], Tilman Matteis [c]

[a] *German Aerospace Center (DLR), Institute of Transport Research, raphael.piendl@dlr.de*
[b] *German Aerospace Center (DLR), Institute of Transport Research, gernot.liedtke@dlr.de*
[c] *German Aerospace Center (DLR), Institute of Transport Research, tilman.matteis@dlr.de*

Abstract

Decisions on shipment size in freight transport are often seen to represent a whole set of logistics decisions made by shippers and recipient. Also, shipment sizes has a large impact on transport mode choice.

Therefore, they are an important aspect in the modeling of freight transport demand, as they allow to display the reactions of various stakeholders on policy measures. In this article, a model for the discrete choice of shipment sizes is applied to interregional road freight transport. Preferences of actors are reflected by a total logistics cost expression. A Latent Class Analysis approach is applied to identify groups of transport cases with similar logistics requirements. The classification reduces significantly heterogeneity in behavior. Reactions of actors on external influences such as policy measures could be predicted more accurately.

***Keywords**: freight transport; shipment size; total logistics cost; discrete choice; latent class.*

1. Motivation

The globally growing relevance of freight transport increases the importance of adequately assessing policy measures for predicting the changes in freight transport demand resulting from external influences. As the logistics activities of involved actors determine to a great extent the characteristics of freight transport, the use of behaviorally sensitive freight transport models based on rational decisions and covering the main dimensions of logistical behavior are becoming more and more necessary.

Generally, the consideration of logistic choices in the context of freight transport is accompanied by a huge variety and diversity of involved actors (shippers, carriers, receiver, operators etc.) and an enormous diversity of transported commodities. As a consequence, each company's logistics operations are differently organized and very detailed on the level of single actors, so that they cannot be inserted into comprehensive freight transport models directly. Instead, simplifications and generalizations have to be found which still allow causes and effects to be traced, and which also preserve the variability in logistics behavior in the model.

A usual practice of finding such simplifications is to model single logistics decisions on a high level of aggregation. One of these proxy decisions is the choice of shipment sizes, which covers the central aspects of the actor's logistical calculus and to some extent explains the behavioral heterogeneity of the actors. In general, the shipment can be seen as one of the simplest links between commodity flows formed from economic interactions on the transport demand side and the vehicle movements that take place on the various transport routes. On the one hand, the size and other properties of the shipments determine the arrangement of the trade relationships between shippers and receivers from which the demand for freight transport results. On the other hand, the requests of the potential customers and the restrictions placed by the transport system are influencing the offers of the transport companies competing on the transport markets. Because of the shipment's occurrence on all stages of decision, the finding of a

[1] Corresponding author. Tel.: +49 30 67055-189; fax: +49 30 67055-283.

shipment size that fulfills the requirements of all involved parties can serve as a model at a high aggregation level of the underlying logistics considerations.

Summarizing the previous aspects, two questions arise in developing a shipment size choice model for large-scale freight transport model systems:

1.) Which influences have to be considered as crucial for the explanation of logistical actors' behavior, and how can they be incorporated in an operational transport model?
2.) How can the omnipresent heterogeneity of actors and decision situations be addressed in a manageable yet realistic way?

The purpose of this paper is to suggest answers to these two questions. We propose a discrete shipment choice model based on a total logistics cost formulation, as it turned out that decisions are oriented on given vehicle and bundle sizes. Additionally, technology-related or market-related influencing factors on transport cost functions, such as the number of transshipments or the characteristics of the goods, support a certain categorical character (Combes et al. (2016)). This orientation results in shipment sizes that fall into a limited number of categories. Moreover, a limited set of size categories from which to choose seems to be more manageable when it comes to integration of the shipment size component in a comprehensive freight transport model.

Discrete choice models are usually estimated from samples that cover a wide range of responses and thus decision situations. Consequently, the estimated model instances reflect the behavior of the sample population as a whole and cannot account for the variability of actor reactions on external influences. A way to alleviate this is to distinguish segments of freight transport demand that exhibit similar behavior. In this case study, we use a novel approach in the context of shipment size choice that segments transport cases according to logistics aspects by means of a latent class model, which in combination with a rational framework of total logistics cost minimization improves the explanatory power of the shipment size choice behavior.

For model estimation a dataset was used which was gathered at the beginning of 2013 within the scope of the research project "'Development of a model for the calculation of freight traffic modal shifting to derive consistent evaluation approaches for German federal infrastructure planning (BVWP)" (BVU – Beratergruppe Verkehr + Umwelt, TNS Infratest (2014)). The data contains in total an amount of 926 transport cases gained from 474 interviews with shippers and receivers. The model in this article has to be restricted to road transports as there are not enough observations for modes other than truck available (23 by rail, 9 by inland waterway, 13 by intermodal rail and 5 by intermodal inland waterway). Nevertheless, the estimated model provides possible links for the comprehensive integration of the mode choice.

This article is organized as follows: Section 2 contains a brief review of the literature dealing with shipment size choice models and the formation of homogeneous groups. Section 3 covers the conceptual framework, which consists of the theoretical background of shipment size choice, the development of a discretized total logistics cost model, and the essentials of determining latent classes. The source data and descriptive analysis take center stage in section 4, whereas section 5 presents the estimation results. Finally a summary and an outlook on possible directions for further research will be given in section 6.

2. Literature review

Most freight transport models assume that the goods to be transported between shippers and recipients do not result from one-off businesses. It is rather assumed that the shipment will be part of a long-established business relationship. From such relationships, a total flow of goods per period results, which has to be split up into several shipments. Two cost components determine the optimal partition. For every order placed at a transport company, a fixed amount has to be payed, regardless of the quantity that is shipped. On the other hand, a small number of large shipments will cause high inventory holding costs. A simple mathematical expression for this tradeoff is given by the Economic Order Quantity (EOQ) model (Harris (1913)). In the context of freight transport, Baumol and Vinod (1970) developed a total logistics cost approach for the determination of shipment sizes. Hall (1985) considered the influence of given vehicle sizes when taking the lower envelope of the transport cost functions of various transport modes as one of the decision criteria. By doing so, the inherently discrete choice of a transport mode was combined with the choice of a (continuous) shipment size. A further reason for this joint consideration is that shipment sizes depend on total logistics cost, which, in turn, are heavily influenced by the physical provision of transport. Thus, most subsequent shipment size models did not consider shipment size choice as an isolated factor, but rather combined it with the choices of mode, carrier or transport chain. However, Holguin-Veras (2002) pointed out that such combined choices are often not taken by the same decision maker, although they are in many cases seen as belonging together.

In large-scale freight transport model systems, econometric models based on the total logistics cost as the actor's rational choice criterion for the determination of shipment sizes prevail. Within the group of these models, there are basically three different ways in which the problem has been addressed:

1.) Modeling of shipment sizes independently of transport mode choice
2.) Modeling of continuous shipment sizes and discrete modes of transport
3.) Modeling of discrete shipment sizes and discrete modes of transport

In the first group of models, single shipments are considered in order to examine the behavior of actors on a microscopic basis. Further decisions that influence total logistics cost are either not in the scope of the model or not coupled tightly to the decisions related to shipment size. Combes (2009) tried to verify the economic order quantity (EOQ) equation on the basis of the French shipment survey ECHO[2]. He added further dummy variables that indicate the chosen mode of transport for the sampled shipment at hand. Given the data from the survey, it was shown that the EOQ model is a good approximation for the choice of a continuous lot size, regardless of what mode of transport was actually chosen. The most striking result was that all transport cases can be dealt with using one set of parameter values, requiring no further segmentation of transport demand. Moreover, further variables were added to explain lot size formation, such as transport distance and whether the shipment was transported directly or in a vehicle tour. As the model was estimated using a comprehensive sample of transport cases of all kinds, the theoretical EOQ model can feasibly be inserted in a more comprehensive freight transport model. Such an insertion was done by Wisetjindawat et al. (2005). In a commodity-based model for urban freight transport, the EOQ formula was applied to create shipments that were fed into vehicle tours. The model of Kawamura et al. (2010) did not take logistics aspects into account when modeling shipment sizes, but rather replicated the distribution of these sizes, which was obtained for the relevant commodity type from the US commodity flow survey. Although the model of Wisetjindawat et al. (2005) also incorporated a feedback on shipment size

[2] Enquête ECHO Envoi-Chargeurs-Opérateurs de transport (Guilbault and Soppé (2009)).

decisions from the transport system, parameters on the single model stages were estimated separately from the remaining parts of the model. This is different from models listed under 2.) and 3.). In these cases, shipment size choice was intertwined with other choices on the level of parameter estimation. The influence that multilevel decisions have hereby on the parameter estimates depends on the model structure.

Models from the second group take into account the influence of total logistics cost caused by the chosen transport mode on the shipment size. An example for this is given by De Jong et al. (2010). Given a flow of goods, several possible delivery frequencies are determined. For each frequency corresponding to a shipment of a certain size, the total logistics cost of all feasible transport chains are calculated and the cheapest combination of frequency and transport chain is chosen. Besides this simulation approach, econometric models prevail in this group.

In these econometric models, two general problems occur. First, as shipment size often enters the mode or vehicle choice submodel as an independent variable, correlation between the submodels has to be accounted for. Further, in models with revealed preference data, only such data records will exist in which the shipment size is conditional on the chosen mode or vehicle. This leads to selectivity bias. Holguín-Veras (2002) and De Jong and Johnson (2009) applied a two-stage procedure that starts with estimating the parameters for the continuous shipment size model. To avoid selectivity bias and feedback between the submodels, shipment size was estimated independently from the chosen mode or vehicle type on regressors that did not occur in the discrete choice submodel. The shipment size obtained by the OLS regression was entered into the discrete choice model for vehicle type or mode choice. As the shipment size was estimated independently, influencing factors of the shipment size choice on the mode choice were not directly integrated into the mode choice model and therefore testing for simultaneity bias was not possible. Abate and de Jong (2014) also pursued a two-step approach but started with transport mode choice according to a MNL model. The choice probabilities obtained served as arguments for a correction function that was added to the shipment choice equation. A completely different path was taken by Abdelwahab and Sargious (1992), who estimated a switching simultaneous equations model for only two modes (rail and truck) with a two-stage least squares regression. They compared their results with those obtained from an estimation in one step with a maximum likelihood estimator.

Models of type 3.) assume that shipment sizes can be classified into discrete categories. This goes along with the finding of Hall (1985) that certain shipment sizes are unfavorable given the vehicles or transport modes to choose from. Moreover, combinations of shipment size categories allow the demand for certain transport solutions to be examined. Medium-sized shipments combined with rail transport indicate single-wagon traffic, for example. In the models of category 3.), random utility models were combined either by nesting (e.g. Windisch et al. (2010), De Jong and Ben Akiva (2007)) or by copula functions (e.g. Pourabdollahi et al. (2013)). Along with the application of nested discrete choice models goes the task of choosing the appropriate nesting structure. According to Windisch et al. (2010), shipment sizes are better placed in the lower level of a nested logit model, indicating that a switch between shipments of various sizes is more usual than between transport chains. Pourabdollahi et al. (2013) refrained from nesting discrete choices, instead using various copula functions to link multinomial logit choices of shipment size and mode choice. By the choice of an appropriate copula function, dependency structures between the two decision problems can be modeled with greater flexibility than with the hierarchical nested logit model.

Regardless of the context in which the formation of shipment sizes was modeled, often three aspects are determined as important: Shippers' preferences, transport cost, and the properties of the good to be transported. Especially the latter two aspects result from logistic or technical restrictions to which the actors have to adapt.

In several cases, logistics components have been mirrored by proxy variables such as characteristics of shipper and/or recipient (Holguín-Veras (2002), Pourabdollahi et al. (2013), De Jong and Johnson (2009)) and the way that a shipment was packaged or handled (De Jong and Abate (2014), Abdelwahab and Sargious (1992), Windisch et al. (2010)). Transport logistics aspects were addressed by Combes (2009), who added information indicating if the shipment was part of a tour or delivered directly. Windisch et al. (2010) added transport logistics via the structure of the choice models. Shipment sizes are nested given certain intermodal transport chains in which they are transported.

Often preferences and logistics requirements are distributed very heterogeneously within the population of all transport cases at hand. Segmentation of the demand population is a possibility to enhance the behavioral foundation of econometric models. In contrast to the findings of Combes (2009), demand has thus been segmented in some operational freight transport models in order to account for this heterogeneity. De Jong et al. (2010) apply different lot size models depending on the considered commodity type. Kawamura et al. (2010) also distinguish between commodity types when fitting shipment sizes to observed distributions. In general, exogenously prescribed segments, such as the NST 2007 classification of commodity types, are likely to contain still a lot of heterogeneity in the context of logistics as they try to integrate different dimensions like packaging categories, sector relations and commodity type-related properties into the respective segments (Liedtke and Schepperle (2004)).

Segmentation of demand can also be data-based to a model to varying degrees. One way to group decision makers into classes in discrete choice situations is with latent class models. In passenger transport, Bhat (1997) endogenously delimited several traveler groups according to socioeconomic characteristics and estimated the model with an expectation-maximization algorithm. Demand segmentation for mode choice in freight transport was done by Gopinath (1995), who classified shippers according to attitudes towards various modes of transport which in turn were derived from logistics figures such as maximum acceptable delay. Arunotayanun (2009) segmented mode choice according to the logistic properties of the relationship between shipper and recipient.

The model proposed in this article combines discrete shipment size choice with latent class segmentation. Due to data availability, only shipments transported by road are considered. Nevertheless, the general approach of the model allows the investigation of the interdependencies between shipment size choice and mode choice in further developments by linking up with other state-of-the-art discrete choice models. However, even for a single transport mode, the specification of size categories allows demand segments for various transport products to be identify, which is preferable to using the prescribed standard categorizations. The logistics properties of these demand segments are incorporated in the model via latent class analysis. To our knowledge, latent class models have not been applied to shipment size decisions in the context of freight transport up to now.

3. Modeling approach

According to basic microeconomic theory, a final consumer with transitive, reflexive and complete preferences chooses a bundle of continuous and positive quantities of goods and services – while satisfying prevalent constraints - for which he receives the maximum utility. The utility itself is a dependency function of the goods and services which mathematically represents the preferences and is, apart from potential order-preserving transformations, unique. Transferred to the logistical context, a decision-maker, which is either a shipper or a recipient, chooses the shipment size for which he receives the maximum utility (minimum costs). This is represented by the optimal distribution of the annual flow of goods in various shipments, leading to an optimal shipment size, which minimizes a decision-maker's logistics cost by balancing the tradeoff between inventory cost and fixed transport cost in an optimal way. The total logistics cost C_n per period (in this case: per year) for a decision maker n is dependent on the shipment size and is made up of four components (in the respective order): fixed transport cost $C_{n,fix}$, variable transport cost $C_{n,variable}$, cost for inventory in transit $C_{n,transit}$ and storage cost $C_{n,storage}$[3]:

$$C_n = C_{n,fix} + C_{n,variable} + C_{n,transit} + C_{n,storage}. \tag{1}$$

Assuming constant and continuous production rate of the shipper and consumption rate of the recipient, the cost expression from equation (1) can now be refined: The fixed transport cost is assumed to be made up of the frequency of the shipments per year multiplied with the cost per shipment F_n [€/shipment]. Moreover, the frequency of transport is given by the ratio of the constant and continuous total flow of transported goods Q_n [ton/year] and shipment size per transport q_n [ton/shipment] to satisfy the total demand Q_n. The variable transport cost is characterized by the function $c_n(q_n)$ [€/ton] for the deciding firm n, which is dependent on the shipment size q_n multiplied with the total amount of transported goods Q_n. Additionally, the cost for inventory in transit represents the cost for the bounded capital – interest rate r times value density of the transported commodities v_n [€/ton] – during the transit time t_n arising for the total amount of the transported goods Q_n. The last component is given by the storage cost, which made up of two parts: warehousing cost w_n and the capital cost rv_n which incur for the average stock $\frac{q_n}{2}$. As the model is based on cost all parameters are assumed to be positive. Altogether, the parametrized total logistics cost in dependency of the shipment size are given by:

$$C_n(q_n) = \frac{Q_n}{q_n} F_n + c_n(q_n)Q_n + rv_n t_n Q_n + \frac{q_n}{2}(w_n + rv_n). \tag{2}$$

The total logistics cost is not a typical utility function in the sense of the microeconomic consumption theory, which increases with a higher amount of the consumed good and is marked by the consideration of multiple goods or/and services. The continuous cost function, in contrast, is one-dimensional with respect to the decision space as well as regarding the image space and reaches its optimum for q_n and has to be in the mathematical left-open and right-closed interval $]0; Q_n]$ depending on the values of the parameters representing the cost for transport and inventory holding.

Assuming a linear non-decreasing function representing proportional variable transport cost $c_n(q_n)$, the minimization of $C_n(q_n)$ leads to the optimal shipment size for decision maker n:

[3] This total logistics cost formulation assumes an immediate and complete replenishment leading to the absence of stock-outs. Extensions, such the availability of a safety stock, damage or deterioration costs or stock-out costs, can be integrated (De Jong and Ben-Akiva (2007)). As this gets mathematically complex and data for estimation is not available, we disregard such auxiliaries in this model.

$$q_n{}^* = \sqrt{\frac{2F_nQ_n}{w_n+rv_n}}. \tag{3}$$

Equation (3) reveals the relationship between the optimal shipment size for decision maker n and its influencing parameters. An increase of the fixed transport cost, such as order, handling and set up costs, leads to a higher optimal shipment size and so does an increasing flow of goods Q_n. As the inventory holding cost increases, the optimal shipment size will decrease due to the relatively higher cost for capital commitment and storage.

3.1 Choice sets and applied model

The model presented in this study is based on discrete choice theory using a random utility approach. A prevalent discrete space of alternatives changes the conditions of decision making: Choosing only one alternative, and therefore not realizing the remaining ones, leads to corner solutions and thus to a non-applicable marginal calculus, which necessitates the consideration of the utility functions of each alternative directly (Ben-Akiva and Lerman (1985)). The decision rule is given by the selection of the alternative which spends the highest utility among all reachable alternatives. The utility functions are assumed to be a summation of attributes (with their coefficients) describing the alternatives and the characteristics of the decision maker. The attributes themselves are weighted with parameters representing the influence on the utility functions as linear in parameter specifications.

The development of the discrete shipment size choice model includes the categorization of the continuous shipment size into different shipment size classes. Following the classification of general cargo, partial loads and (multiple) full loads on road transports and the realized shipments in the RP data, we divided the shipment sizes into three categories:

$$q_1 = \begin{cases} 1, & if\ 0t < q_n{}^* \leq 3t\ and \\ 0, & otherwise. \end{cases}$$

$$q_2 = \begin{cases} 1, & if\ 3t < q_n{}^* \leq 12t\ and \\ 0, & otherwise. \end{cases} \tag{4}$$

$$q_3 = \begin{cases} 1, & if\ 12t < q_n{}^*\ and \\ 0, & otherwise. \end{cases}$$

This yields to the global choice set $S = \{q_1, q_2, q_3\} = S_n,\ \forall n$ assuming all alternatives are accessible for every decision maker n.

The discretized logistics cost $C_n(q_i)$ – adapted from equation (2) – for actor n choosing shipment size class q_i can be expressed by:

$$C_n(q_i) = \left(\frac{F_n}{q_i} + c_n(q_i) + rv_nt_n\right)Q_n + \frac{q_i}{2}(w_n + rv_n). \tag{5}$$

In general, micro-level data sets in freight transport are all globally very scarce with respect to transports. The data set used in this study is lacking in regard to the amount of useable observations for other modes than truck transports. For that reason, the model in this case is restricted to long-haulage distance road transports which in Germany are predominantly conducted by semi-trailers (between 55% and 90% of transports, increasing with the haulage distance[4]). Generally, the main proportion of fixed transportation cost regarding road transports falls upon the number of stops and the distance between them during pre-carriage and onward carriage. Due to the exclusive consideration of long-haulage distance road transports and the predominant use of semi-trailers during the main run, we assume the fixed transport cost F_n to be constant for each shipment size category q_i. This formulation on the one hand covers the differing amount of stops caused by collection of piece goods, partial loads and full truck loads and, on the other, affords the opportunity to relax the assumption in further enhancements of the model, with multiple means of transport or varying fixed cost within each of the modes.

The warehousing costs mainly depend on the handling of the good (heating, cooling, packaging etc.) and the respective space requirement (sizing of warehouses, opportunity costs of space consumption etc.) during the storage processes. The handling and the space used strongly go along with the characteristics, weight and respective density of the commodity. Due to data constraints regarding the density of the transported goods, and as a useful simplification without loss of model validity, we suppose the warehousing costs w in this model only to vary between the different shipment size classes and not between each decision maker n. As can be seen in equation (8), the warehousing cost, for which no data are available in the estimation sample, then become a part of the parameter $\beta_{q_i,2}$ which is alternative-specific. This leads to the following specification:

$$C_n(q_i) = \left(\frac{F_{q_i}}{q_i} + c_n(q_i) + rv_n t_n\right) Q_n + \frac{q_i}{2}\left(w_{q_i} + rv_n\right). \tag{6}$$

The division of expression (6) by the constant and continuous flow of goods Q_n resulting in the total costs per ton represents a model formulation, which on the one hand can be empirically estimated, and on the other is properly interpretable with respect to the alternative-specific constants:

$$\frac{C_n(q_i)}{Q_n} = \frac{F_{q_i}}{q_i} + c_n(q_i) + rv_n t_n + \frac{q_i w_{q_i}}{2} \cdot \frac{1}{Q_n} + \frac{q_i r}{2} \cdot \frac{v_n}{Q_n}. \tag{7}$$

The variable transport costs $c_n(q_n)$ from equation (2) are originally characterized by a positive dependency on the haulage distance and on the shipment size (Abate and De Jong (2014)). As the shipment size in our model is discretized, the variable transport costs $c_n(q_i)$ are dependent on the shipment size categories and the haulage distance. For that reason, the variable transport costs are approximated by a function of haulage distance, varying for the respective shipment size categories. The distance without detailed cost functions then represents the relative attractiveness of the respective shipment size categories.

Choosing the alternative with maximum utility, which is in our case equivalent to minimum costs, requires the formulation of a utility function $U_{q_i,n}$ of decision maker n for the alternatives

[4] The shares of semi-trailers regarding laden journeys are estimated based upon data gained from the research project 'Motor traffic in Germany 2010' (WVI – Prof. Dr. Wermuth Verkehrsforschung und Infrastrukturplanung GmbH et al. (2012)).

in S_n. Adding an i.i.d. extreme value type I distributed stochastic component $\varepsilon_{q_i,n}$, a vector with the separately estimated latent classes $\boldsymbol{L}_n$ and interpreting increasing costs per ton for shipment size class q_i as negative utility results in the following parametrized function:

$$-U_{q_i,n} = \frac{C_n(q_i)}{Q_n} + \varepsilon_{q_i,n} = \alpha_{q_i} + \beta_{q_i,1} \cdot d_n + \beta_{q_i,2} \cdot \frac{1}{Q_n} + \beta_{q_i,3} \cdot \frac{v_n}{Q_n} + \boldsymbol{\beta}_{q_i,L} \cdot \boldsymbol{L}_n + \varepsilon_{q_i,n} \quad (8)$$

The inventory in transit (bounded capital times transport time) has been dropped in the estimation of the total logistics cost functions as the parameters in initial estimations were insignificant. This is also in accordance with theoretical considerations predicting independence between the shipment size and the inventory in transit, which is reflected by equation (2). As we consider truck transports exclusively, the transport times do not significantly vary between the different shipment sizes although transhipments take place. Dealing with different modes, the inventory in transit may become a significant influencing factor (Combes and Tavasszy (2016)).

According to the discrete choice theory, in association with the random utility approach, a shipment size class q_i gets chosen by an individual *n* if $U_{q_i,n} \geq U_{q_k,n} \quad \forall k \in S_n,\ i \neq k$. This means that the sum of the observable and unobservable parts of the costs of a chosen alternative has to be lower than the costs of all other alternatives in the choice set S_n.

3.2 Theoretical framework of homogeneous segments

In general, the total amount of transported goods per period, which represents the firm's commodity flows, empirically contributes a big portion to the explanatory power of the shipment size choice (Abate and de Jong (2014), Combes (2009)). In other words, varying periodic demand is the main cause of heterogeneity regarding the shipment size choice. Inherent approaches to control for effects not captured by the variables of the estimated total logistics cost function are given by the integration of variables describing different characteristics of the transport and/or using standard classifications of transported goods. These nominally scaled variables are integrated as dummy variables into the utility equation measuring the effect for the different characteristics. The standard classification of transported goods in the European Union, represented by the NST 2007 taxonomy and primarily used for mode choice or shipment size choice models within in the European Union, is made up of 20 different commodity types at the highest level of aggregation. On closer examination of the commodity types, they tend to contain a lot of heterogeneous actors and transports within the individual categories. Liedtke and Schepperle (2004) mention the example of milk, where milk as a liquid bulk good is assigned to the same category as palletized milk, despite representing a completely different level of the production chain. In addition to the suboptimal classification of commodity types with respect to logistical decisions, the multiplicity of variables which capture transport characteristics makes the integration of a lot size component difficult, as data are not available for the whole population of firms. Due to the above-named aspects, the formation of subgroups with similar characteristics offers a possibility to represent the behavioral heterogeneity in a manageable way and therefore improves the shipment size choice model.

To enhance the core of the shipment size choice with homogeneous clusters we applied a so-called "exogenous segmentation" approach. According to Sharman and Roorda (2013) an exogenous segmentation is characterized as a procedure which conducts the segmentation and the model estimation sequentially with no feedback between the two estimation processes. As there is no feedback from segmentation to modeling there is no guarantee of minimizing heterogeneity within, and maximizing heterogeneity between, the derived segments. Contrary to that, an endogenous segmentation approach simultaneously estimates segments and the respective choice model for each segment.

One method to determine segments exogenously is the application of the Latent Class Analysis (LCA). Latent classes are thereby characterized as an unobservable and categorical variable which has a nominal level of measurement and is measured by categorical indicators. The notation in this paper is associated with the one from Collins and Lanza (2013). The basis of the LCA is a contingency table, which contains the response categories $r_j = 1, \ldots, R_j$ of the indicator variables $j = 1, \ldots, J$ and their absolute frequencies. Each cell of the multidimensional contingency table represents a specific response pattern $\boldsymbol{y} = (r_j, \ldots, r_J)$ recording the answer to each of the J indicator variables. The LCA clusters individuals with similar response patterns and is essentially determined by two sets of parameters: the latent class prevalences γ describing the proportion of individuals in the respective class and the item-response probabilities ρ. The item-response probabilities express the relation between each indicator variable and the latent classes, thus providing the basis for the interpretation of the latent classes. Let $\boldsymbol{Y}$ be the matrix with all possible response patterns $\boldsymbol{y}$ and L be the set of all latent classes $c = 1, \ldots, C$. The core of the LCA is the given by the probability of a specific response pattern

$$P\,(\boldsymbol{Y} = \boldsymbol{y}) = \sum_{c=1}^{C} \gamma_c \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{I(y_j=r_j)} \tag{9}$$

where γ_c is the latent class prevalence of class c, $\rho_{j,r_j|c}$ is the item-response probability for responding r_j to indicator variable j conditional on the membership to class c, and $I(y_j = r_j)$ represents an indicator variable, being one if the response y_j on variable j is given by r_j and zero otherwise. The parameters are estimated by maximizing the log-likelihood function of equation (9) via an expectation-maximization algorithm. Based on the results of the LCA, each individual is classified into the group for which it reaches the maximum membership probability

$$\begin{aligned} P\,(L = c|\boldsymbol{Y} = \boldsymbol{y}) &= \frac{P\,(\boldsymbol{Y} = \boldsymbol{y}|\, L = c) P\,(L = c)}{P\,(\boldsymbol{Y} = \boldsymbol{y})} \\ &= \frac{\left(\prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{I(y_j=r_j)}\right) \gamma_c}{\sum_{c=1}^{C} \gamma_c \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j|c}^{I(y_j=r_j)}}. \end{aligned} \tag{10}$$

These classes are integrated – as can be seen in equation (8) – by adding supplements and deductions to the utility of each alternative, depending on the membership in the respective class of each transport and the related decision maker n. These supplements and deductions account for effects which are not captured by the other observed variables of the total logistics cost function. Such a formulation assumes that the decision makers choose according to the same cost-minimizing pattern and that deviations from this can be captured by the latent classes or disappear in the error term. In the context of shipment size choice, these could be either the above-mentioned warehousing costs, which strongly depend on the characteristics of the commodities, or the fixed and variable transport costs, which might be affected by factors regarding the commodity types or the commodity characteristics. This is formally represented by the vector $\boldsymbol{L}_n$ in this model, which includes the classes with the respectively assigned transports. Although exogenously derived clusters neither guarantee the maximization of homogeneity within the derived segments nor the maximization of heterogeneity between the segments, they lead in our model to an intuitive interpretation of the segments themselves and to a meaningful enrichment to the analysis of shipment size decisions.

The results of the LCA are subsequently integrated into the choice model. Let $L_{c,n}$ be a binary variable indicating if shipper n belongs to the estimated latent classes $c = 1, \dots, C$ then are the different variables $L_{c,n}, c = 1, \dots, C$ part of the attribute vector $\boldsymbol{L}_n$. For each actor n and each class, the vector $\boldsymbol{L}_n$ therefore contains information about the membership which is mathematically expressed by zeros and ones. The above-mentioned supplements and deductions to the utility are then added in reference to a specific base class if, and only if, the entry in the matrix of shipper n is given by one.

4. Data and descriptive analysis

Revealed preference (RP) data provide the base frame of both the survey above-mentioned and the model presented in this paper.[5] The chosen enterprises of a quota sample were drawn from a German-wide business directory with about 10,000 addresses, whereby unsuitable members of the sample were excluded through multi-criteria screening. Respondents were classed as unsuitable if one of the following aspects were answered in the negative:

- Does the firm's scope of decision-making cover the mode choice?
- Are transports mainly conducted with an interregional scope? (above 100km)
- Are the shipment sizes mostly above two tons?
- Are basic alternative modes selectable?

The observations of the dataset used were ascertained via computer-assisted personal interviews (CAPI) with responsible logistics employees of shipping and receiving companies from all areas of extraction of raw materials, manufacturing, and wholesale.

In each interview, two representative transport cases and the corresponding attributes were recorded. The attributes contained information about the type, the weight, the value and the properties of the commodity. Relevant properties of the transported goods described the handling during the transport processes as well as other characteristics of the commodities which potentially influenced the execution of the transport. Further, the frequency of the transports and the position within the logistic chain were logged. In addition to this, commodity information and the duration, distance and costs of each transport for the respective mode were ascertained. Restricting the model to road transports with realistic shipment sizes (in this case: $q_n \leq 48t$) decreases the amount of observations to 794.

Table 1 presents the summary statistics of the main variables of interest which are initially revealed by the EOQ model. The constant and continuous flow of goods Q_n was not

[5] Within this survey stated preference (SP) experiments were also performed based on the information obtained from the revealed preference part. The survey used can be accessed via the corresponding report of the research project.

directly inquired in the datasets. It was instead calculated from the shipment size of the representative transports and their frequency of occurrence per year. The value density v_n was calculated by dividing the value of the transported goods by the shipment size. Because of the occurrence of missing values, the calculation of the flow of goods per year and the value density finally narrowed the data basis to 487 useable observations. Additionally used variables in the econometric estimation of the model are the transport distance d_n and the attributes regarding the properties of the transported goods presented in table 1.

Table 1: Summary statistics of used variables for 487 observations

Variable	*Definition*	*Mean/Freq*	*Std. dev.*
q	Shipment size (t)	13.47	9.75
Q	Flow of goods (t/year)	1 845.45	2 369.30
v	Value density (€/t)	11 311.84	55 958.44
1/Q	Inverse of Flow of goods	0.0038	0.0076
v/Q	Relation of value density and flow of goods	69.53	414.67
d	Haulage distance (km)	447.81	363.40
Fragile	1 if commodity is fragile	76	
Valuable	1 if commodity is extremely valuable	175	
Bulky	1 if commodity is bulky (unhandy, voluminous)	150	
Temperature	1 if commodity must be handled in temperature-controlled manner	48	
Food	1 if commodity is a food product	58	
Dangerous	1 if commodity is inflammable, explosive, poisonous, caustic etc.	62	
Bulk cargo	1 if commodity is unpacked bulk cargo	35	
Liquid	1 if commodity is unpacked liquid good	5	
Standard	1 if commodity is transported in standard unit loads	213	
Custom	1 if commodity is a custom-made item	84	
Accumulation	1 if commodity is an accumulation of several articles	170	

Note: Regarding the categorical attributes, no responses as well as multiple responses were possible.

Table 2 contains information about the shipment size classification and the distribution of the continuous variables in each of the classes. The positive influence of the total flow of goods and the negative impact of the value density revealed by equation (2) tend to become apparent by considering the values and the distribution in each shipment size category. While the mean and the quantiles of the annual flow of goods Q_n throughout increase with the shipment size categories, the mean and the quantiles of the value density v_n decrease in accordance with the EOQ formula. It is obvious that the distributions of the variables in each class are right-skewed. On top of that, the variances of Q_n increase with the class width, whereas the statistical scatter regarding value density v_n of the commodities decrease with the shipment size classes. Also, the inverse of Q_n and the ratio v_n/Q_n show a varying dispersion whereby the relative variance of the value density is different to the variance of the annual flow of goods. These insights are also quite intuitive, as the growing category widths can cause stochastic dependency of the error terms $\varepsilon_{q_i,n}$ and the variables used, which is associated with the possible violation of a main assumption of the logit model: the error terms $\varepsilon_{q_i,n}$ are independent and identically Gumbel distributed. However, we proceeded by taking the natural logarithm of the variables cushioning the distributions nearly to a bell-shaped curve and also approaches the variance between the different shipment size classes. Shrinking the variance of the explaining variables affects the variance of the entire cost expression, which therefore tends to cushion the variance of the error terms.

Table 2: Shipment size categories and distribution of variables

q_i		Q_n	v_n	$1/Q_n$	v_n/Q_n	$\ln(1/Q_n)$	$\ln(v_n/Q_n)$	d_n	No. obs.
< 3t	*Min*	*12*	*366.7*	*0.00133*	*0.815*	*-6.620*	*-0.2048*	*40*	*104*
	Q1	*100*	*2 250*	*0.00333*	*9.944*	*-5.704*	*2.2970*	*200*	
	Med.	*150*	*5 000*	*0.00667*	*33.333*	*-5.011*	*3.5066*	*303*	
	Mean	*238*	*18 544*	*0.01045*	*250.54*	*-5.055*	*3.5663*	*378*	
	Q3	*300*	*10 000*	*0.01000*	*101.04*	*-4.605*	*4.6154*	*500*	
	Max	*750*	*500 000*	*0.08333*	*5 000*	*-2.485*	*8.5172*	*1 480*	
3t – 12t	*Min*	*30*	*43.6*	*0.00033*	*0.0159*	*-8.006*	*-4.1434*	*5*	*154*
	Q1	*250*	*1 531*	*0.00080*	*2*	*-7.131*	*0.6931*	*205*	
	Med.	*600*	*3 661*	*0.00167*	*5.7060*	*-6.397*	*1.7414*	*398*	
	Mean	*892*	*14 021*	*0.00358*	*42.602*	*-6.303*	*1.9492*	*450*	
	Q3	*1 250*	*7 500*	*0.00400*	*21.276*	*-5.521*	*3.0570*	*550*	
	Max	*3 000*	*625 000*	*0.03333*	*1 562.5*	*-3.401*	*7.3540*	*3 000*	
> 12t	*Min*	*90*	*3.7*	*0.00004*	*0.0005*	*-10.13*	*-7.5090*	*20*	*229*
	Q1	*1 100*	*500*	*0.00018*	*0.1736*	*-8.613*	*-1.7509*	*220*	
	Med.	*2 400*	*1 250*	*0.00042*	*0.6250*	*-7.783*	*-0.4700*	*380*	
	Mean	*3 217*	*6 206*	*0.00094*	*5.4366*	*-7.648*	*-0.6090*	*478*	
	Q3	*5 500*	*3 000*	*0.00091*	*1.8333*	*-7.003*	*0.6061*	*600*	
	Max	*25 000*	*752 381*	*0.01111*	*358.28*	*-4.500*	*5.8813*	*3 000*	

5. Model estimation

In this section several multinomial logit models based on the specification of equation (8) have been estimated. First, models without latent classes, which represent the core of the shipment size choice, will be presented; the enhanced models with latent classes are given in the second subsection. We used the statistical software “R” with its supplemental packages “mlogit” and “poLCA” for the estimation of the models.

5.1 Basic model

6. Table 3: Shipment size choice model 1 – non-standardized variables.

	q_2: $3t - 12t$	q_3: $> 12t$
Constant	0.6219*	1.6698***
	(0.2525)	(0.2765)
d_n	0.0013*	0.0020***
	(0.0005)	(0.0006)
$1/Q_n$	-118.47***	-639.68***
	(25.670)	(85.497)
v_n/Q_n	-0.0006	-0.0090*
	(0.0005)	(0.0044)

Final log-likelihood: -395.26
McFadden R^2 (ρ^2): 0.2260
Adj. McFadden $\overline{R}^2$ ($\overline{\rho}^2$): 0.2103
Likelihood ratio test: χ^2 = 230.30 (p.value = < 2.22e-16)

7. Note: Significance levels: . p<0.1; * p<0.05; ** p<0.01; *** p<0.001; 487 observations; standard errors are given in brackets.

Table 3 shows the estimation results with d_n, $\frac{1}{Q_n}$ and $\frac{v_n}{Q_n}$ as independent variables. The value 0.23 of ρ^2 indicates – despite the menace of heterogeneity – a well-performing shipment size choice model which was estimated in relation to the smallest shipment size class with its referencing parameter values equal to zero. Every coefficient has the expected sign respective the expected order of magnitude. The ratio of value density and annual flow of goods does not have a statistically significant impact, whereas the other estimated coefficients are highly significant. As can be seen, the constants have a positive sign and an increasing order regarding the shipment size classes, which describe the average influences of non-considered attributes such as the transportation fixed costs $\frac{F_n}{q_i}$.

The positive sign and the order of magnitude of the constants show on average decreasing costs per unit as shipment size increases. A decreasing annual flow of goods leads to an increase of the inverse $\frac{1}{Q_n}$, which induces higher costs for higher shipment size categories. This goes along with the theoretically positive impact of the annual flow revealed by equation (3). The transport distance d_n as an approximation of the variable transport costs has a positive influence on the shipment size choice, which is in accordance with other studies (Abate and de Jong (2014), Combes (2009), Jansson and Shneerson (1982)). In general there is some dissent on the interpretation of the distance effect. On the one hand, the increasing coefficients in our estimation can be interpreted as a concave variable cost function, having a declining slope for bigger shipment sizes. Abate and De Jong (2014) explain this effect by less than proportional increasing fuel/time cost per shipment for larger vehicles. On the other hand, this effect might be caused by cultural properties or safety needs which take the increasing risk of delays accompanying longer haulage distances into account. Combes (2009) illustrates the effect with a potential decoupling of production location and regional retail center via a regional distribution center. If in general the transports have larger shipment sizes from the production location to the distribution center and they are dispatched, the positive influence of the transport distance could be explained.

Table 4: Shipment size choice model 1 – natural logarithm of $\frac{1}{Q_n}$ and $\frac{v_n}{Q_n}$.

	q_2: $3t - 12t$	q_3: $> 12t$
Constant	-4.9081***	-8.4198***
	(1.2925)	(1.6387)
d_n	0.0015**	0.0028***
	(0.0005)	(0.0006)
$\ln(1/Q_n)$	-0.8993***	-1.4232***
	(0.1887)	(0.2347)
$\ln(v_n/Q_n)$	-0.1434	-0.7027***
	(0.1092)	(0.1413)

Final log-likelihood: -330.03
McFadden R^2 (ρ^2): 0.35371
Adj. McFadden $\overline{R}^2$ ($\overline{\rho}^2$): 0.3380
Likelihood ratio test: χ^2 = 361.25 (p.value = < 2.22e-16)

Note: Significance levels: . p<0.1; * p<0.05; ** p<0.01; *** p<0.001; 487 observations; standard errors are given in brackets.

As mentioned before, the danger of a misspecification concerning the potential heterogeneity is omnipresent. We therefore estimated the model again, transforming the independent variables $\frac{1}{Q_n}$ and $\frac{v_n}{Q_n}$ by taking the natural logarithm. Table 4 contains the results, which show an improved performance, although we are moving away from the core of the model represented by the total logistics cost per ton. The alternative-specific constants have become negative with relatively high values. This is induced by the transformation of the variables, which decreases small values (< 1) strongly to negative values and flattens out the slope with increasing values. Again all signs and orders of magnitude of the coefficients are as expected, although we also must point out that the coefficient for the $\frac{v_n}{Q_n}$-relationship regarding the second shipment size class has no significant influence. The coefficient of $\frac{v_n}{Q_n}$ explains the variation of the cost per ton not already covered by the reciprocal annual flow of goods. The impact is negative but has a relatively weak influence on the decision of the shipment size choice. This result supports the conclusion that the costs for capital tied up do not completely pervade the shipment size choice.

7.1 Enhancement of the discretized EOQ-model with exogenously derived latent classes of shipments' attributes

The reasonable explanatory power of the discretized EOQ model enables the enhancement of the model with further characteristics influencing the shipment size systemically. In light of the already acknowledged influences of commodity characteristics, commodity types and/or economic activities taking place at the origin or the destination of the transport on the shipment size choice behavior the model will be expanded in this section. We therefore use an approach which has to our knowledge not been undertaken up to now in the modelling of shipment size choice: clustering the realized transports on the basis of similar characteristics, in order to model the shipment size choice behavior more accurately, and to reduce the dimensions of possible heterogeneity with respect to practicability in large-scale freight transport models.

To this end, we estimated different specifications of classification criteria including commodity characteristics and activities such as storage, transshipment or further processing of the goods. The variables describing the position within the logistic chain or the activities at the origin and destination of the transports neither improved the explanatory power of the model by individual integration nor could be used properly for the grouping of transports[6]. A distinction of receivers' and shippers' questionnaires, resulting in different attributes, lead, with respect to an algorithmic segmentation, to a bias of separation concerning shippers and receivers instead of the respective attributes. This is caused by the nature of the LCA grouping individuals according to similar response patterns, which in this case would be different for shipper and receiver, and would therefore bias the segmentation result. Furthermore, a merge of shipper and receiver attributes is problematic, as they are slightly different in terms of the content which the respective items cover. While estimating different specifications for the LCA, it was found that the commodity characteristics as a basis for segmentation lead to the best result regarding interpretability, the trade-off between estimated parameters and number of observations, and the variance reduction in the context of the shipment size choice. Although the economic activity at the origin and

[6] Attributes describing the activity at the origin or the destination on the side of shippers are given, for example, by upstream or downstream transhipments and storage at the destination. For instance, receivers were asked if the origin of the transport was the manufacturer, wholesaler or a distribution center, and if the good was processed immediately (same day) after the stock receipt. The integration of these categorical variables into the shipment size choice model did not show up significantly.

destination of the transports in this study – also due to the relatively low number of observations and the increasing number of parameter estimates with further attributes – was not appropriate for segmentation, the effect of an integration of such attributes should not be neglected for prospective investigations of segmentation regarding shipment size choice.

Table 5: Information criteria of LCA's.

No. classes	*log-likelihood*	*BIC*
2	-2 284.309	4 710.949
3	-2 232.398	4 681.386
4	-2 191.574	4 673.995
5	-2 175.075	4 715.257
6	-2 158.490	4 756.347
7	-2 146.076	4 805.778

Because of the above-mentioned reasons, we used an LCA based on the characteristics of the transported goods to group the individual transports. Due to the propensity of the expectation maximization algorithm to get stuck in local optima, we repeated the estimation 1000 times with randomly chosen initial values for each number of classes. The determination of the proper number of classes is not endogenous. We therefore had to decide by reference to the values of the Bayes information criterion (BIC) which are displayed in table 5 for several counts of classes. We picked the classification with four classes, for which the results of the LCA can be obtained in table 6. The conditional probabilities of answering "Yes" to a specific indicator denote the probability to answer "Yes" if an individual is assigned to the respective class. High or low values – also in comparison to the other classes - are therefore characteristic and affect the interpretation of it. We marked in bold notation the conditional probabilities being characteristic for the class. This means, first, that they are relatively high compared to other probabilities in the same class and, second, quite different regarding the same indicators across the other estimated classes. The bold values are therefore used to apply the necessary subjective interpretation of each class.

We called class 1 "(temperature-controlled) food products", as every individual belonging to this class will answer assuredly "Yes" to the "food" item and more than 60% to the "temperature" item. Also, the transported goods in this class are never dangerous or bulky and are transported with probability 0.555 on standard unit loads. Typical examples for this class are palletized flour, milk, meat or miscellaneous frozen food. Class 2 is mainly characterized by the items "standard" and "dangerous"; we thus interpreted it as "miscellaneous standard cargo," including dangerous commodities transported in the same. This class also inherits the largest proportion of the population, with 36%. Included in the second class are transports containing a wide range of goods, such as pharmaceutical and chemical products, machine parts and various textiles. The next class has high loads on the items "fragile", "valuable", "bulky" and "custom," which we have interpreted as "special goods," and mostly comprises the mechanical engineering and manufactured goods sector. The proportion of this class, about 35%, seems relatively high and indicates a general overrepresentation of it in the sample. Exemplary transports of "special goods" are machines, cars, furniture, and diverse metal and electronic products. Finally, we named class 4 "unpacked bulk goods," as it has both relatively high probabilities for the "bulk cargo" and "liquid" items and also low probabilities of "custom", "standard", "accumulation" and "fragile". Representatives of this class are mainly raw materials such as wood, cement, paper, chemicals, steel or synthetic materials.

Table 6: LCA result with four classes.

	Class 1	*Class 2*	*Class 3*	*Class 4*
γ_c	*0.1070*	*0.3600*	*0.3475*	*0.1854*
	Conditional probability of answering "Yes"			
Fragile	0.1727	0.0634	**0.3302**	**0.0000**
Valuable	0.2631	0.2395	**0.6373**	0.1266
Bulky	**0.0000**	0.2376	**0.4942**	0.2734
Temperature	**0.6331**	0.0855	0.0000	0.0000
Food	**1.0000**	0.0000	0.0000	0.0651
Dangerous	**0.0000**	**0.2079**	0.0599	0.1707
Bulk cargo	0.0398	**0.0000**	0.0000	**0.3647**
Liquid	0.0555	**0.0000**	0.0000	**0.1800**
Standard	**0.5550**	**0.8113**	0.2308	**0.0305**
Custom	0.0384	0.0841	**0.3974**	**0.0000**
Accumulation	0.4781	0.4134	0.3992	**0.0557**
Observations: 487				
Parameter estimated:47				

Note: Bold parameters are characteristic for the class with respect to the class itself and across classes.

The assignment of the different transports to the classes is accomplished by calculating the membership probability shown in equation (10) and taking the maximum of the probabilities as allocation rule. This procedure denotes a probabilistic approach in contrast to a deterministic assignment and therefore needs a validation of applicability. One established indicator is the so-called Odds of Correct Classification (OCC), which sets the average probability of the individuals assigned to a class in ratio to the general proportions of each class and has a ratio higher than five as threshold (Nagin (2005)). In our classification this is reached for every class, which can be seen in table 7.

Table 7: Key figures of classification.

Class	*Mean*	*Variance*	*OCC*
1	0.93	0.02	109.45
2	0.85	0.02	10.01
3	0.84	0.02	10.12
4	0.75	0.04	13.11

Note: This table includes means and variances of the maximum probabilities from the assigned individuals.

The next step is to integrate the identified classes into the into the discrete choice model for shipment size. Table 8 shows the estimation results for the comprehensive model. The performance of the enhanced model increased, which can be deducted from the value of $\overline{\rho}^2$. Also, all coefficients become at least significant with respect to 0.1 level.

The impacts of the latent classes have all been estimated in comparison to the first class of "(temperature-controlled) food products". As expected, all other classes tend to choose bigger shipment sizes, as the warehousing of the goods is more expensive and the perishability does not allow high order quantities. At first glance, the order of the coefficients for the classes is somehow contradictory. "Miscellaneous standard cargo" has in comparison with the group with "special goods" smaller coefficients, which from a superficial point of view is unexpected. But the interviews showed that the members of the German mechanical engineering sector need to use higher shipment sizes because of the weight of their products. Also the goods in class 2 are mainly transported in standard unit loads, which supports the splitting into smaller shipment sizes, in contrast to the goods of class 3. Finally, the “unpacked bulk goods” show the highest effect on the largest shipment size class. This behavior is reasonable, as the warehousing costs should be relatively low and the production of the goods is performed in big batches. In general, the results reveal that the effects of the groups do not vary much with respect to the medium shipment size class. Solely the food products are less likely to be sent in medium shipment sizes.

Table 8: Shipment size choice model with latent classes

	q_2: $3t - 12t$	q_3: $> 12t$
Constant	-5.5253***	-10.027***
	(1.3670)	(1.7849)
d_n	0.0013*	0.0027***
	(0.0005)	(0.0006)
$\ln(1/Q_n)$	-0.8900***	-1.4964***
	(0.1933)	(0.2451)
$\ln(v_n/Q_n)$	-0.2239.	-0.7521***
	(0.1175)	(0.1516)
Miscellaneous standard cargo	0.9168.	1.3603*
	(0.4884)	(0.5799)
Special goods	1.2750*	1.9700**
	(0.5004)	(0.6004)
Unpacked bulk goods	1.1386*	2.4094***
	(0.5776)	(0.6741)
Log-likelihood: -321.15		
McFadden R^2 (ρ^2): 0.3711		
Adj. McFadden $\overline{R}^2$ ($\overline{\rho}^2$): 0.3482		
Likelihood ratio test: χ^2 = 379 (p.value = < 2.22e-16)		

Note: Significance levels: . p<0.1; * p<0.05; ** p<0.01; *** p<0.001; 487 observations; standard errors are given in brackets.

8. Conclusion and further research

In this paper, a discrete shipment size choice model for truck transports on the basis of discretized total logistics cost was developed, taking into account the different requirements regarding implementation in a large scale freight transport model. It has turned out that the shipment sizes are oriented on given vehicle and bundle sizes, which is concomitant with discrete technology-related and market-related factors influencing the total logistics cost. The estimation of the different models was performed on a database which was collected within the scope of the German federal infrastructure planning (BVWP). The framework used also provides possible links to other state-of-the-art decision models such the discrete mode choice, which was unfeasible in this study due to an insufficient amount of observations with modes other than truck transports. In order to reach an appropriate microscopic representation of the choice situations’ heterogeneity, the choice behavior was addressed separately for exogenously derived clusters of homogeneous transport cases. The conducted transports were classified by using a Latent Class Analysis approach on the basis of the commodity characteristics leading to a meaningful and comprehensible enrichment of the shipment size choice. Four homogeneous segments were identified, which were integrated as an additional component to the systematic framework of the total logistics cost formulation.

From a transport analyst's point of view, the interesting results are that the volumes of the underlying commodity flows can explain the choice of shipment sizes to a large extent. The integration of the latent classes improved the model and provided reasonable signs and orders of magnitude for the estimated coefficients. Being estimated on a sample of very heterogeneous transport cases, the latent class model qualifies for incorporation in operational large-scale freight transport models. Moreover, the latent class part shows that a shipment size model is a way to incorporate logistics aspects into freight transport models on the required coarse-grained level of detail. It remains to be examined whether the same categorization applies to other decisions drawn on the shipments such as transport mode choice. Moreover, the influence of distance hints that spatial aspects play a role in the choice of shipment sizes.

A ubiquitous problem in the context of this model is the heteroscedasticity, which is caused by the categorization of the shipment size classes and should be attended to by future research in a systematic way, potentially accompanied by a modification of the total logistics cost to control for the increasing class widths. Furthermore, the integration of economic activities taking place at the origin and the destination of the transports as classification criteria needs to be investigated in depth regarding further model developments. Another possible enhancement could be given by the usage of an endogenous segmentation approach, or the estimation of different models for exogenously derived segments, which lead to variations of the estimated parameters across the groups of decision-makers and reveals different influences on the costs for different groups. In this study, the central assumption was that the decision makers would, en masse, choose the shipment size according to the total logistics cost. Further information reached by the segmentation is added to the model via supplements and deductions for the respective class membership. Altering this assumption to different models for the segments may cause unexpected outcomes, contradicting the rational core which is represented by the total logistics cost approach. Therefore, the investigation of various models for different segments has to be treated carefully. Nevertheless, both the implementation of different segmentation approaches and the usage of further classification attributes need an adequate data base enabling a reliable estimation, which was not given in this study. Finally, an additional aspect which should be considered is obviously extending the model to different modes of transport and combining shipment size choice with mode choice.

Acknowledgements

We are very thankful to the editor and two anonymous referees for helpful comments. Also a thank-you to the participants of the URBE 2015 – Urban Freight and Behavior Change conference for the comments. The study was funded by the Federal Ministry of Transport and digital Infrastructure (BMVI) within the scope of the German federal infrastructure planning (BVWP).

References

Abate, M. and De Jong, G., (2014). The optimal shipment size and truck size choice – The allocation of trucks across hauls. *Transportation Research Part A, 59*, pp. 262-277.

Abdelwahab, W. and Sargiuos, M., (1992). Modelling the demand for freight transport: a new approach. *Journal of Transport Economics and Policy*, 26 (1), pp. 49-70.

Arunotayanun, K., (2009). Modelling freight supplier behavior and response. PhD thesis. Centre for Transport Studies, Imperial College of London.

Baumol, W.J. and Vinod, H.D., (1970). An Inventory Theoretic Model of Freight Transport Demand. *Management Science*, 16 (7), pp. 413- 421.
Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.
Bhat, C.R., (1997). An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. *Transportation Science*, Vol. 31, No. 1, pp. 34-48.
BVU – Beratergruppe Verkehr + Umwelt and TNS Infratest (2014). *Entwicklung eines Modells zur Berechnung von modalen Verlagerungen im Güterverkehr für die Ableitung konsistenter Bewertungsansätze für die Bundesverkehrswegeplanung*, FE-Projekt 96.0996/2011, vorläufiger Endbericht. http://www.bmvi.de/SharedDocs/DE/Anlage/VerkehrUndMobilitaet/bvwp-2015-modalwahl-zeit-zuverlaessigkeit-gueterverkehr-vorl-schlussbericht.pdf?__blob=publicationFile
Collins, L. M. and Lanza, S. T., (2013). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (Vol. 718). John Wiley & Sons.
Combes, F., Harache, J., Koning, M., and Morau, E. (2016). Empirical Analysis of Freight Transport Prices Using the French Shipper Survey ECHO. In *Commercial Transport* (pp. 321-335). Springer International Publishing.
Combes, F. and Tavasszy, L. A. (2016). Inventory theory, mode choice and network structure in freight transport. *EJTIR*, 16(1), 38-52.
Combes, F., (2009). The shipment size in interurban freight transport: from microeconomic modelling to operational simulation. Ph.d. Thesis, Université Paris-Est Ecole Doctorale Ville Environnement, Laboratoire Ville Mobilité Transport.
De Jong, G. and Ben-Akiva, M., (2007). A micro-simulation model of shipment size and transport chain choice. *Transportation Research Part B*, Vol. 41, 2007, pp. 950-965.
De Jong, G.C. and Johnson, D., (2009). Discrete mode and discrete or continuous shipment size choice in freight transport in Sweden. Presented at European Transport Conference, Netherlands, 2009.
De Jong, G., Ben Akiva, M. and Baak, J. (2010). *Method Report - Logistics Model in the Swedish National Freight Model System (Version 2)*, Deliverable 6B for the Samgods Group
Gopinath, D.A., (1995). Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand. MIT Ph.D. Thesis.
Guilbault, M. and Soppé, M., (2009). "Grandes tendances d'évolution des pratiques de transport et logistique au travers des enquêtes chargeurs 1988 et ECHO 2004". In M. Guilbault, M. Soppé (Eds), *Connaissance des chaînes de transport de marchandises et de leurs déterminants logistiques* (pp 33-56). Actes INRETS no 121.
Hall, R. W., (1985). Dependence between shipment size and mode in freight transportation. *Transportation Science*, Vol. 19, No. 4, 1985, pp. 436-444.
Harris, F. W. (1913). How many parts to make at once. *Factory, the Magazine of Management*, 10(2), pp. 135-152.
Holguin-Veras, J., (2002). Revealed preference analysis of commercial vehicle choice process. *Journal of Transportation Engineering*, Vol. 128, No. 4, 2002, pp. 336-346.
Jansson, J. O. and Shneerson, D. (1982). The optimal ship size. *Journal of transport economics and policy*, 217-238.
Kawamura, K., Mohammadian, A., Samimi, A., Pourabdollahi, Z. and Klekotka, M. (2010). *Development of Freight Policy Analysis Tool for Northeastern Illinois and the United States*, National Center for Freight and Infrastructure Research and Education (CFIRE), Report No. CFIRE 02-29.

Liedtke, G. and Schepperle, H. (2004). Segmentation of the transportation market with regard to activity-based freight transport modelling. *International Journal of Logistics Research and Applications*, 7(3), 199-218.

Nagin, D. (2005). *Group-based modeling of development*. Harvard University Press.

Pourabdollahi, Z., Karimi, B. and Mohammadian, A., (2013). Joint Model of Freight Mode and Shipment Size Choice. *Transportation Research Record: Journal of the Transportation Research Board*, (2378), pp. 84-91.

Sharman, B. W. and Roorda, M. J. (2013). Multilevel modelling of commercial vehicle inter-arrival duration using GPS data. *Transportation Research Part E: Logistics and Transportation Review*, 56, 94-107.

Windisch, E., de Jong, G. C., van Nes, R. and Hoogendoorn S. P., (2010). A disaggregate freight transport model of transport chain and shipment size choice. Presented at the European Transport Conference, Glasgow, Scotland, UK, 2010.

Wisetjindawat, W., Sano, K. and Matsumoto, S., (2005). Supply Chain Simulation for Modeling the Interactions in Freight Movement. *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 6, pp. 2991 – 3004.

WVI – Prof. Dr. Wermuth Verkehrsforschung und Infrastrukturplanung GmbH, IVT - Institut für angewandte Verkehrs- und Tourismusforschung e.V., DLR - Deutsches Zentrum für Luft- und Raumfahrt e.V., KBA - Kraftfahrt-Bundesamt (2012), *Kraftfahrzeugverkehr in Deutschland 2010*, Projekt-Nr. 70.0829/2008, Schlussbericht.